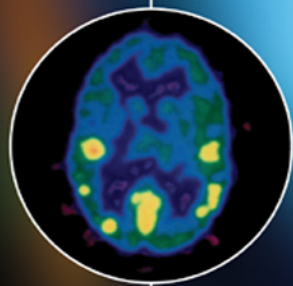
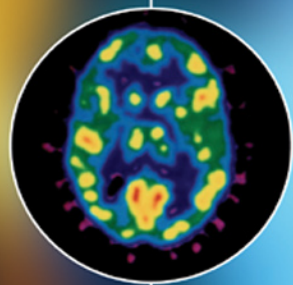
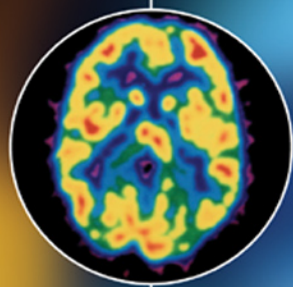


# HANDBOOK OF PSYCHOLINGUISTICS

SECOND EDITION



Edited by  
Matthew J. Traxler and  
Morton Ann Gernsbacher



**HANDBOOK OF  
PSYCHOLINGUISTICS**  
2nd EDITION

This page intentionally left blank

# HANDBOOK OF PSYCHOLINGUISTICS

2nd EDITION

EDITED BY

**MATTHEW J. TRAXLER**  
*Department of Psychology,  
University of California at Davis,  
Davis, USA*

**MORTON A. GERNSBACHER**  
*University of Wisconsin,  
Madison, USA*



ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford  
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo  
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier  
84 Theobald's Road, London WC1X 8RR, UK  
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands  
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK  
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA  
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA

First edition 1994  
Second edition 2006

Copyright © 2006 Elsevier Inc. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: [permissions@elsevier.com](mailto:permissions@elsevier.com). Alternatively you can submit your request online by visiting the Elsevier web site at <http://elsevier.com/locate/permissions>, and selecting *Obtaining permission to use Elsevier material*

#### Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

ISBN-13: 978-0-12-369374-7  
ISBN-10: 0-12-369374-8

For information on all Academic Press publications  
visit our website at [books.elsevier.com](http://books.elsevier.com)

Printed and bound in USA

06 07 08 09 10 10 9 8 7 6 5 4 3 2 1

Working together to grow  
libraries in developing countries

[www.elsevier.com](http://www.elsevier.com) | [www.bookaid.org](http://www.bookaid.org) | [www.sabre.org](http://www.sabre.org)

ELSEVIER

BOOK AID  
International

Sabre Foundation

## TABLE OF CONTENTS

Preface	vii
List of Contributors	ix
1. Observations on the Past and Future of Psycholinguistics <i>Alan Garnham, Simon Garrod, and Anthony Sanford</i>	1
<b>Section 1: Language Production</b>	<b>19</b>
2. Properties of Spoken Language Production <i>Zenzi M. Griffin and Victor S. Ferreira</i>	21
3. Syntax and Production <i>Fernanda Ferreira and Paul E. Engelhardt</i>	61
4. Speech Disorders <i>Gary Weismer</i>	93
5. Functional Neuroimaging of Speech Production <i>Thomas A. Zeffiro and Jennifer L. Frymiare</i>	125
<b>Section 2: Language Comprehension</b>	<b>151</b>
6. Speech Perception within a Biologically Realistic Information-Theoretic Framework <i>Keith R. Kluender and Michael Kiefte</i>	153
7. The Perception of Speech <i>Jennifer S. Pardo and Robert E. Remez</i>	201
8. Spoken Word Recognition <i>Delphine Dahan and James S. Magnuson</i>	249
9. Visual Word Recognition: The Journey from Features to Meaning (A Travel Update) <i>David A. Balota, Melvin J. Yap, and Michael J. Cortese</i>	285
10. Lexical Processing and Sentence Context Effects <i>Robin K. Morris</i>	377
11. Semantic Memory <i>Beth A. Ober and Gregory K. Shenaut</i>	403
12. Syntactic Parsing <i>Martin J. Pickering and Roger P. G. van Gompel</i>	455
13. Prosody <i>Shari Speer and Allison Blodgett</i>	505

14.	The Syntax–Semantic Interface: On-Line Composition of Sentence Meaning <i>Liina Pylkkänen and Brian McElree</i>	539
15.	Constraint Satisfaction Accounts of Lexical and Sentence Comprehension <i>Maryellen C. MacDonald and Mark S. Seidenberg</i>	581
16.	Eye-Movement Control in Reading <i>Keith Rayner and Alexander Pollatsek</i>	613
17.	Psycholinguistics Electrified II (1994–2005) <i>Marta Kutas, Cyma K. Van Petten, and Robert Kluender</i>	659
18.	Discourse Comprehension <i>Rolf A. Zwaan and David N. Rapp</i>	725
19.	Neuroimaging Contributions to the Understanding of Discourse Processes <i>Robert A. Mason and Marcel Adam Just</i>	765
20.	Comprehension Ability in Mature Readers <i>Debra L. Long, Clinton L. Johns, and Phillip E. Morris</i>	801
21.	Figurative Language <i>Raymond W. Gibbs, Jr. and Herbert L. Colston</i>	835
22.	Eye Movements and Spoken Language Comprehension <i>Michael K. Tanenhaus and John C. Trueswell</i>	863
23.	Perspective Taking and the Coordination of Meaning in Language use <i>Dale J. Barr and Boaz Keysar</i>	901
24.	Comprehension Disorders in Aphasia: The Case of Sentences that Require Syntactic Analysis <i>David Caplan and Gloria Waters</i>	939
25.	Language Processing in Bilingual Speakers <i>Ana I. Schwartz and Judith F. Kroll</i>	967
26.	Psycholinguistics and Neurolinguistic Perspectives of Sign Languages <i>David P. Corina and Heather P. Knapp</i>	1001
	<b>Section 3: Language Development</b>	<b>1025</b>
27.	Language Learning in Infancy <i>Anne Fernald and Virginia A. Marchman</i>	1027
28.	Acquisition of Syntax and Semantics <i>Stephen Crain and Rosalind Thornton</i>	1073
29.	Learning to Read <i>Richard K. Wagner, Shayne B. Piasta, and Joseph K. Torgesen</i>	1111
30.	Cognitive and Linguistic Issues in the Study of Children with Specific Language Impairment <i>Laurence B. Leonard and Patricia Deevy</i>	1143
	Index	1173

## PREFACE

It has been a lucky 13 years since the publication of the first edition of this *Handbook*. Cognitive psychology in general, and psycholinguistics in particular, have experienced tremendous growth and change during this time (an overview sketch appears in Chapter 1 by Garnham, Garrod, & Sanford). One of our goals in amassing the second edition was to document and survey the most important of these theoretical and empirical developments. To this end we recruited approximately 50 of the very best researchers and theoreticians in the field. We hope you will agree that these scholars have done a superb job, not merely describing their own research enterprises, but reviewing and explaining activity across their sub-disciplines.

Another of our main goals in creating this second edition was to sample as broadly across the field of psycholinguistics as possible. One small flaw in the first edition was an over-representation of chapters on higher level processing. Because higher level language processing is still a fascinating and critically important facet of the study of human language understanding, and because there have been tremendous advances in this sub-discipline since the previous edition, we have tried to ensure that higher-level processes are still adequately represented (e.g., chapters by Zwaan & Rapp, Mason & Just, and Gibbs). However, we hope that the second edition strikes a better balance between different areas of research.

We hope that the reader will, therefore, obtain a broad and general overview of the field, as well as an introduction to the most important recent experimental findings.<sup>1</sup> Of course, there is a great deal of important work in psycholinguistics that we have not been able to include due to space limitations.<sup>2</sup> This a regrettable but inevitable consequence of the economics of publishing.

We have organized this edition into three sections plus Garnham, Garrod, and Sanford's historical review and future directions chapter. The first section comprises four chapters on language production. The chapter by Griffin and V. Ferreira focuses on the production of words, while F. Ferreira and Englehardt's chapter focuses primarily on production of phrases and sentences. Weismer admirably reviews speech disorders. Zeffiro and Frymiare complete the section with an explanation of how to exploit fMRI techniques to study language production, an area that we are likely to see more of in the next decade.

The second and largest section reviews issues in language comprehension. Without listing each chapter individually, we note that this section is organized in a roughly

---

<sup>1</sup> Readers will also learn the meaning of the Swedish word "ko-tätaste."

<sup>2</sup> For example, I have already been comprehensively hazed for not dedicating a chapter specifically to reference assignment. It goes without saying that the next edition will have a chapter on reference assignment.



bottom-up way, starting with the lower-level processes involved in speech processing (e.g., Kluender & Kieft), leading to comprehension of individual words (e.g., Balota, Yap, & Cortese), proceeding to phrase- and sentence-level issues (e.g., Speer & Blodgett), and moving upward to discourse processing and interaction in dialogue (e.g., Barr & Keysar). Along the way, these authors review a number of methodological issues. This section culminates with a trio of chapters that review language comprehension in specific populations (i.e., aphasic patients, Caplan, & Waters; bilingual speakers, Schwartz & Kroll; and deaf signers, Corina).

The final section reviews language development, beginning with a chapter on language development in infancy by Fernald and Marchman, followed by Crain and Thornton's review of syntactic development in early childhood.<sup>3</sup> Wagner, Piasta, and Torgesen then provide concrete advice on how to teach children to read in their chapter.<sup>4</sup> We conclude with a chapter by Leonard and Deevy on specific language impairment.

A volume of this scope would not be possible without the support and cooperation of many people. We, as editors, express our strongest gratitude to each of the authors for producing such high-quality work. Sarah Oates at Elsevier has also been instrumental in bringing this volume to fruition.

Psycholinguistics has been enriched by the addition of many young and talented scientists in the past 13 years. It has also been diminished by the loss of some of our best and liveliest individuals, including Peter Jusczyk, Ino Flores d'Arcais, Elizabeth Bates, and Marica de Vincenzi. Our sadness at their passing is tempered by the fact that their contribution to our understanding of psycholinguistics endures.

*Matthew J. Traxler  
Morton Ann Gernsbacher*

---

<sup>3</sup> The reader who moves immediately from Fernald & Marchman to Crain & Thornton may experience epistemological whiplash, but that is one of the risks you take when reading about an active and exciting field like psycholinguistics.

<sup>4</sup> I have it on unimpeachable authority that Joe Torgesen was to meet with President Bush to discuss reading interventions on September 11, 2001.

## LIST OF CONTRIBUTORS

- David A. Balota* Department of Psychology, Washington University, St. Louis, MO, USA
- Dale J. Barr* University of California, Riverside, CA, USA
- Allison Blodgett* Ohio State University, Columbus, OH, USA
- David Caplan* Neuropsychology Laboratory, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
- Herbert L. Colston* University of Wisconsin, Kenosha, WI, USA
- David P. Corina* Center for Mind and Brain, University of California, Davis, CA, USA
- Michael J. Cortese* University of Nebraska at Omaha, Omaha, NE, USA
- Stephen Crain* Macquarie University, Sydney, Australia
- Delphine Dahan* Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA
- Patricia Deevy* Purdue University, West Lafayette, IN, USA
- Paul E. Engelhardt* Department of Psychology and Cognitive Science Program, Michigan State University, MI, USA
- Anne Fernald* Department of Psychology, Stanford University, Stanford, CA, USA
- Fernanda Ferreira* Department of Psychology, University of Edinburgh, UK
- Victor S. Ferreira* Department of Psychology, University of California, San Diego, CA, USA
- Jennifer L. Frymiare* Department of Psychology, University of Wisconsin, Madison, WI, USA

<i>Alan Garnham</i>	Department of Psychology, University of Sussex, Brighton, UK
<i>Simon Garrod</i>	Department of Psychology, University of Glasgow, UK
<i>Raymond W. Gibbs</i>	University of California, Santa Cruz, CA, USA
<i>Zenzi M. Griffin</i>	School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA
<i>Clinton L. Johns</i>	Department of Psychology, University of California, Davis, CA, USA
<i>Marcel Adam Just</i>	Center for Cognitive Brain Imaging, Carnegie Mellon University, Pittsburgh, PA, USA
<i>Boaz Keysar</i>	University of Chicago, Chicago, IL, USA
<i>Michael Kieft</i>	School of Human Communication Disorders, Dalhousie University, Nova Scotia, Canada
<i>Keith R. Kluender</i>	Department of Psychology, Madison, WI, USA
<i>Robert Kluender</i>	Department of Linguistics, University of California, San Diego, CA, USA
<i>Heather P. Knapp</i>	Department of Psychology, University of Washington, Seattle, Washington, DC, USA
<i>Judith F. Kroll</i>	Department of Psychology, Pennsylvania State University, PA, USA
<i>Marta Kutas</i>	Departments of Cognitive Science and Neurosciences, University of California, San Diego, CA, USA
<i>Laurence B. Leonard</i>	Purdue University, West Lafayette, IN, USA
<i>Debra L. Long</i>	Department of Psychology, University of California, Davis, CA, USA
<i>Maryellen C. MacDonald</i>	Department of Psychology, University of Wisconsin, Madison, WI, USA
<i>James S. Magnuson</i>	Department of Psychology, University of Connecticut, Storrs, CT, USA; and Haskins Laboratories, New Haven, CT, USA

- Virginia A. Marchman* Department of Psychology, Stanford University, Stanford, CA, USA
- Robert A. Mason* Center for Cognitive Brain Imaging, Carnegie Mellon University, Pittsburgh, PA, USA
- Brian McElree* Department of Psychology, New York University, NY, USA
- Phillip E. Morris* Department of Psychology, University of California, Davis, CA, USA
- Robin K. Morris* University of South Carolina, SC, USA
- Beth A. Ober* Department of Human and Community Development, University of California, Davis, CA, USA; VANCHCS, Mather, CA, USA
- Jennifer S. Pardo* Department of Psychology, Barnard College, New York, NY, USA
- Shayne B. Piasta* Florida Center for Reading Research, Florida State University, FL, USA
- Martin J. Pickering* Department of Psychology, University of Edinburgh, UK
- Alexander Pollatsek* Department of Psychology, University of Massachusetts, Amherst, MA, USA
- Liina Pylkkänen* Department of Linguistics & Department of Psychology, New York University, NY, USA
- David N. Rapp* University of Minnesota, Minneapolis, MN, USA
- Keith Rayner* Department of Psychology, University of Massachusetts, Amherst, MA, USA
- Robert E. Remez* Department of Psychology, Barnard College, New York, NY, USA
- Anthony Sanford* Department of Psychology, University of Glasgow, UK
- Ana I. Schwartz* The University of Texas at El Paso, El Paso, TX, USA
- Mark S. Seidenberg* Department of Psychology, University of Wisconsin, Madison, WI, USA

- Gregory K. Shenaut* Department of Human and Community Development,  
University of California, Davis, CA, USA; VANCHCS,  
Mather, CA, USA
- Shari Speer* Ohio State University, Columbus, OH, USA
- Michael K. Tanenhaus* Department of Brain and Cognitive Sciences, University  
of Rochester, Rochester, New York, USA
- Rosalind Thornton* Macquarie University, Sydney, Australia
- Joseph K. Torgesen* Florida Center for Reading Research, Florida State  
University, FL, USA
- John C. Trueswell* Department of Psychology, University of Pennsylvania,  
Philadelphia, PA, USA
- Roger P. G. van Gompel* Department of Psychology, University of Dundee, UK
- Cyma K. Van Petten* Department of Psychology, University of Arizona,  
Tucson, AZ, USA
- Richard K. Wagner* Department of Psychology, Florida State University, FL,  
USA
- Gloria Waters* Communication Sciences and Disorders, Sargent  
College, Boston University, MA, USA
- Gary Weismer* University of Wisconsin, Madison, WI, USA
- Melvin J. Yap* Washington University, St. Louis, MO, USA
- Thomas A. Zeffiro* Department of Psychiatry, Massachusetts General  
Hospital, Boston, MA, USA
- Rolf A. Zwaan* Department of Psychology, Florida State University,  
Tallahassee, FL, USA

## Chapter 1

### Observations on the Past and Future of Psycholinguistics

*Alan Garnham, Simon Garrod, and Anthony Sanford*

“Do you mean am I a structuralist or a Leavisite or a psycho-linguistician or a formalist or a Christian existentialist or a phenomenologist?”...“Well, I’m none of them”... “I’m a nineteenth-century liberal” (Malcolm Bradbury, *The History Man*, p. 106).

## 1. HISTORICAL PERSPECTIVES ON PSYCHOLINGUISTICS

### 1.1. Philosophical Beginnings

How and when can we distinguish “psycho-linguisticians” from other people who might be interested in language or, more generally, from “nineteenth-century liberals”? Interest in language is a long-standing one, though when it became a psycholinguistic one, is a harder question. North Americans tend to date the history of psycholinguistics from the 1950s. Being European, our natural instinct is to trace intellectual origins to Ancient Greece: to Plato, in fact, since Socrates did not write anything, and the pre-Socratics are too fragmentary and difficult to interpret. Plato had a theory of concepts. In fact he had the same theory as Jerry Fodor (1987). Plus ça change. Except that Plato tried to say something about where the “innate” concepts came from – from our (mysterious) contact with the world of ideal forms. Fodor remained silent about this matter. Perhaps he was heeding Wittgenstein’s (1921/1961) advice to keep quiet when it is patently obvious that nothing sensible can be said (or however one wants to translate proposition 7 of the *Tractatus*). We mention Plato’s theory of concepts because Plato was clearly concerned with the mental. Indeed his theory of concepts is more like Fodor’s than much of the so-called psycholinguistic research of the 1950s is like today’s work on sentence processing, for example.

Plato aside, or rather theories of “ideas” aside, much of the interest in language before the late nineteenth century was not psychologically oriented. To a modern psycholinguist, particularly one influenced by Chomsky’s view of linguistics, that may seem strange. Nevertheless, even though language use is clearly (primarily and almost entirely) a

human activity, and a mental activity at that, most people throughout most of the history of the study of language have treated language as, in Jerry Katz's (1981) phrase, an "abstract object". Ironically, perhaps, this view is called Platonist. So, there is a long history of interest in languages per se, going back, according to recent scholarship, over 2500 years. There were flourishing traditions in Mesopotamia, China, the Arabic-speaking world, ancient Greece, and, perhaps most notably, India, in the study of grammar, broadly construed. In some of these traditions, but not all, the link between the study of language and the study of logic was strong. This strand of work on language led, eventually, to the development, in the work of Boole, Frege, and others, of formal logical systems that bore certain resemblances to natural languages. And eventually, formal tools were applied to something roughly approximating to natural languages, with the first serious attempt to capture some of the complexity of real languages in the work of Richard Montague (Thomason, 1974).

Other traditions focused more closely on the details and intricacies of natural languages, leading eventually to the comparative method of William Jones and others in the nineteenth century, and then to Saussure, structuralism and modern linguistics proper. Another strand of this work, one interwoven with more general issues that can be traced back to Plato, sowed the seeds of the Chomskian revolution, or at least is retrospectively seen as doing so.

These developments began with some of Aristotle's many disputes with Plato. Aristotle did not like Plato's theory of concepts, and who can blame him? Since Aristotle we have had over two millennia of rationalism versus empiricism. The debate centered primarily on ideas (concepts) on the one hand and knowledge on the other. It became heated in so-called modern (post-Renaissance) philosophy and in particular in the works of Descartes, Spinoza, Leibniz, Kant and other continental rationalists, and Bacon, Hobbes, Locke, Berkeley, Hume and other British empiricists. In this philosophical tradition, other aspects of language received little attention, or so one might conclude from the standard histories of philosophy. However, Chomsky famously picked up on Descartes remarks about the creative nature of language, and discovered precursors of his own ideas in the work of "lesser" Cartesian philosophers, such as Cordemoy, and in the (rationalist influenced) Port Royal grammar (Arnauld & Lancelot, 1660), which is now seen as proposing the notion of a universal grammar.

## 1.2. Psychological Beginnings

Psychology did not exist as a discipline in the first half of the nineteenth century. By the end of the century it clearly did. It is traditional to identify the foundation of Wundt's lab in Leipzig as the beginning of psychology as an independent discipline. And it is certainly true that the division of faculties into departments in rich nineteenth century German universities both freed psychologists from some of their philosophical shackles, and allowed them to begin or expand programs of empirical research. Development in the early part of the nineteenth century are also pertinent to the discipline of psycholinguistics, Medicine saw spectacular changes and spectacular growth, with detailed case studies appearing of

psychological deficits of various kinds. Of particular importance to psycholinguistics were the original descriptions of Broca's (1861) and Wernicke's (1874) aphasias.

The foundation of Wundt's lab, and its notional date of 1879, is well known. Less well known that there was a thriving tradition of experimental work on the psychology of language, particularly in Wundt's own lab. Wundt himself published a book on "die Sprache" in 1900, which appeared in an enlarged two-volume edition in 1912–1913.

Wundt's early psycholinguistic work, and that of other German-speaking or German-influenced psychologists in the late nineteenth- and early twentieth centuries has been documented in detail by Arthur Blumenthal (1970). In a later piece, Blumenthal (1987) suggested a symmetry between the breakdown of the early period of Sprachpsychologie, as he called it, and the prematurely announced (Reber, 1987) demise of modern psycholinguistics. According to Blumenthal, some linguists were initially greatly attracted to the empirical and philosophical work being carried out on language in Wundt's Leipzig laboratory. Regular attendees at Wundt's lectures included Bloomfield, Mead, Saussure, and Boas. Also attracted were the so-called Junggrammatiker, many of whom were also based in Leipzig. These "young grammarians" (the common translation "neogrammarians" is misleading) were reacting against the stuffiness of Germanic university traditions in the humanities. As is well known, the different approaches and philosophies of the various German labs appeared to produce irreconcilable problems within psychology itself. And according to Blumenthal this led some linguists, notably Delbrück (1901), to argue that linguists should seek to work independently of psychologists. Reber (1987) similarly argued that one of the reasons for the demise of modern (Chomskian) psycholinguistics was that linguists could not agree among themselves, and that psychologists therefore thought they would be better working on their own.

As is well known, the First World War and its aftermath had a profoundly negative effect on psychological science in continental Europe. In North America, too, Wundt's influence waned dramatically with the advent of behaviorism. In Europe, Wundt became involved in the kind of arguments that led directly to behaviorism in the USA – about the use of introspective techniques, for example. Wundt, himself, favored strict experimentation, but the Würzburg group favored the use of introspective techniques. It was from this group that the next great European psycholinguistic, Karl Bühler, emerged. Bühler was a functionalist and, although he publicly opposed them, his ideas had much in common with the Gestalt psychologists, who also emerged from the Würzburg school. Bühler was forced to flee from the Nazi regime to the USA, but never established himself in an academic post there.

Ironically, behaviorism affected the study of language both within psychology and within linguistics, but without producing a continuing rapport between them. Bloomfield's *Introduction to the study of language* (1914) was decidedly Wundtian in orientation. But by the time it had metamorphosed into *Language* (1933), the behaviorism for which Bloomfield is renowned had come to the fore, though in the preface he states that "since that time (1914 – AG)...we have learned...that we can pursue the study



of language without reference to any one psychological doctrine". In psychology, although behaviorists were anxious to analyze thought as subvocal speech, they had comparatively little to say about speech itself, or any other aspect of language. Eventually, in the late 1950s, Skinner published his little read but much cited (as routed by Chomsky) *Verbal behavior* (1957). Behaviorists still maintain that Skinner's purpose was entirely different from what Chomsky construed it to be, and that it is a functional analysis, in a broader sense of that term than is common in linguistics (see, e.g., Catania, 1998, 2005).

Bloomfieldian linguistics made little contact with psychology, behaviorist or otherwise. But one strand of early- to mid-twentieth century North American linguistics did make an impact on the emerging cognitive psychology of the 1950s: The Sapir – Whorf hypothesis. Sapir was an academic linguist. Whorf was not. His passion and the fact that he was self-taught give his work a certain appeal. And this appeal was only compounded by stories, apocryphal or otherwise, such as the one about the "empty" gasoline cans that exploded when spent matches were discarded in them, because they were full of gasoline vapor. The question of how language and thought are related is an old and difficult one. Students like the topic and they like the Sapir – Whorf hypothesis. However, regardless of its truth or falsity, at least one of the authors of this article (AG) has always been embarrassed in presenting some of Whorf's arguments for this position.

Wundt's work was wide-ranging, and influential in its time. Yet it was rapidly forgotten. However, it should be remembered that there are other antecedents of modern psycholinguistics in late nineteenth- and early twentieth-century "psychological" scholarship. In *The psychopathology of everyday life* Freud (1975) introduced the idea that has come to be known as the Freudian slip, one type of which is the slip of the tongue. Freud had particular ideas about the genesis of speech errors, and little concern for the form they took. However, those he reported largely conform to modern notions of what speech errors should sound like (Ellis, 1980). There is, of course, no contradiction between the idea that speech errors have certain linguistic properties and the idea that they are generated as expressions of unconscious intentions. Indeed, the latter might be considered a description of the causes of speech errors as "verbal behavior", though perhaps not one that behaviorists would relish. Slightly earlier than Freud, Meringer and Meyer (1895) published what would now be seen as a more orthodox analysis of a large collection of speech errors, noted down from everyday speech. This technique was revived in the 1960s and some of its limitations noted (see, e.g., Cutler, 1982).

Corpus techniques were also being applied, at a similar time, to questions about language acquisition, again foreshadowing work that started in the 1960s in North America. As Arthur Blumenthal (1970) points out in the introduction to Chapter 3 of his *Language and psychology: Historical aspects of psycholinguistics*, diary studies of child development are often traced back to Rousseau's 1762 book *Émile*, and to a more formal report published by an academic, Dietrich Tiedemann, in 1787. With the emergence of psychological laboratories in Germany in the late 19th century, more sophisticated accounts began to be kept, including one by Wundt. The most systematic of these was that of Wilhelm Preyer (1881). More important than these works, however, is the diary

study (and its analysis) of Clara and Wilhelm Stern (Stern & Stern, 1907), which was considerably more linguistically sophisticated.

A not entirely separate set of precedents for another line of research can be traced back to a similar period: on eye-movements in reading. Émile Javal, in Paris, first observed that the eyes do not move smoothly in reading, but in a series of jerky movements (saccades) interspersed with pauses in which the eye are effectively still (fixations). J. M. Cattell (1886), working with Wundt in Leipzig, used a new instrument, the tachistoscope, to show that words can be recognized in a single glance in which the eye does not have time to move across the word. Various attempts were then made to record eye movements in reading, culminating in the work of Edmund Huey (1908). Huey had met Javal in France, and carried out much of his work in the laboratory founded by G. Stanley Hall at Clark University in Massachusetts. Hall had also travelled in Europe and studied with Wundt in Leipzig. In addition to Huey, Erdmann and Dodge, in Halle, made detailed observation of eye movements in reading, observing those movements via mirrors. Dodge also showed that little information is taken in while the eye is moving. Even in the behaviorist period, there were important studies of eye movements by Buswell and Tinker. Buswell (1922) devised a much less intrusive method (than Huey's Plaster of Paris cup on the cornea) for making a physical record of eye movements. Tinker (1936) attempted to establish that eye movements in ordinary reading were similar to those obtained with the rather intrusive laboratory techniques then in use.

### 1.3. The Modern Era

Just as psychology as a science is traditionally traced to the founding of Wundt's laboratory in Leipzig, modern (largely Anglo-Saxon) psycholinguistics has its quasi-mythical founding moment. Actually, there are three related moments. Two seminars sponsored by the Social Science Research Council (US) and the subsequent publication of the original version of Osgood and Sebeok's (1965) *Psycholinguistics: A survey of theory and research problems*. A leading figure in the instigation and organization of these seminars was John B. Carroll, editor of the collected papers of Benjamin Lee Whorf (Carroll, 1956), and a psychologist who is associated with attempts to establish the Sapir – Whorf hypothesis using psychological techniques. The leading idea was “reuniting linguistics and psychology” (Osgood & Sebeok, 1965, p.v), thus recognizing that they had previously been much closer than they were in 1950. John W. Gardner, another psychologist who played a leading role in setting up the seminars, hoped that the reunion would have profound implications for problems in education.

In retrospect, the seminars and Osgood and Sebeok's survey have an almost surreal feel to them. There is little, if any, hint of the impending impact of Chomsky's work on either linguistics itself or psycholinguistics. Of the three “approaches to language behavior” identified, one is the linguistic approach and the other two are the learning theory approach (behaviorism) and the information theory approach. Both information theory and learning theory were very quickly dismissed, in the period following the seminars, as too narrow to encompass language behavior.

While learning theory approaches never recovered from withering attacks by Chomsky (1959), Fodor (1965) and others, information theory has been important for cognitive psychology more generally. It influenced work on attention, short-term memory and, to some extent work on language – one thinks of Miller and Selfridge’s (1950) use of texts with different orders of approximation to English, and Yngve’s (1962) work on transition probabilities. One might even argue that it is (however tenuously!) linked to the notion of information processing, which was hugely important in the cognitive revolution of the 1950s and 1960s.

Miller, information theory’s most important psychological proponent, was soon lured away from that approach by the idea that a theory like that outlined in Chomsky’s *Syntactic Structures* (1957) could form the basis of a processing theory (see, e.g., Miller & Chomsky, 1963). This 1963 paper is the origin of a set of ideas that later came to be dubbed as the Derivational Theory of Complexity:

The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring an appreciation of the structure of transformed sentences is some function of the nature, number and **complexity** of the grammatical transformations involved. (Miller & Chomsky, 1963, p. 481).

Unfortunately the Derivational Theory was never formulated in a testable way, and it is unclear how it could be (Garnham, 1983).

The 1965 reprint of Osgood and Sebeok’s survey contains a follow-up Survey of Psycholinguistic Research, 1954–1964 by A. Richard Diebold Jr. (1965) Its bibliography, which runs to nearly 16 pages, is informative. There are references to work on language and thought, language acquisition, verbal learning, and information theory as represented at the original meetings. There are a good many references to work by linguists, including Chomsky, and there is mention of the early experimental work inspired by Chomsky’s linguistic theory. What is noticeable, however, is how few of these references would appear in a modern text on psycholinguistics. Psycholinguistics as we know it really got started in the mid-to late 1960s.

As we have already mentioned, the first major wave of work looked at the psychological reality of transformations and led to the Derivational Theory of Complexity. Chomsky’s ideas were also influential in empirical work on language acquisition (e.g., Roger Brown’s First Language project, see Brown, 1973, and the notion of the Logical Problem of Language Acquisition, see Baker & McCarthy, 1981), and on the biological foundations of language (e.g., Lenneberg, 1967). Chomsky, notoriously, soon backed away from the idea that experimental work in psychology might have implications for linguistic theory. And, indeed, other linguists, including contemporary cognitive linguists, who claim that cognitive considerations are important for language, have proved similarly reluctant to engage with psychological methods (as opposed to psychological considerations). Reber’s (1987) claim of a “(surprisingly rapid) fall of psycholinguistics”

is misleading in that all it is really saying is that the particularly strong link between Chomskyan theory and psychological research on language, which existed briefly in the 1960s, was broken and it was not replaced by a similar link to another framework. So, for example, neither Generative Semantics, nor any of the Phrase Structure Grammars of the 1970s and 1980s, both of which have an obvious psychological appeal, have inspired much in the way of psycholinguistic research. And neither, as we have already said, has cognitive linguistics. Reber's comments do, however, open a debate about the link between linguistics and psychology and the extent to which the psychology of language should be psycholinguistics. Can language use be explained partly or wholly in terms of general cognitive principles, or do we have special language processing devices? And if the latter, what linguistic concepts are needed to describe them? We do not have answers to these questions. We believe that proper descriptions of languages are important for the psychology of language, but the relation between linguistic descriptions and descriptions of processing mechanisms is likely to be a complex one. Nevertheless, even without definitive answers, work on sentence processing has continued apace in the 1980s and 1990s and through into the 21st century.

Another major influence on psycholinguistic research from the 1960s was work in artificial intelligence, and in particular research from Minsky's (1968) semantic information processing framework, which culminated, as far as language processing was concerned, with Terry Winograd's (1972) SHRDLU. The other major influence from the semantic information processing literature on psycholinguistics research was Ross Quillian's (1968) notion of a semantic network for representing meaning. Perhaps more generally influential, first in setting unreasonable expectations for AI, and then for the backlash against it, was Joseph Weizenbaum's (1966) ELIZA program, in its various manifestations. DOCTOR, the Rogerian therapist version of ELIZA, engaged in convincing conversations with people, and led to claims (not by Weizenbaum) that the problem of understanding language had been solved. ELIZA is widely claimed to have inspired the creation of HAL, the computer in *2001: A Space Odyssey*. Weizenbaum (1976) later made strong representations against the claims made for ELIZA.

Since the 1980s, however, the GOFAI ("good old-fashioned AI") that inspired psychological research has dried up with the advent of what has been called the "AI winter". Perhaps foolishly in retrospect, one of us published an AI textbook for psychologists at that time (Garnham, 1988). To some extent the place of GOFAI in psychology has been taken by neural network ("connectionist") modelling. Within psycholinguistics, the major impact of neural network modelling has been in the domain of word recognition. Connectionism has also sparked renewed debate about modularity.

A refreshing development in psycholinguistics since the 1960s has been a greater sophistication in dealing with questions about meaning. Although semantic networks capture some interesting facts about word meaning, the "theory" theory (Murphy & Medin, 1985) is more subtle. And more recently there has been a renewed interest in questions of polysemy, metonymy and the like. However, we remain unconvinced that most psycholinguists appreciate the enormous complexities of questions about the

meanings of words and how they relate to representations of information in individual minds (an issue touched on in Hilary Putnam's, 1975, famous discussion of natural kind terms such as "larch" and the linguistic division of labor). The use of Wittgenstein's ideas in psychology (e.g., Garnham, 1980), or at least psychologists' interpretations of those ideas, is another interesting development. However, Wittgenstein's (1953) own ideas are notoriously opaquely presented, and it is almost certain that he would not have endorsed the various uses of his ideas, given his dismissive view that "in psychology there are empirical methods and conceptual confusion" (Part II, Section xiv). Much earlier in his career, Wittgenstein had experience of psychological research. Another set of ideas that continue to intrigue psycholinguists are those of pragmaticists, and in particular the work of Grice (see, in particular, 1975). Nevertheless, there is a great deal more work to be done in determining how the pragmatic aspects of meaning are produced and understood.

The mental model theory (Johnson-Laird, 1983) has revolutionised thinking about text meaning. For example, it gives a much clearer idea of what is meant by the integration of information in comprehension than the "bizarre" Bransford and Franks (1971) experiment. Nevertheless, reading a novel is surely more about engaging with the characters than about constructing an internal model of the situation(s) described in the text. There is a great deal more to be learned about the understanding of extended texts.

There is much more that could be said about the modern era in psycholinguistics. Much of it is said in this Handbook. There is no doubt that – psycholinguistics is alive and kicking. We all have our favored questions and techniques, but there is plenty to keep us all, and more, busy for the foreseeable future. We should learn whatever we can however and from whomever.

## 2. FUTURE DIRECTIONS

The original handbook chapter on Future Directions opened with the statement that there is nothing more foolish than trying to predict the future. And that is still as true today as it was then (although it is a fair bet that we shall see much more from the laboratories of cognitive neuroscientists). Instead of trying to predict the future we pick some issues in psycholinguistics that we feel call out for future study. The issues are not in any way intended to be exclusive. They are just issues that we feel are important, unresolved and relate directly to the primary goal of psycholinguistics in elucidating psychological mechanisms of language use.

The first issue concerns the range of language use addressed by the subject. The modern era of psycholinguistics has concentrated almost exclusively on one kind of language use: namely, that associated with monologue settings. Yet, the most natural and basic form of language use is dialogue: Every language user, including young children and illiterate adults, can hold a conversation, yet reading, writing, preparing speeches and even listening to speeches are far from universal skills. Therefore, we feel that a central

goal of psycholinguistics of the future should be to provide an account of the basic processing mechanisms that are employed during natural dialogue.

The second issue we consider also concerns scope, but in this case in relation to the more conventional topic of reading comprehension, and the extent to which standard psycholinguistic approaches do justice to the complexity of texts that people read in everyday life. In both cases the issue is about taking seriously how language processing in the psycholinguistic sense relates to the wide range of uses to which language may be put.

## 2.1. Language Processing and Dialogue

There are many reasons why psycholinguists have avoided dialogue in the past, both theoretical and practical. The theoretical reason relates to something which we have already alluded to in covering the history of the subject and its grounding in linguistics. Theoretical linguistics, at least in the generative tradition, has developed theories about the structure of isolated, decontextualized sentences that are used in texts or speeches – in other words, in monologue. In contrast, dialogue is inherently interactive and contextualized: Each interlocutor both speaks and comprehends during the course of the interaction; each interrupts both others and himself; on occasion two or more speakers collaborate in producing the same sentence (Coates, 1990). So it is not surprising that generative linguists commonly view dialogue as being of marginal grammaticality, contaminated by theoretically uninteresting complexities. Dialogue sits ill with the competence/performance distinction assumed by most generative linguistics (Chomsky, 1965), because it is hard to determine whether a particular utterance is “well-formed” or not (or even whether that notion is relevant to dialogue). Thus, linguistics has tended to concentrate on developing generative grammars and related theories for isolated sentences; and psycholinguistics has tended to develop processing theories that draw upon the rules and representations assumed by generative linguistics. However, the situation in linguistics is changing and linguists are beginning to explicitly take dialogue into account (see e.g., Ginzburg & Sag, 2001; Keysar, this volume). So there is less theoretical excuse for psycholinguists to ignore dialogue.

The practical reason is that dialogue is generally assumed to be too hard or impossible to study, given the degree of experimental control necessary. Until quite recently it was also assumed that imposing a sufficient level of control in many language production studies was impossible. Thus, Bock (1996) points to the problem of “exuberant responding” – how can the experimenter stop subjects saying whatever they want? However, it is now regarded as perfectly possible to control presentation so that people produce the appropriate responses on a high proportion of trials, even in sentence production (e.g., Bock, 1986; Levelt & Maassen, 1981)

Contrary to many people’s intuitions, the same is true of dialogue. For instance, Branigan, Pickering, and Cleland (2000) showed effects of the priming of syntactic structure during language production in dialogue that were exactly comparable to the priming shown in isolated sentence production (Bock, 1986) or sentence recall (Potter & Lombardi, 1998). Similar control is exercised in studies by Clark and colleagues (e.g.,

Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992; also Brennan & Schober, 2001; Horton & Keysar, 1996). Well-controlled studies of language processing in dialogue may require some ingenuity, but such experimental ingenuity has always been a strength of psycholinguistics.

Also, there has been a steady development in techniques that make it much easier to study language processing 'in the wild' that could contribute to a mechanistic psycholinguistics of dialogue. First, there is the development of more extensive dialogue corpora available in electronic form. Some corpora have been elicited in semi-controlled conditions, such as the HCRC map task corpus (Anderson et al., 1991), and have been extensively coded and time-stamped. Such rich sources of naturalistic data open up new ways of testing processing hypotheses. There is also the development of more sophisticated behavioral measures of on-line processing during dialogue, such as the head-mounted or remote eye-tracking systems now available. In fact, such equipment has already been used to investigate referential processing in a constrained dialogue setting (see Brown-Schmidt, Campana, & Tanenhaus, 2004; Tanenhaus & Trueswell, this volume).

Therefore from both the theoretical and the practical point of view there is every reason to hope for a more focused study of language processing during dialogue in future years. But how might this contribute to a better understanding of basic language processing mechanisms? The key difference between a dialogical and a monological approach to language processing is in how they define the system under investigation. In a monological approach there are two basic systems one for language production the other for language comprehension. The only relation between the two is that the output of one system is taken as the input to the other. In other respects the two systems are to all intents and purposes independent. However, the dialogical approach treats the system as minimally bounded by two interlocutors engaged in both production and comprehension of the language being used. Communication and language processing is taken to be a joint activity between both interlocutors (Clark, 1996). Hence, how one interlocutor formulates her message is inevitably influenced by how the other interlocutor has formulated his. More generally, Pickering and Garrod (2004; see also Garrod, 1999) have argued that the basis of successful communication is somewhat different in monologue and dialogue settings. In monologue readers and listeners attempt to establish a coherent interpretation or situational model of what the texts are about. However, successful dialogue depends on interlocutors aligning their respective models or representations.

Pickering and Garrod (2004) argue that one of the consequences of this alignment process is that comprehension and production become coupled. This is not a new idea, at least in some areas of psycholinguistics. For instance, it has long been argued that there is a close relationship between perception and articulation of speech (Liberman & Whalen, 2000). Although the claim about speech is still controversial, the debate has recently been rejuvenated in neuroscience as a result of evidence for activation of articulators during speech perception (Fadiga, Craighero, Buccino, & Rizzolati, 2002).

## 2.2. An Enriched Approach to Reading Comprehension

Just as dialogue poses a challenge for future psycholinguistics, we think that even continued research into monologue and reading settings has its own challenges that require considerable attention. A prevalent perception of psycholinguistics by many academics outside of the discipline, but interested in language use, is that the materials used in most experiments are short, typically dull, de-contextualised, and generally unrelated to anything in real life. As psycholinguists, we would of course defend the subject against any negative construal of these facts, pointing out that adequate control over sentence structure and content is essential if we are to understand the basic mechanisms of comprehension. It is of central interest, for instance, to determine the syntactic and semantic interpretations given particular sentence structures.

However, we also think that it is important for psycholinguistics to take a broader view of interpretation. While the determination of the processes and principles underlying how single, controlled sentences are understood, along with analyses of the time course of processing, is grist to the mill of psycholinguistics, it can harbour dangers for progress. For instance, there has been a small but steady amount of work on the interpretation of doubly quantified sentences, almost all of which (to our knowledge) relies on principles that operate at sentence level only, so that the sentences of interest occur *in vacuo*. Although some sentence-level grammatical constraints undoubtedly apply in these situations, we think it likely that most examples of double quantification occur in specific contexts, where pragmatics will constrain ultimate interpretations, and will probably constrain intermediate possibilities, and semantic interpretations, as well. Studies of *in situ* processing may well provide a different and more useful picture of how interpretation occurs from studies of sentences *in vacuo*. Let us immediately say that we are not advocating lack of experimental control: rather, we are advocating the study of processing within situations where sentences are treated as utterances within a setting. While hopefully there will be many instances in which processing is identical whether the sentence in question is *in vacuo* or *in situ*, it is a question that needs an answer for any specific proposed processing mechanism.

Another aspect of greater realism in materials concerns claims of richness in interpreting these materials. The kind of inferential activity that goes along with reading has been a topic of interest in the psychology of language for many years, and has had a checkered history. Inferences made in the service of coherence (causal chaining, discourse anaphora, etc.) probably form the biggest group that have been investigated, although there has been a steady interest in elaborative inference as well. Work on causal chaining and on elaborative inferences has tended to be the domain of discourse psychologists, who are somewhat on the edge of what is currently mainstream psycholinguistics. McKoon and Ratcliff's (1991) paper on minimal inference-making while reading occurred at a time when (since the mid-1970s) psychologists along with colleagues in AI were hypothesizing very large scale inferential activity in the service of comprehending. McKoon and Ratcliff showed how careful one has to be in claiming that



this or that inference is made. Beyond that, they claimed that in many laboratory reading studies, many types of inference simply are not made.

This minimalist position holds little appeal for the growing number of psychologists interested in what we might term the richness of experience when immersed in a story-world. For instance, adopting what on the face of it is a quite different viewpoint, Zwaan (1999) suggested:

When reading a story, we may “experience” cold wind blowing in our face, the smell of stale beer, a kiss on our lips... (Zwaan, 1999, p. 83).

Certainly, if readers experienced no excitement, no prediction of what might happen next, and no ego and emotional involvement, then there would be no sales of novels, detective fiction, and popular magazines. Expanding the notion of interpretation to include these possibilities takes one well beyond the limits of what is typical in mainstream psycholinguistics. It is of interest to note evidence that descriptions do indeed have very direct influences on people, and these go beyond the purely cognitive. For instance, listening to accounts of after dinner smoking elicits cognitive and physiological responses consistent with smoking urges in smokers (e.g., Drobles & Tiffany, 1997). Writing is just as often emotive and dramatic as it is meant to be instructional.

We suggest that a better understanding of how written communication works is possible only by exploring a fuller range of written material, let us call it realistic written material. Once again, we should emphasize that we are not advocating a lack of controlled experimental materials, but rather, an expansion of the types of questions that are being asked. One growing area that should rather obviously benefit from a broader perspective is what might be termed the grounding program.

In the past two or three decades there has been an upsurge of interest in the problem of how meaning is grounded in the world of perception and action (e.g., Searle, 1980; Harnad, 1990; Glenberg, 1997; Barsalou, 1999; Ziemke & Sharkey, 2001). Consideration of the symbol grounding problem is of course having an impact on questions about language comprehension and is a likely area for a major upsurge of future research. Most of the existing work is on the periphery of mainstream psycholinguistics at present, and only the most basic of demonstration work has so far been carried out. First, pioneering efforts by Glenberg and his colleagues, and the later work of others, have demonstrated that when actions are carried out that are incompatible with the direction of movement implied by a description, then interference occurs. For instance, Glenberg and Kaschak (2002) had people judge whether utterances like *You handed Courtney the notebook* was a sentence or not. If the judgement was made by moving the hand to a button away from the body, then it was initiated more rapidly than if the judgement was made by moving toward the body. The opposite holds for *Courtney handed you the notebook*. More recently, Zwaan and his colleagues (Zwaan, Madden, Yaxley, & Aveyrard, 2004; Kaschack et al., 2005) have demonstrated similar interference/facilitation patterns for perceptual displays changing in ways that represent motion away from or toward the observer.

Secondly, studies in neuroscience have also been recruited as evidence of embodied cognition. There is now good evidence that areas of the brain near the appropriate motor cortex areas are activated when words denoting certain bodily actions are presented in a variety of tasks. For instance, the verb *walk* activates areas near those associated with movements of the lower limbs, while *talk* activates areas associated with control of verbal articulation (Pulvermuller, Harle, & Hummel, 2001). This work is seen as supporting the belief that understanding is somehow rooted in experience (action and perception).

While this angle constitutes an interesting way of approaching at least some of the fundamental problems of meaningful, grounded interpretation, it also appears to provide a basis for linking reading to the kinds of phenomenological experiences described by Zwaan. We think that a broader construal of the notion of interpretation to include bodily correlates is a likely area of expansion. However, demonstrations of the role of action and perceptual systems in interpretation is certainly in its infancy, and, like mainstream psycholinguistics, uses single sentences in most experiments. Whether comprehending long sequences that describe complex actions requires close coupling to action and perceptual systems throughout the narrative still requires demonstration.

Once one begins to address the possibility of dealing with naturalistic materials like real stories and literary materials, many other questions arise. One is simply whether processing is uniformly full (“complete”, if you have a theory of what constitutes a complete interpretation). There is a growing literature within more conventional psycholinguistics showing that semantic processing is often rather shallow (see Ferreira, Bailey, & Ferraro, 2002; Sanford & Sturt, 2002, for reviews). When one considers reading long pieces of work, selective processing is likely to be much more important than in the typical single sentence/short paragraph work of traditional psycholinguistics, although this is an open question, of course. However, much more study of what it is that controls how thoroughly processes of reference resolution, causal chaining, and representation of discourse takes place, seems to be due. Writers typically have the problem of causing their readers to think about A and not about B – that is, they have a problem of controlling the processing patterns of their readers, if they are to effectively put over an impression or a message. Selectivity in processing must underlie success in the face of this problem.

### 2.3. Scope and Interest

We see a major problem for a proper psychology of language as being one of lack of interaction between different sub-disciplines. For instance, the lack of overlap of attendance at the major conferences on sentence processing (e.g., the CUNY series of conferences) and those on discourse (e.g., the series on Text and Discourse) is very noticeable to those of us interested in both perspectives. While sentence processing and text processing and dialogue fail to fall under integrating umbrellas, there will never be a Language Science comparable to the recently emerged Vision Science. What we cannot tell about the future is whether there ever will be a Language Science (or even an integrated Psychology of Language). But it makes a fine goal.

## REFERENCES

- Anderson, A. H., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech, 34*, 351–366.
- Arnauld, A., & Lancelot, C. (1660). *Grammaire générale et raisonnée*. Paris: Port-Royal.
- Baker, C. L., & McCarthy, J. J. (Eds). (1981). *The logical problem of language acquisition*. Cambridge, MA: MIT Press.
- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.
- Bloomfield, L. (1914). *Introduction to the study of language*. New York: Henry Holt and Company.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Blumenthal, A. L. (1970). *Language and psychology: Historical aspects of psycholinguistics*. New York: Wiley.
- Blumenthal, A. L. (1987). The emergence of psycholinguistics. *Synthese, 72*, 313–323.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.
- Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*(4), 395–421.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition, 75*(2), B13–B25.
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2*, 331–350.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 1482–1493.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for dysfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274–296.
- Broca, P. (1861). Remarques sur le siege de la faculté du langage articulé, suivés d’une observation d’aphémie. *Bulletin de la Société Anatomique de Paris, 36*, 330–357.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2004). Real-time reference resolution by naive participants during a task-based unscripted conversation. In: J. C. Trueswell, & M. K. Tanenhaus (Eds), *Processing world-situated language: Bridging the language as product and language as action traditions* (pp. 153–172). Cambridge, MA: MIT Press.

- Buswell, G. T. (1922). *Fundamental reading habits: A study of their development*. Chicago, IL: University of Chicago Press.
- Carroll, J. B. (Ed.) (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.
- Catania, A. C. (1998). The taxonomy of verbal behavior. In: K. A. Lattal, & M. Perone (Eds), *Handbook of research methods in human operant behavior* (pp. 405–433). New York: Plenum.
- Catania, A. C. Introduction to verbal behavior. Retrieved June 27, 2005 from Cambridge Centre for Behavioral Studies, Verbal Behavior. <[http://www.behavior.org/vb/index.cfm?page=http%3A%2F%2Fwww.behavior.org/vb/verbal\\_behavior\\_catania.cfm](http://www.behavior.org/vb/index.cfm?page=http%3A%2F%2Fwww.behavior.org/vb/verbal_behavior_catania.cfm)>
- Cattell, J. (1886). The time it takes to see and name objects. *Mind*, 11, 63–65.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1959). A Review of B. F. Skinner's *Verbal Behavior*. *Language*, 35, 26–58.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Coates, J. (1990). Modal meaning: The semantic pragmatic interface. *Journal of Semantics*, 7(1), 53–64.
- Cutler, A. (1982). The reliability of speech error data. In: A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7–28). Amsterdam: Mouton.
- Delbrück, B. (1901). *Grundfragen der Sprachforschung; mit rücksicht auf W. Wundt's Sprachpsychologie*. Strassburg: Traber.
- Diebold, A. R., Jr. (1965). A survey of psycholinguistic research, 1954–1964. In: C.E. Osgood, & T. A. Sebeok (Eds), *Psycholinguistics: A survey of theory and research problems* (pp. 205–291). Bloomington, IN: Indiana University Press.
- Drobes, D. J., & Tiffany, S. T. (1997). Induction of smoking urge through imaginal and in vivo procedures: Physiological and self-report manifestations. *Journal of Abnormal Psychology*, 106, 15–25.
- Ellis, A. W. (1980). On the Freudian theory of speech errors. In: V. A. Fromkin (Ed.), *Errors in linguistic performance "Slips of the tongue, ear, pen, and hand"* (pp. 123–131). New York: Academic Press.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolati, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15, 399–402.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Fodor, J. A. (1965). Could meaning be an  $R_m$ ? *Journal of Verbal Learning and Verbal Behavior*, 4, 72–81.

- Fodor, J. A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Freud, S. (1975). *The psychopathology of everyday life* (Trans. A. Tyson). Harmondsworth, UK: Penguin (originally published 1901).
- Garnham, A. (1980). Default values, criteria and constructivism. *Cognitive Science*, 4, 427–433.
- Garnham, A. (1983). Why psycholinguists don't care about DTC: A reply to Berwick and Weinberg. *Cognition*, 15, 263–269.
- Garnham, A. (1988). *Artificial intelligence: An introduction*. London: Routledge and Kegan Paul.
- Garrod, S. (1999). The challenge of dialogue for theories of language processing. In: S. Garrod, & M. Pickering (Eds), *Language processing* (pp. 389–415). Hove: Psychology Press.
- Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.
- Glenberg, A. M., & Kashak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9, 558–565.
- Ginzburg, J., & Sag, I. A. (2001). *Interrogative investigations*. Stanford, CA: CSLI.
- Grice, H. P. (1975). Logic and conversation. In: P. Cole, & J. Morgan (Eds), *Syntax and semantics III: Speech acts* (pp. 41–58). New York: Academic Press.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335–346.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Huey, E. B. (1908). *The psychology and pedagogy of reading*. New York: Macmillan.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.
- Kaschak, M. P., Madden, C. J., Theirrault, D. J., Yaxley, R. H., Aveyrard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, 94, B79–B89.
- Katz, J. J. (1981). *Language and other abstract objects*. Totowa, NJ: Rowman and Littlefield.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In: W. Klein, & W. J. M. Levelt (Eds), *Crossing the boundaries in linguistics. Studies presented to Manfred Bierwisch*. Dordrecht: D. Riedel.
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Science*, 4(5), 187–196.

- McKoon, G., & Ratcliff, R. (1991). Inferences during reading. *Psychological Review*, 99, 440–466.
- Meringer, R., & Meyer, K. (1895). *Versprechen und Verlesen; Eine Psychologische-Linguistische Studie*. Stuttgart: Göschen.
- Miller, G. A., & Chomsky, B. (1963) Finitary models of language users. In: R. D. Luce, R. R. Bush, & E. Galanter (Eds), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Wiley.
- Miller, G. A., & Selfridge, J. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology*, 63, 178–185.
- Minsky, M. (Ed.) (1968). *Semantic information processing*. Cambridge, MA: MIT Press.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Osgood, C. E., & Sebeok, T. A. (1965). *Psycholinguistics: A survey of theory and research problems*. Bloomington, IN: Indiana University Press (originally published as supplements to *International Journal of American Linguistics* and *Journal of Abnormal and Social Psychology*, 1954).
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–225.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in Immediate recall of sentences. *Journal of Memory and Language*, 38, 265–282.
- Preyer, W. (1881). *Die Seele Des Kindes (The mind of the child)*. Leipzig: Th. Grieben's Verlag (L. Fernau).
- Pulvermuller, F., Harle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and electrophysiological correlates of action verb processing. *Brain and Language*, 78, 143–168.
- Putnam, H. (1975). The meaning of “meaning”. In: K. Gunderson (Ed.), *Language, mind and knowledge*, Minnesota Studies in the Philosophy of Science (Vol. VII, pp. 131–193). Minneapolis: University of Minnesota Press.
- Quillian, M. R. (1968). Semantic memory. In: M. Minsky (Ed.), *Semantic information processing* (pp. 000-000). Cambridge, MA; MIT Press, 227–270.
- Reber, A. S. (1987). The rise and (surprisingly rapid) fall of psycholinguistics. *Synthese*, 72, 325–339.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382–386.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

- Stern, C., & Stern, W. (1907). *Die Kindersprache* (Child Language). Leipzig: Barth.
- Thomason, R. H. (Ed.) (1974). *Formal philosophy: Selected paper of Richard Montague*. New Haven, CT: Yale University Press.
- Tinker, M. A. (1936). Reliability and validity of eye-movement measures of reading. *Journal of Experimental Psychology*, *19*, 732–746.
- Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Communication of the Association for Computing Machinery*, *9*, 36–45.
- Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco: W. H. Freeman.
- Wernicke, C. (1874). *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis*. Breslau: Cohn and Weigert.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*, 183–194.
- Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, *3*, 1–191.
- Wittgenstein, L. (1921). *Tractatus logico-philosophicus* (Trans. D. F. Pears & B. F. McGuinness, 1961). London: Routledge & Kegan Paul.
- Wittgenstein, L. (1953). *Philosophical investigations* (Trans. G. E. M. Anscombe). Oxford: Basil Blackwell.
- Yngve, V. H. (1962). Computer programs for translation. *Scientific American*, *206*(6), 68–76.
- Ziemke, T., & Sharkey, N. E. (2001). A stroll through the world of robots and animals: Applying Jakob von Uexkuell's theory of meaning to adaptive robots and artificial life. *Semiotica*, *134* (1–4), 701–746.
- Zwaan, R. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88.
- Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyrard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*, 611–619.

## **SECTION 1**

### **LANGUAGE PRODUCTION**



This page intentionally left blank

## Chapter 2

### Properties of Spoken Language Production

*Zenzi M. Griffin and Victor S. Ferreira*

Although a common caricature of speaking is that it is the reverse of listening, language production processes fundamentally differ from comprehension processes in many respects. Whereas people typically recognize the words in their native language quickly and automatically, the same words require an intention to speak and can take over five times longer to generate than to recognize. For example, listeners begin to direct their gaze to the referent of a spoken noun (even in the absence of highly predictable speech) before the speaker completes articulation of the word (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), whereas speakers typically take about 900 ms to begin to generate a noun in isolation based on a pictured object (e.g., Snodgrass & Yuditsky, 1996).

Language production is logically divided into three major steps: deciding what to express (*conceptualization*), determining how to express it (*formulation*), and expressing it (*articulation*; Levelt, 1989). Although achieving goals in conversation, structuring narratives, and modulating the ebb and flow of dialogue are inherently important to understanding how people speak (for review, see Clark, 1996), psycholinguistic studies of language production have primarily focused on the formulation of single, isolated utterances. An utterance consists of one or more words, spoken together under a single intonational contour or expressing a single idea (e.g., Boomer, 1978; Ferreira, 1993). While Ferreira and Englehart's chapter in this volume on syntax describes processes that allow speakers to produce their words in grammatical utterances, we focus instead on the processing of the words themselves. Indeed, most theories of multi-word utterance or sentence production ultimately boil down to an account of how sentences acquire their word orders and structures, how the dependencies between words are accommodated (e.g., subject-verb agreement), and a functionally independent account of how individual content words are generated (e.g., Chang, Dell, Bock, & Griffin, 2000; Ferreira, 2000; Kempen & Hoenkamp, 1987). In this chapter, we describe the basic properties of spoken word production, outlining empirical data that demonstrate the properties of the processes resulting in speech and discussing the processing assumptions that models of language production invoke to account for these properties. Although it could easily fill a chapter of its own, we conclude by discussing timing in multi-word utterances.

### 1. GENERATING WORDS

The simplest meaningful utterance consists of a single word. Generating a word begins with specifying its semantic and pragmatic properties. That is, a speaker decides upon an intention or some content to express (e.g., a desired outcome or an observation) and encodes the situational constraints on how the content may be expressed (e.g., polite or informal speech, monolingual or mixing languages; see Levelt, 1989). This process, termed *conceptualization* or *message planning*, is traditionally considered pre-linguistic and language neutral (Garrett, 1975; Levelt, 1989). However, speakers may include different information in their messages when preparing to speak different languages (see Slobin, 1996, on *thinking for speaking*). The next major stage is formulation, which in turn is divided into a word selection stage and a sound processing stage (Fromkin, 1971; Garrett, 1975). Deciding which word to use involves selecting a word in one’s vocabulary based on its correspondence to semantic and pragmatic specifications. The relevant word representation is often called a *lemma* (Kempen & Huijbers, 1983), *lexical entry*, *lexical representation*, or simply a *word*, and it marks the presence of a word in a speaker’s vocabulary that is capable of expressing particular semantic and pragmatic content within a particular syntactic context. Sound processing, in contrast, involves constructing the phonological form of a selected word by retrieving its individual sounds and organizing them into stressed and unstressed syllables (*phonological encoding*) and then specifying the motor programs to realize those syllables (*phonetic encoding*). The final process is articulation, that is, the execution of motor programs to pronounce the sounds of a word. This sequence of stages is illustrated in Figure 1.

This gross analysis of language production serves to illustrate the complexity of expressing an idea in words. The challenges posed by this complex problem give rise

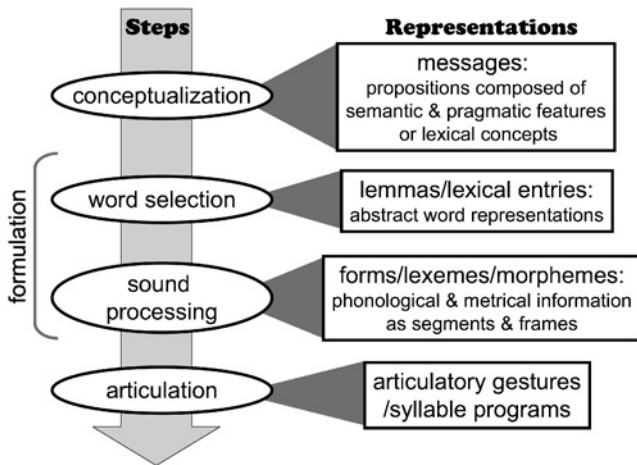


Figure 1. Major steps and representations in language production.

to the fundamental properties of the word production process, the descriptions of which below form the bulk of this chapter. Roughly, these properties delineate steps in processing (Property 1), describe how speakers deal with the relationship between meaning and word (Properties 2–9), explain how speakers represent and assemble the sounds of words (Properties 10–13), and how these processes play out in time (Properties 14 and 15).

Interestingly, current models of word production agree on the basic facts about how the system works to a surprising extent, with only minor variations in explanatory mechanisms. When models differ, the tendency is to focus on different stages of production, such as word selection or phonological encoding, and different aspects of these stages such as speed of processing or how processing may go awry to yield speech errors. This means that of the properties of production described below, most are accounted for (at least to some level of detail) by most models of production. Other properties of production have yet to receive detailed attention, although we feel that much can be gained if these properties are addressed in future theories. Next, we describe each of these properties of word production in turn.

## 1.1. Basic Steps of Word Production

### 1.1.1. *Property 1: Word selection precedes sound assembly*

When speakers access word representations, they do so first based on meaning and then focus on assembling their sounds. Several sorts of evidence suggest this. The first and strongest evidence comes from analyses of errors made during spontaneous speech (Fromkin, 1971; Garrett, 1975), which reveal that speech errors most frequently involve units that can be most conservatively considered to correspond to whole words, morphemes (i.e., minimal units of meaning such as *cran* and *berry* in *cranberry*), or individual speech sounds (i.e., phonemes or segments such as the *b-* and *oo-*sounds in *boo*). In particular, error patterns suggest that a speaker may err in selecting a word but correctly assemble and pronounce its component sounds, or they may successfully select a word that can express an intended meaning, but then err in assembling its sounds. Table 1 lists examples of word, morpheme, and sound errors. In addition, the word production process occasionally falters at a point where speakers seem to have selected a word to express what they want to say but have not yet retrieved all of its sounds (see Property 8 for details).

Another sort of evidence that is often cited as showing that meaning-based word selection precedes the processing of words' sounds comes from experiments exploring the time course of word production (Schriefers, Meyer, & Levelt, 1990). Specifically, when speakers name pictures as they ignore distractor words, semantically related distractors (e.g., reading *tiger* when naming a picture of a lion) primarily slow object-naming latencies when the distractor word appears simultaneously with the object or precedes it by up to 400 ms (Glaser & Döngelhoff, 1984; Roelofs, 1992; Starreveld & La Heij, 1995, 1996; note that the effects of semantically and phonologically related words on word production

Table 1

Speech error examples with **error** in bold, *correction* in italics, and [intended words] in square brackets.

Error type	Transcription	Source
Semantically related word substitution	“and Robbie delivers the blow that might give him the <b>yel-</b> er the <b>pink</b> oh dear let’s start again might give him the <i>green</i> jersey”	Phil Liggett, Outdoor Life Network commentator, 47 minutes into stage 13 of the 2005 Tour de France, July 15, 2005
Blend	“ <b>Justi:ce</b> , <i>Justin and Travis</i> ”	Dr. Arthur D. Fisk in Cognitive Aging Brown Bag December 4, 2001
Phonologically related word substitution	“The battle for the green <b>journey</b> [jersey] currently on ice, until the next sprint perhaps”	Phil Liggett, Outdoor Life Network commentator, 1st sprint on Stage 21 of the 2005 Tour de France, July 24, 2005
Morpheme substitution	“It’s much more useful on occasions like that. It gives the rider and the coach an awful lot of <b>useless</b> information. <i>useFUL information</i> ”	Paul Sherwen, Outdoor Life Network commentator, 1:45 into Stage 19 of the 2005 Tour de France, July 22, 2005
Sound substitution	“The <b>raid</b> – uh . the <i>roadie</i> is accepting uh . guitar from a young man”	Participant in an experiment, Griffin (2004)

are the reverse of what is typically observed for primes in word recognition – for review, see Balota, Cortese, & Yap, this volume). When a semantically related word is presented after the object appears, it either has no impact on latencies or results in faster latencies relative to an unrelated control word. In word translation (which takes a bit longer than object naming), a semantically related distractor word speeds production when it precedes the to-be-translated word by 400 ms and slows it when presented 200 ms after the to-be-translated word (Bloem & La Heij, 2003). In contrast, phonologically related distractor words (e.g., hearing *liar* when naming a picture of a lion) sometimes have no effect on object naming latencies when they precede the presentation of the objects, but consistently facilitate naming latencies when they appear at the same time or after the objects (Damian & Martin, 1999; Schriefers et al., 1990). That said, phonologically related distractors presented as early as 300 ms before objects may facilitate naming (Starreveld, 2000). Interpretation of effects in the picture–word interference paradigm relies heavily on uncertain assumptions about word recognition and how it interacts with word

production (cf. Levelt, Roelofs, & Meyer, 1999; Roelofs, Meyer, & Levelt, 1996; Starreveld, 2000; Starreveld & La Heij, 1995).

That speakers produce words first by processing their meaning-level properties, then by processing their sound-level properties is arguably the fundamental property of word production. Models of word production incorporate this property by assuming two stages in producing words as well as separate word-level and sound-level representations (e.g., Caramazza, 1997; Dell, 1986; Levelt et al., 1999; but see Starreveld & La Heij, 1996, for an exception). Meaning-level representations make lexical-level representations available, which in turn provide access to individual sound/segment representations. This implies that it is not possible to go from meaning to sounds except through a lexical representation.

## 1.2. Selecting A Content Word

Despite the apparent simplicity of the resulting utterance, production of a single word can go awry in a number of ways and can take a surprisingly long time under some circumstances. Studies of isolated word production have focused primarily on nouns (e.g., person, place, or thing) with some studies of verbs (i.e., action words and predicates), ignoring other grammatical classes of content words that are less often spoken alone. In one-word utterances, the properties of word production processes appear similar for nouns and verbs (e.g., MacKay, Connor, Albert, & Obler, 2002; Vigliocco, Vinson, Damian, & Levelt, 2002). There is no reason to suspect that other types of content words are prepared differently in single word production.

### 1.2.1. Property 2: The intention to produce a word activates a family of meaning-related words

Speech error analyses suggest that the most common error in word selection occurs when a speaker substitutes a *semantically related* word for the intended one, such as calling a van *bus* (Dell et al., 1997). A related type of speech error is a *blend* in which two words that could sensibly fill a particular slot in an utterance are spliced together to form an unintended string of sounds, such as *behavior* and *deportment* emerging as *behortment* (Wells, 1951/1973). Often the words that form a blend are not true synonyms out of context but are interchangeable only within the context of the utterance (Garrett, 1975).

Such observations suggest that accessing word representations by meaning or message representations is not surgical. Specifically, the intent to produce a particular word will lead to the activation of a family of words, all sharing some aspect of the intended meaning. This leads to the issue of how meaning is represented in models of word production, and on this issue, two major theoretical positions have been staked. On one side are *decompositional* views of semantic representation (Bierwisch & Schreuder, 1992; Katz & Fodor, 1963). Decompositional views portray the primitives of semantic representation as being entities that are smaller than the words whose production they ultimately support.

For ease of exposition, such features are sometimes described as themselves meaningful, so that the meaning of *bird*<sup>1</sup> might include *HAS WINGS*, *HAS FEATHERS*, *SINGS SONGS*, and the like (e.g., Cree & McRae, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004). However, decompositional approaches can equally (and perhaps more realistically) assume that lexicalizable concepts consist of organized collections of arbitrary features (bearing a non-trivial resemblance to parallel-distributed processing accounts of cognition; McClelland & Rumelhart, 1986; Rogers & McClelland, 2004) rather than semantic primitives (or other tidy features). By either account, semantic similarity arises straightforwardly from feature overlap – lexical items are similar to one another to the extent that the semantic features that promote their use are the same. In turn, the activation of a family of words, all related in meaning, also follows straightforwardly – if to produce *bird*, a speaker must activate the features *HAS WINGS*, *HAS FEATHERS*, and *SINGS SONGS*, then other words that also share those features will become activated (e.g., airplanes, because they have wings, opera singers, because they sing songs), and to an extent that increases with the number of shared features. Furthermore, the observation that many meaning-level influences on lexical production are context-dependent (e.g., semantic word substitutions and blends typically only involve words that can be interchanged within a particular context; Garrett, 1975; Levelt, 1989) follows from the natural assumption that representations of meaning too are activated in context-dependent fashion. Only words of the same level of specificity interfere with one another in the picture–word interference tasks, which also suggests that specificity is a important feature or constraint on word activation (Costa, Mahon, Savova, & Caramazza, 2003; Vitkovitch & Tyrrell, 1999). Theories of production that posit decompositional semantic features include Osgood (1963), Fromkin (1971), Dell (1986), Stemberger (1985), Butterworth (1989), Caramazza (1997), and Chang et al. (2000). In addition, decompositional theories have played an important role in the development of connectionist models of word processing (e.g., Strain, Patterson, & Seidenberg, 1995) and language deficits (e.g., Hinton & Shallice, 1991).

On the other side are *non-decompositional* views of semantic representation. The philosophical case for non-decompositional views has been forwarded most prominently by Fodor (1975). With respect to word production, the *WEAVER++* model (Levelt, 1989, 1992; Roelofs, 1993) and other models (e.g., Bloem & La Heij, 2003; Garrett, 1982; Starreveld & La Heij, 1996) have adopted non-decompositional representations. According to non-decompositional views, the representational bases of words and their meanings bear a one-to-one relationship, so that the word *bird* is fed by an atomic meaning representation of *BIRD*, the word *airplane* is fed by an atomic meaning representation for *AIRPLANE*, and so forth. These atomic meaning representations are often called *lexical concepts*. Within such accounts, the activation of a family of words, all similar in meaning, is not quite as straightforward as it is with decompositional accounts. Specific

---

<sup>1</sup> Words referring to words or lexical entries are printed in lowercase italics, whereas words referring to semantic representations appear in uppercase italics. The meaning of a word is a semantic representation although word is written in lowercase italics in such a context.

claims as to how multiple meanings become activated have been presented by Roelofs (1992) and Levelt et al. (1999). The idea is that activating the concept *BIRD* activates the concept *FISH* because *BIRD* will be connected within a semantic network to the concept *ANIMAL* (through what is sometimes called an “is-a” link), which will then spread activation to *FISH*. The concept *FISH* can then activate the word *fish*. Other links through other mediating properties will cause other meaning-similar words to become active. Figure 2 illustrates these two different forms of semantic representations.

1.2.2. Property 3: Words that express similar meanings compete for selection

Both error patterns and studies of object naming indicate that when a speaker strives to generate the most appropriate word for a particular occasion (often referred to as *target* or *intended* words), other words with very similar meanings in the given context become available and compete for selection as well. Intruding words in semantically related word substitutions share grammatical class, taxonomic category, and level of specificity with intended words (e.g., Fromkin, 1971; Nooteboom, 1973). Also, as described above, speakers take longer to label objects, actions, or colors in the presence of semantically related distractors relative to unrelated distractors. Although associated words such as *dog* and *bone* are related in meaning and often co-occur in speech, they do not show any

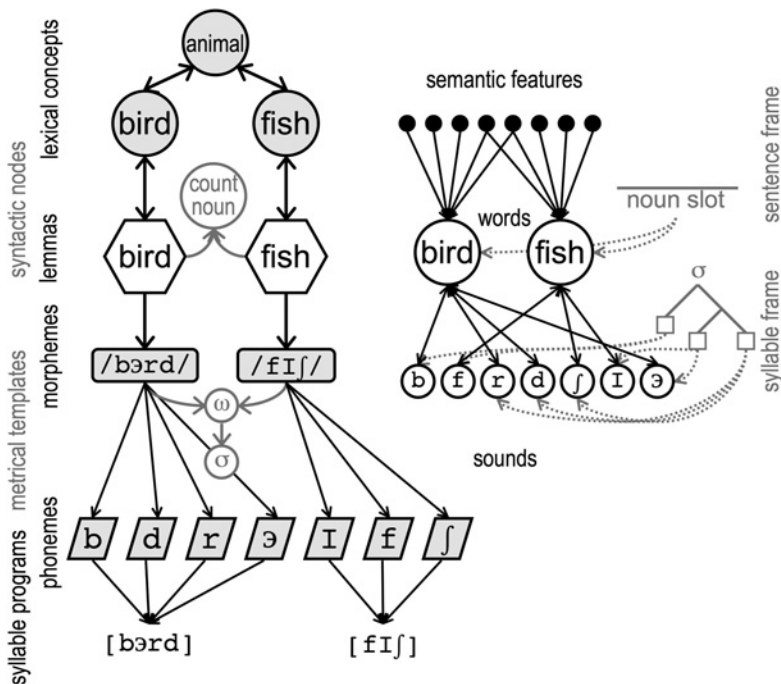


Figure 2. Diagrams based on WEAVAR++ model (Levelt et al., 1999) on the left and models by Dell (e.g., Dell et al., 1997) on the right.



tendency to compete for selection (e.g., Cutting & Ferreira, 1999; Lupker, 1979). That is, competition is restricted to words that express similar meanings rather than simply related meanings.

According to nearly all models of word production, the availability or activation level of one word affects the speed and/or likelihood that a speaker will select another word. The simplest way of modeling this is to have the probability of selecting a word (or other unit) directly related to its level of activation, so that if an unintended word has too high a level of activation, it will be erroneously selected. Due to patterns of connectivity between semantically (and phonologically) related words, these are the most likely to be highly activated and selected in error. Several models use activation levels and connectivity alone to account for patterns of speech errors (e.g., Dell, 1986; Dell et al., 1997; MacKay, 1982). Although this type of selection mechanism suffices to account for error patterns, additional assumptions are needed to account for latencies.

A number of models explain differences in naming latencies with lateral inhibition between activated word representations (e.g., Cutting & Ferreira, 1999; Harley, 1993; Stemmer, 1985). The more activated one word (or unit) is, the more it inhibits (decreases) the activation of other words. Words must reach some threshold before they are selected and the time it takes for one word to win out by suppressing others is reflected in naming latencies (for discussion of selection mechanisms in localist networks, see Schade & Berg, 1992). By contrast, in WEAVER++ (Levelt et al., 1999), the timing of word selection is influenced by two factors. The first is the activation of the to-be-selected word relative to all other activated lexical representations in a response set (embodying what is often termed a Luce ratio; Luce, 1959). As in many other models, the more activated a word is relative to other words, the more likely it is to be selected. The other factor is termed a *critical difference*, whereby a lexical representation can only be selected if its activation exceeds the activation of all other representations by some minimum amount. This critical difference is important, as it implies that one alternative representation with high activation might be a more formidable competitor than two alternative representations each with half of its activation (because the one with higher activation is more likely to exceed the critical difference threshold).

Word-production models also must account for why semantically related distractor words interfere more with word production than unrelated words do. Almost every model of word production that has aimed to explain semantic interference attributes it generally to the fact that as a speaker tries to select the most appropriate word for a stimulus, the word representations of semantically related distractors are activated more strongly and so more formidably compete for selection of the alternative form than semantically unrelated distractors (for one exception, see Costa et al., 2003). Generally, this is assumed to occur because the lexical representations of semantically related distractors receive activation from two sources: the distractor words themselves and through their semantic relation to intended words. The representations of unrelated distractor words, in contrast, do not receive this latter source of activation. For example, when naming a picture of a lion, the lexical representation of *tiger* would be activated not only by the distractor word

*tiger* but also by the semantic representation of *LION*, whereas the lexical representation of *table* would only be activated by the distractor word *table*.

All else being equal, it takes more time for speakers to generate the names of objects that have multiple context-appropriate names (such as *TV/television* or *weights/barbells*) than those that have a single dominant name (such as *tooth* or *bomb*; Lachman, 1973; Lachman, Shaffer, & Hennrikus, 1974). Within the language of spreading activation, less codable meanings initially activate the representation of the word that is eventually produced (e.g., *TV*) less than highly codable meanings do and simultaneously provide more activation to at least one other word (*television*), which will act as a competitor. Thus, models can explain these effects of codability on word production although the effect is not usually addressed.

### 1.2.3. *Property 4: Competition for selection is constrained by grammatical class and contextual features*

Not just any semantically related word competes for selection with the most appropriate word to express a meaning. Critically, the competition is limited to words of the same grammatical class. That is, only nouns substitute for nouns, verbs for verbs, and so on (Fromkin, 1971; Garrett, 1975; Nooteboom, 1973). Likewise, distractor words only appear to interfere with word production when they share grammatical class, verb transitivity (e.g., Schriefers, Teurel, & Meinhausen, 1998), and other context-relevant syntactic features with the most appropriate or target word (Schriefers, 1993; Tabossi, Collina, & Sanz, 2002). Even when substituting words are only phonologically related to the word they replace, they show a strong tendency to share the same grammatical category (e.g., Fay & Cutler, 1977). Thus, constraints on maintaining the grammatical class of an intended word appear stronger than the constraints on expressing the intended meaning.

Models of word production typically invoke different processing mechanisms to impose syntactic constraints on word selection as opposed to semantic constraints. Several models posit syntactic frames in which content words are inserted after selection. The selection mechanism is blind to word representations that do not fit the slot it is trying to fill, such that only a noun can fill a noun slot (e.g., Dell, 1986; MacKay, 1982; Stemberger, 1985).

### 1.2.4. *Property 5: The speed and accuracy of selection is affected by specific meaning-level properties*

A number of factors related to semantic representations and their mapping to word representations affect the speed and accuracy with which a word is selected and produced. The concreteness or imageability<sup>2</sup> of a word is one such factor (Morrison, Chappell, &

---

<sup>2</sup> The heaviness of verbs seems to be a similar variable. Specific verbs such as *run* are considered heavier than general verbs such as *go* (see e.g., Gordon & Dell, 2002).

Ellis, 1997). Presumably, words with concrete, imageable meanings such as *vampire*, *wind*, and *pine* have richer semantic representations that guide word selection more efficiently than words with more abstract meanings such as *fear*, *sense*, and *spirit* (De Groot, 1992). Higher imageability or greater concreteness facilitates word translation (De Groot, 1989), generating associations (Cattell, 1889), and word recall (e.g., Martin, Lesch, & Bartha, 1999). However, highly imageable words appear more prone to substitution by semantically related words, perhaps due to a tendency to share more meaning features with other words (Martin, Saffran, & Dell, 1996). Similarly, names of objects from structurally similar categories and, in particular, living things seem more error-prone than artifacts (for comprehensive review and a theory, see Humphreys & Forde, 2001).

In addition, sentence context clearly influences the speed of word selection, probably through the influence of a combination of pragmatic, semantic, and syntactic constraints. The more predictable a word is in an utterance (based on other people's attempts to guess it from context), the less likely a speaker is to silently pause or say *um* before saying it in spontaneous speech (Goldman-Eisler, 1958b), in laboratory settings (Goldman-Eisler, 1958a), and the faster a speaker will label a corresponding object (Griffin & Bock, 1998).

Explaining such effects on word selection within the context of decompositional views is fairly straightforward. The properties that make particular word meanings imageable or concrete also bestow those meanings with additional features and thus richer meanings (e.g., Gordon & Dell, 2002, 2003; Hinton & Shallice, 1991). Likewise, sentence contexts may increase or sharpen the features specifying the meaning to be lexicalized. In turn, these additional features should increase the activation levels of consistent word representations while activating fewer potential competitors. With respect to imageability and concreteness, non-decompositional views might take an analogous stance, but rather than propose that imageable and concrete word meanings have richer meanings, they might propose that those meanings participate in more richly interconnected meaning networks, which might selectively promote the activation of imageable or concrete word-meanings.

#### 1.2.5. *Property 6: Attended objects do not necessarily lead to lexical activation*

In the semantic interference effect, hearing or seeing a semantically related word interferes with generating another word. This suggests that word representations might be easily activated based on any strongly associated stimulus in the environment, even objects. For example, seeing a banana might activate the word *banana* to some extent, even in the absence of any intention to talk about it. Indeed, several models make this prediction (e.g., Humphreys & Forde, 2001; Roelofs, 1997). In contrast to the semantic interference effect that these theories would predict, viewing a semantically related object facilitates production relative to viewing an unrelated object (e.g., viewing a banana while trying to generate the word *apple*; Bloem & La Heij, 2003; Damian & Bowers, 2003). This suggests that word representations may only be activated by meaning or visual representations in the presence of an intention to communicate about them. Speakers even tend to gaze directly at objects while generating words to inaccurately label them (i.e., lie about them; Griffin & Oppenheimer, 2003). So, it seems that distractors

easily influence production processing via language comprehension processes but not via object recognition processes (for some exceptions that may be due to failure to focus on what to express, see Harley, 1990).

However, the data regarding name activation for ignored visual information are not completely clear-cut. At the same time as there is evidence of perceived objects failing to show any semantic interference effects (Bloem & La Heij, 2003; Damian & Bowers, 2003), other researchers have found results that suggest phonological information about ignored objects becomes available (Morsella & Miozzo, 2002; Navarrete & Costa, 2005). Specifically, speakers are faster to name an object (e.g., a cat) in the presence of a distractor object with a phonologically similar name (*cap*) than an unrelated name (*shoe*). This is problematic given that all models require that phonological activation be mediated by word representations, so that under the same conditions that one sees phonological activation of names, it should be possible to detect semantic interference just as in other situations.

The possibility that speakers only linguistically activate words they intend to speak is an important characteristic for models of word production to take account of. The solution is to restrict activation from freely flowing from semantic representations to word representations, limiting the flow to meanings within a pre-verbal message (for examples, see Bloem & La Heij, 2003; Bloem, Van Den Boogaard, & La Heij, 2004). Specifically, Bloom and La Heij (2003) propose that until a semantic-level representation reaches a threshold level of activation, it is unable to influence word representations (only other semantic representations), and that an intention to speak is necessary to achieve that threshold activation. Note that this in essence introduces a kind of discreteness to the word-production process between the levels of meaning representation and linguistic representation.

Restricting activation flow is likely to have additional consequences and potentially explain other observations. For example, selective activation may resolve the seeming contradiction that on the one hand, imageable words are produced more quickly and accurately, whereas they are also relatively more prone to word substitution errors. Specifically, if imageable words have richer meanings or participate in richer semantic networks, then when accessed, they probably lead to the activation of a wider cohort of semantically related word meanings. When that widely activated cohort of word meanings does not extend to lexical activation, the most appropriate word representation should not suffer lexical-level competition, and so the greater activation of the intended word meaning should be free to be easily selected. At the same time, if we assume that speakers sometimes select the wrong word meaning for production or fail to highly activate important features for distinguishing similar objects, then because of the wider semantic cohort, speakers should be more likely to select the wrong meaning to lexicalize or activate a wider range of competing words when generating highly imageable word meanings than less imageable meanings.

Restrictions on activation flow between semantic and word representations cannot be blindly added to any model of production. For instance, WEAVER++ requires that meaning-level representations freely activate lexical representations in order to explain

lexical-competition effects. For example, if the concept *FISH* (as activated through “is-a” links from the concept *BIRD*) is unable to activate the lexical representation of *fish* because *FISH* is not in the message, lexical competition between *fish* and *bird* is not possible and the lexical competition effects described above go unexplained.

*1.2.6. Property 7: Selecting words is subject to long-term repetition effects that resemble learning*

Selecting a word to name an object or express a meaning has long-lasting effects on how easy it is to retrieve that word again to express a similar meaning. One manifestation of this is in repetition priming for naming objects that lasts over months (Cave, 1997) and retrieving the same name for different exemplars of the same type of object (e.g., multiple cars) over the course of an experimental session (Bartram, 1974). This increase in availability is also reflected in perseveratory speech errors, such as calling a giraffe *zebra* after correctly naming a zebra (Vitkovitch & Humphreys, 1991). Also relevant are observations of increased latencies to name objects or actions when presented with other items from the same taxonomic category (the semantic homogeneity effect; e.g., Vigliocco et al., 2002). Note that these semantic interference and strengthening effects only occur when speakers must select words to label pictures, as sentence completions, or in some other way that is based on meaning. The perseveratory effects are not produced by reading words aloud or categorizing words as names for artifacts or natural objects (e.g., Vitkovitch & Humphreys, 1991).

Generally speaking, such long-term effects invite explanations in terms of learning. This highlights a notable gap in models of word production: Unlike the subfields of word reading (e.g., Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996), phonological assembly (Dell, Juliano, & Govindjee, 1993), and even sentence production (Chang, 2002; Chang, Dell, & Bock, 2006), no major model of word production has emerged that accounts for learning effects during word production. Long-term repetition priming effects and the semantic homogeneity effect could be explained via automatic strengthening of connections between meaning representations (most readily imagined as semantic features) and a selected word representation whenever the generated word successfully expresses the meaning. The fact that such long-term effects occur only with conceptually mediated production is consistent with an explanation that requires a mapping and selection rather than simple activation of meaning or word representations (as in categorization and reading). This implies that expressing the same meaning with the same word should subsequently be more efficient, but expressing a similar meaning with a different word should be more difficult. The landscape of models of word production would benefit if it included an implemented and fully developed model with learning principles that could explain these kinds of effects.

*1.2.7. Property 8: Word production can halt part of the way through the process*

Slips of the tongue such as saying *frish gotto* for *fish grotto* (Fromkin, 1971) suggest that speakers may correctly select the words that they intend to say but then err in

assembling their sounds. Sometimes, speakers fail to retrieve any of the sounds of a word that they want to say. Researchers refer to this as being in a *tip-of-the-tongue* or TOT state, from the expression “to have a word on the tip of one’s tongue.” TOT states occur most often for people’s names (Burke, MacKay, Worthley, & Wade, 1991), but can also be elicited in the laboratory using definitions of obscure words such as “What do you call a word that reads the same backwards as forwards?”<sup>3</sup> TOTs are more likely for meanings with low imageability (Astell & Harley, 1996) and uncommon words (Burke et al., 1991; Harley & Bown, 1998). Some forms of brain damage cause people to experience similar word retrieval problems for common words. When trying to come up with one of these elusive words, both brain-damaged and unimpaired speakers are able to provide a great deal of general world knowledge associated with the word and, moreover, information about the word’s syntactic properties, such as whether it is a count or mass noun in English (Vigliocco, Martin, & Garrett, 1999), its grammatical gender in Italian, French, or other languages with fairly arbitrary grammatical categories for nouns (Badecker, Miozzo, & Zanuttini, 1995; Miozzo & Caramazza, 1997b), and identify the correct form of the auxiliary for sought-for verbs in Italian (Miozzo & Caramazza, 1997a). Although speakers cannot say the whole word they seek, they often can identify its first letter or sound, its number of syllables, and words that sound similar (Brown & McNeill, 1966). Provided with candidates for missing words, the speaker can (with exasperation) reject unintended words and (with gratitude) identify intended ones. Diary studies indicate that in normal life outside the laboratory, speakers typically come up with the TOT word minutes or days later. In the lab, cueing the speaker with sounds from the missing word increases the likelihood that it will “spontaneously” occur to them (James & Burke, 2000; Meyer & Bock, 1992) and providing a homophone before a TOT-eliciting stimulus makes a TOT less likely to occur (e.g., eliciting *cherry pit* makes speakers more likely to successfully name Brad Pitt; Burke, Locantore, Austin, & Chae, 2004).

Two models of production have sought to specifically explain TOT states: WEAVER++ (Levett et al., 1999) and Node Structure Theory (MacKay, 1987; Burke et al., 1991). TOT states starkly illustrate the architectural assumptions of WEAVER++. After having selected word representation to express a meaning, the retrieval of the next required representation, the lexeme or phonological form of the word, fails. The successful selection of the lemma representation explains speakers’ confidence that they know a word to express and their ability to report the word’s syntactic characteristics (which are stored at the same level in the theory), whereas the failed selection of the lexeme representation explains speakers’ inability to articulate the word. Node Structure Theory takes a similar stance, except without the lexeme (i.e., a complete lexical–phonological representation). Specifically, it postulates that a (non-phonological) lexical node fails to fully activate (allow selection of) phonological information (syllables, segments, etc.) due to weakened connections between representations. According to both accounts, partial access to phonological knowledge is accounted for by claiming partial activation of phonological representations, either as mediated by a lexeme node (WEAVER++) or

---

<sup>3</sup> Answer: palindrome.

not (Node Structure Theory; for further discussion, see Harley & Bown, 1998). Analogous explanations within either account can also explain the similar anomic states of brain-damaged individuals, in which they can report the grammatical genders of words that they cannot articulate (Badecker et al., 1995).

### 1.3. Generating function words and morphemes

#### 1.3.1. *Property 9: The selection of some words and morphemes is not primarily driven by meaning*

The above-described properties can be relatively uncontroversially ascribed to the production of content words such as nouns and verbs. The production of function words such as articles (*a, the*) and morphemes such as past tense *-ed*, however, seems less driven by properties of meaning and highly dependent on the grammatical and phonological properties of accompanying words. For example, in English, the forms of indefinite determiners vary depending on whether they are used with count or mass nouns (e.g., *some pasta* vs. *a noodle*); in Swedish, indefinite determiners vary with the grammatical gender of the nouns they modify (e.g., *ett bord* [a table] vs. *en stol* [a chair]); and in French, possessive pronouns vary with the grammatical gender and phonological form of the accompanying words (e.g., *mon chapeau* [my hat] and *mon arbre* [my tree] for masculine nouns but *ma table* [my table] and *mon ampoule* [my light bulb] for feminine; from Janssen & Caramazza, 2003). This dependence on other words results in gender interference effects for determiners in which presenting a distractor noun that has a different grammatical gender than the intended object name delays speech onset for a noun phrase. So, for speakers of Dutch, production of a noun phrase such as *het groene huis* [the green house] is delayed by seeing the word *tafel* [table] that takes the definite determiner *de* relative to seeing the word *been* [leg] which takes the same determiner, *het* (Schriefers, 1993).

Other evidence also points to an important dissociation between content- and function-word production. One is that the well-known observation that in agrammatic aphasias, function word production is notably impaired, despite the high frequency and phonological simplicity of function words (for cross-linguistic review, see Bates, Wulfeck, & MacWhinney, 1991). Another is that function words participate in speech errors in different ways than content words do (Dell, 1990; Garrett, 1975; Stemberger, 1985).

Two different classes of explanation have been proposed to account for the function-content word difference. The original explanation was that function word production is associated with syntactic production, in particular a stage termed *positional processing* (Garrett, 1975). Function words are integral parts of the positional frames with which speakers bind their to-be-produced content words. The use of particular function words is conditional upon meaning-level properties, so positional selection must be sensitive to meaning-level features. Nonetheless, such retrieval is performed differently than is done with content words, which are claimed to be retrieved in a manner like that described earlier. A second related proposal is that function words are selected after the retrieval of the

content words that license their use (Ferreira, 2000; Levelt, 1989). The general idea is that a speaker might retrieve, say, a noun, which in turn will trigger the retrieval of knowledge (Ferreira, 2000) or execution of a procedure (termed *indirect election* by Levelt, 1989) that retrieves the needed function words. Again, this amounts to a claim that a function word is not retrieved directly by meaning, but is instead mediated by the syntactic properties of content-word knowledge.

#### 1.4. Assembling the Sounds of a Word

##### 1.4.1. *Property 10: The sounds of a word are assembled anew*

A potentially counterintuitive idea is that the individual sounds of words are assembled anew each time they are spoken rather than retrieved as intact wholes. Yet, patterns of speech errors and latency data suggest that this is the case. According to one estimate, errors involving sounds occur approximately 2.6 times per 1000 sentences or 1.5 times per 10,000 words in spontaneous speech, whereas word errors occur at a rate of 4.4 per 1000 sentences or 2.5 per 10,000 words (Deese, 1984). When unimpaired speakers name isolated objects, errors involving the sounds of words are much less likely than word substitution errors (Dell et al., 1997). Sound errors include omissions, additions, and exchanges of individual sounds. The most common type of error is the anticipation of an upcoming sound (Nooteboom, 1973) as in *alsho share* for *also share* (Fromkin, 1971).

Levelt and colleagues have argued that one reason that a word's sounds must be assembled anew each time is due to changes in metrical structure contingent on the accompanying words and inflections (e.g., Levelt et al., 1999). For example, whereas the /d/ is syllabified with the other sounds in *hand* when the word is spoken alone, it is syllabified with the following word, *it*, in the utterance *Hand it* ["han-dit"]. The importance of metrical structure can also be seen in benefits of repeating syllable structure independent of syllable content (Sevald, Dell, & Cole, 1995). That is, speakers can repeatedly produce *kem-til-fer* much faster than they can produce *kem-tilf-ner*, because the first two syllables of the first sequence share syllable structure (CVC) whereas the first two syllables of the second sequence do not (CVC and CVCC).<sup>4</sup>

##### 1.4.2. *Property 11: Experience strongly affects speed and accuracy of assembling words*

A striking fact about slips of the tongue is the way they reflect both long- and short-term experience with language patterns. Words fall apart in ways that reflect the sequences of sounds a speaker is most familiar with. Slips of the tongue are more likely to create words that exist in a speaker's language rather than create novel sequences of sounds, a phenomenon known as *lexical bias* (Baars, Motley, & MacKay, 1975; Dell & Reich, 1981). Even when novel sequences are created, sounds in these new sequences only occur in syllable positions that they occupy in existing words of the language.

---

<sup>4</sup> There were controls for length and syllable complexity, of course.



The sounds that slip tend to be those in the least predictable positions within the language of the speaker (Berg, 1998; Nooteboom, 1973). For example, word initial consonants (e.g., /b/ in the word *bicycle*: /baj.sI.kl/) are less predictable than other consonants that begin syllables in English (e.g., the /s/ and /k/ in *bicycle*), and in Germanic languages, slips of the tongue are more likely to separate a word initial consonant from a word than another syllable initial consonant (Shattuck-Hufnagel, 1987). This also shows up in word games and sayings where everything except the initial consonant of a word is repeated (e.g., Pig Latin and “helter skelter” or “piggely wiggly”). Languages like Spanish that do not have this difference in distribution of word-initial vs other syllable-initial consonants do not show this difference in phonological speech errors (Berg, 1991). Whole syllables often participate in speech errors in Mandarin Chinese (Chen, 2000, as cited by Chen, Chen, & Dell, 2002) but rarely in English. Likewise, in Semitic languages such as Arabic, where morphemes are discontinuous and the syllable positions of consonants change more often across words, speech error and poetic rhymes pattern very differently from languages like English where morphemes tend to be concatenated and maintain syllabic position across words (for comprehensive review, see Berg, 1998).

Even short-term experience with particular sound-ordering conventions affects the likelihoods of making different types of errors. So, when in the context of a particular experiment, sounds only occupy particular positions, speakers’ speech errors come to reflect these biases even when they are not part of the language in general. For example, when /f/ only occurs at the beginning of syllables in an experiment, speech errors involving /f/ nearly always involve the beginning of syllables (Dell, Reed, Adams, & Meyer, 2000). Even individual phonological features such as place of articulation (e.g., the difference between /b/, /d/, and /g/) are sensitive to these effects of experience (Goldrick, 2004).

Speakers appear sensitive to the frequency of whole words in addition to sequences of sounds within them. Unsurprisingly, children tend to learn common words earlier than uncommon words (e.g., Huttenlocher et al., 1991). Thus, it is difficult to determine whether it is the age at which a word is typically learned (its *age-of-acquisition*), how often it tends to be used (its *word frequency*), or both that affect word production (for a discussion of attempts, see e.g., Brysbaert & Ghyselinck, in press).<sup>5</sup> More common or earlier learned words are generated as much as 100 ms rapidly than less common words (Oldfield & Wingfield, 1964). This speed advantage for common words may be due to the benefits of experience in word selection and phonological encoding, but several results suggest that the impact of frequency and age-of-acquisition is greater in phonological encoding than in word selection (see Brysbaert & Ghyselinck, in press). For

---

<sup>5</sup> The same objects and concepts tend to be mentioned equally often across languages such that a word that is common in one language will tend to have a translation with a similar frequency (Bates et al., 2003). However, the lack of word frequency or age-of-acquisition effects in tasks that only require identifying objects or retrieving semantic information suggests that their effects arise during word production rather than in accessing semantic representations (e.g., Oldfield & Wingfield, 1964).

example, lower word frequency increases the likelihood of phonological word substitutions, slips of the tongue (Dell, 1990; Stemberger & MacWhinney, 1986), and TOTs (Burke et al., 1991; Harley & Bown, 1998), but only seems to affect the likelihood of semantic word substitutions in unimpaired speakers when they are under heavy time pressure to speak (Vitkovitch & Humphreys, 1991).

Although they sound the same, the meanings and therefore the initial stages of word production differ for generating *homophones*, such as *ball* meaning a spherical object and *ball* meaning a formal dance. Given that they sound the same, one might suspect that their phonological processing, phonological representations, or both would be the same. Indeed, several experiments suggest that homophones share phonological processing and representations. Although *ball* meaning a formal dance is a very uncommon word, it is less prone to slips of the tongue than equally unusual words (Dell, 1990). In fact, providing an aphasic with practice generating one meaning of a homophone carries over to improved performance in generating the other for about a week (Biedermann, Blanken, & Nickels, 2002). Likewise, as described above, producing the meaning of one member of a homophone pair (*pit* as in cherry pit) makes a TOT less likely for the other member of the pair (e.g., *Pitt* as in Brad Pitt; Burke et al., 2004). Hearing a word related to one homophone (e.g., *dance* related to the formal dance meaning of *ball*) speeds the naming of objects corresponding to the other homophone meaning (i.e., *ball* meaning a spherical object; Cutting & Ferreira, 1999) and speeds reading aloud the ambiguous printed word (e.g., *ball*; Balota & Paul, 1996). All of these results are readily accounted for by assuming that different meanings of homophones share a common representation of phonological form but differ in the lexical representations accessed through meaning (e.g., Dell, 1990; Jescheniak & Levelt, 1994). Alternatively, all lexical and morphological representations could be separate for the two homophone meanings, but bi-directional connections to shared sounds might lead to strong priming effects by allowing activation to spread back and forth between separate representations. Only Caramazza's (1997) word production model explicitly excludes both shared representations for homophones and bi-directional activation, making it difficult to see how it could account for homophone effects. However, this feature of the model readily allows it to account for experiments suggesting that low-frequency homophones might not be produced as quickly as their high-frequency partners or their combined frequencies when important factors are controlled (Caramazza, Costa, Miozzo, & Bi, 2001; Jescheniak, Meyer, & Levelt, 2003).

Unlike the above-noted effects of the frequency of words or word patterns, the effect of the frequency of syllables upon production is less clear. Levelt and Wheeldon (1994) reported that the frequency of the final syllables of words influences production time independently of word frequency or whole-word naming time. More recent experiments showed that for disyllabic non-word production, Dutch-naming latencies (the same language assessed by Levelt and Wheeldon) were influenced by the frequencies of first syllables but not second syllables (Cholin, Levelt, & Schiller, in press). More work is necessary to sort out exactly when syllable frequency does and does not affect production times.

#### 1.4.3. *Property 12: Aspects of sound assembly proceed sequentially*

Processing seems to start earlier in time for sounds at the starts of words than for sounds at the ends. For example, in picture–word interference tasks, distractors that share the initial sounds in object names have effects at earlier points in time than distractors that share later sounds of names (Meyer & Schriefers, 1991; e.g., an initially overlapping word like *tile* facilitates naming of a tiger at earlier points in time relative to a word like *liar* that only overlaps in final sounds). When speakers are told in advance that they will be asked to articulate one word out of a set that share initial sounds, they begin speaking earlier than if the word comes from a group that does not share initial sounds (Meyer, 1990, 1991). That is, speakers appear able to prepare the shared parts of the words in advance, leaving less material to prepare and allowing faster production than when nothing is known in advance about the form of the upcoming word. However, this foreknowledge is only helpful in Dutch when words share initial sounds and metrical structure (specifically, number of syllables and stress pattern; Roelofs & Meyer, 1998). In contrast, knowing in advance the final sounds or syllables of words provides no benefit (Meyer, 1990, 1991). Knowing just a phonological feature such as place of articulation at the start of a word also does not provide any detectable benefit, suggesting that the relevant level of preparation involved is whole phonemes or sounds (Roelofs, 1999). Somewhat similarly, generating a word with word-initial overlap (e.g., *tile* and *tiger*) slows naming of an object on a subsequent trial, whereas generating a word with word-final overlap (*anger* and *tiger*) speeds naming relative to generating an unrelated word (Wheeldon, 2003). When speakers repeat word pairs multiple times, it takes more time per pair for combinations that share initial sounds relative to those that share no sounds, which in turn take more time than combinations that only share final sounds (Sevold & Dell, 1994). Although there are intriguingly different patterns of effects for hearing a word with overlapping initial segments vs. generating one, all of these results suggest a sequential process associated with retrieving, organizing, or programming speech sounds.

The WEAVER++ model covers all bases by having both simultaneous retrieval of all segments in a word, followed by a step in which each segment is associated with a syllable position in sequential order (Levelt et al., 1999). Thus, speed of processing in the model is sensitive to the availability of all phonemes in the first part of phonological encoding and the time needed to sequentially associate them with a syllable position. In addition, there is a final stage of phonetic encoding in which the phonologically specified syllables of words are sequentially associated with stored articulatory gestures. While having two stages that show sequential processing allows the model to account for sequential effects in production, it also makes it difficult for the model to simultaneously account for the absence of certain length effects. Specifically, object naming latencies and gaze durations on objects suggest that when the many potential confounds with word length are controlled, speakers take the same amount of time to prepare a multi-syllabic word as a monosyllabic one (Bachoud-Levi, Dupoux, Cohen, & Mehler, 1998; Bonin, Chalard, Meot, & Fayol, 2002; Griffin, 2003; Sternberg, Knoll, Monsell, & Wright, 1988; for discussion, see Meyer, Roelofs, & Levelt, 2003).

#### 1.4.4. Property 13: The effect of similar sounding words is highly situation-dependent

The effect of recent experience with a word that is phonologically similar to an intended word sometimes speeds (e.g., Starreveld, 2000) and sometimes slows word production (e.g., Wheeldon, 2003), indicating that such effects depend on a complex set of factors. These differing effects may be due in part to experimental paradigms differentially calling on phonological subprocesses, such as sound retrieval as opposed to sound sequencing, and similar sounding words having different effects in sound retrieval, associating sounds with metrical structure, translating these phonological plans into motor programs, and articulation (see Levelt et al., 1999; O'Seaghdha & Marin, 2000). Complicating the interpretation of phonological priming effects in production (e.g., Starreveld, 2000), similar sounding words compete with one another in word recognition (e.g., Tanenhaus et al., 1995). In addition to the position in the words where sounds are shared (Sevold & Dell, 1994), the duration and type of processing that the first or priming word undergoes appears critical in shaping effects (O'Seaghdha & Marin, 2000).

When a speaker unintentionally asks for *balaclava* in a Mediterranean restaurant (rather than *baklava*), it is tempting to conclude that similar word forms compete against one another. Instead, speakers may say words that sound similar to their intended words as near-misses, in which they fail to retrieve all of the sound information for an intended word and default to a very similar form (Burke et al., 1991). In this category of effects, one can list the tendency for slips to be real words rather than novel sequences of sounds, the tendency for intruders in phonological word substitutions to have the same number of syllables and other characteristics as intended words (Fay & Cutler, 1977; Gagnon et al., 1997), the tendency for speakers in TOT states to often come up with similar sounding words (Burke et al., 1991), the tendency of slips of the tongue to involve sounds that share many phonological features, such as /t/ and /k/ rather than /t/ and /v/ (Fromkin, 1971; Shattuck-Hufnagel & Klatt, 1979), and the tendency for sounds to exchange between similar sounding words (e.g., Dell & Reich, 1981). That is, words that sound alike do not appear to interfere and compete with one another during phonological encoding in the way semantically related words do in word selection. Indeed, although selecting a word from a semantically dense neighborhood seems to take more time than selecting one from a sparse neighborhood, the opposite seems to hold for phonologically defined neighborhoods. Words that share many sounds with other words take less time to generate than words that are more unusual (Vitevitch, 2002) and appear more likely to be successfully retrieved in terms of fewer phonologically related word substitutions (Vitevitch, 1997) and TOTs (Harley & Bown, 1998). Also supporting the idea that similar sounding words support each other rather than compete is the observation that priming with phonologically related words can resolve and prevent TOTs (James & Burke, 2000; Meyer & Bock, 1992). Likewise, presenting phonologically related distractor words during object naming speeds naming latencies relative to unrelated distractors (e.g., *liar* vs. *ankle* for a lion; Schriefers et al., 1990). Simulation studies conducted with interactive activation models suggest that feedback of activation from phonological neighbors may aid intended words in competing against their semantic neighbors (Dell & Gordon, 2003).

That said, there are situations in which having similar sounding words slows speech or increases the likelihood of errors. The most obvious case of such interference is in tongue twisters such as *The sixth sick sheik's sixth sheep's sick*. Experiments indicate that repeating words with similar initial sounds is significantly more difficult than repeating sequences with unrelated sounds (Sevold & Dell, 1994). In addition, speakers are slower to generate a name for an object (e.g., *hoed* meaning *hat* in Dutch) when they generated a word with overlapping initial sounds (*hond*) on the preceding trial rather than an unrelated word (Wheeldon, 2003). Dell and O'Seaghdha (1992) suggested that these and related phenomena reflect sequentially cued phonological competition, whereby having completed a sequence of phonemes (e.g., /i/) with one ending (/k/, in *sheik*), cueing of the recently used ending makes it more difficult to subsequently complete that sequence (/i/) with a different ending (/p/, in *sheep*). Such sequential competition is readily accounted for with the subclass of connectionist models called simple recurrent networks and control signal networks that output phonological segments one at a time for a given input (e.g., Dell et al., 1993; Vousden, Brown, & Harley, 2000). Although such models do an excellent job of producing some phenomena associated with phonological word assembly (particularly the effects of experience, similarity, and order on speech errors), it is unclear how they would be integrated with other parts of the production system to account for phenomena such as phonological influences on word selection.

## 1.5. Time course of processes in word production

### 1.5.1. Property 14: Semantic competitors activate their sounds

Despite the near consensus on the need for two stages to the production process, a famous controversy among theories of word production concerns the extent to which processing of sound and meaning overlap in time. In one manifestation of this, researchers have debated whether sound-related information is only processed after word selection is complete (e.g., Dell & O'Seaghdha, 1992; Dell & Reich, 1981; Harley, 1993; Levelt et al., 1991; Peterson & Savoy, 1998). On one side are models that characterize the flow of information during production as strictly staged – speakers first use activated meaning-level representations to perform word selection and only access sound information after the completion of the selection process. The most prominent model of this strictly discrete sort is the WEAVER++ model presented in Levelt et al. (1999), which was developed computationally in Roelofs (1992, 1997). Other theorists have also argued for the strict separation of word selection and sound processing stages (e.g., Butterworth, 1989; Caramazza, 1997). On the other side are models that assume staged processing, but allow activation to flow relatively freely among meaning, lexical, and sound representations, making multiple types of information relevant to both word selection and sound assembly (e.g., Dell, 1986; Harley, 1993). Specifically, partially activated but ultimately unselected lexical representations are permitted to influence sound assembly (via *cascading activation*). For example, before ultimately naming an object as *couch*, a speaker should activate both the word representation for *couch* and its synonym *sofa* (see Property 2, that speakers activate a family of meaning-related words) and, via cascading, the sounds of these words. Indeed, speakers are faster to read aloud words that are

phonologically related to dispreferred synonyms of object names (e.g., *soda* for *sofa* when preparing to name a couch) when they are presented after beginning to prepare to name a drawing of a couch rather than a completely unrelated object (Peterson & Savoy, 1998; replicated by Jescheniak & Schriefers, 1997; see also Jescheniak & Schriefers, 1998). The WEAVER++ model makes the post-hoc assumption that word selection is delayed until after sound processing begins only in the case of synonyms (Levelt et al., 1999).

Another manifestation of this controversy has concerned whether the sounds of ultimately unselected words may influence which word is selected. In models with bi-directional flow of activation or *feedback*, partially activated but ultimately unselected phonological representations are allowed to send activation backwards to affect lexical (and perhaps even semantic) levels of representation. The most prominent implemented model of this interactive sort is presented in Dell (1986), but this type of interactive activation has been incorporated in many theories and models (e.g., Dell et al., 1997; Eikmeyer, Schade, Kupietz, & Laubenstein, 1999; Harley, 1993; MacKay, 1982, 1987; Stemberger, 1985). Explaining the *mixed error effect* is one of the primary motivations for assuming this type of interactivity. It turns out that the intruding words in semantically related word substitutions bear a greater than chance phonological similarity to the intended words that they replace (Brédart & Valentine, 1992; Dell & Reich, 1981; Harley, 1984; Martin, Weisberg, & Saffran, 1989). In interactive-activation models with feedback, when generating the word *cat*, activation spreads to words related in meaning to *cat* such as *dog*, *mouse*, and *rat*, and via their word representations to their sounds. The sounds that form the word *cat* are highly activated by their link to *cat's* word node and they relay a portion of that activation to other words containing the same sounds such as *cap*, *kit*, and *rat*. Thus, a word that is both semantically and phonologically related to the intended word such as *rat* receives converging activation from both semantic and phonological representations, making it more likely to be selected by mistake than a word activated by only one of these sources. In contrast, discrete two-stage models account for mixed errors uses an independently motivated error-checking mechanism (see e.g., Motley, Camden, & Baars, 1982). The basic idea is that the more a substituting word resembles an intended word, the less likely a pre-articulatory editing mechanism is to detect the error and prevent it from being uttered. Thus, under this account, mixed errors are not made disproportionately often, it is just that other errors are more likely to be detected and prevented, making the types of errors observed unrepresentative of those created in the language production system (Butterworth, 1982; Levelt, 1989). Thorough treatments of the issues of discreteness and interactivity in word production can be found in Rapp and Goldrick (2000) and Vigliocco and Hartsuiker (2002).

### 1.5.2. *Property 15: The scope of message planning is greater than the scope of sound assembly*

Early in the study of speech error patterns, researchers noted that there was a greater distance between words that exchange places than between sounds that exchange places. For example, Nooteboom (1973) noted that 2.1 syllables separated exchanging sound

segments, whereas 4.1 intervening syllables was the average distance between exchanging units of greater size such as morphemes and words. Such observations support the distinction between word and sound representations and separate processing stages that operate on them (e.g., Fromkin, 1971; Garrett, 1975). In addition, it suggests that abstract properties of words are specified further in advance than their sounds are.

Speakers seem to create proposition-sized pre-linguistic messages before they begin articulating an utterance. A rational argument is based on observing that the order of mention and syntactic structure of clauses<sup>6</sup> depends on factors, such as the relative animacy, humanness, and agency of its message elements, suggesting that speakers may compare these prior to making any ordering or syntactic choices (e.g., Ferreira, 1994; Griffin & Bock, 2000). Empirically, speakers are more likely to pause, repeat words, and say *um* before articulating complex utterances and clauses than less complex ones (e.g., Beattie, 1979; Clark & Wasow, 1998; Ford, 1982). Syntactic complexity presupposes a complex message representation, so it is more parsimonious to attribute such effects to message planning than syntactic planning. Studies of simultaneous translators also suggest that speakers prefer to have a proposition-sized message prior to initiating an utterance (Goldman-Eisler, 1972). Speakers start long utterances with a higher pitch than shorter utterances (e.g., Cooper, Soares, & Reagan, 1985), indicating some degree of advance knowledge of content. At the start of a clause, speakers are slower to respond in secondary tasks (Ford & Holmes, 1978) and tend to avert their gaze from a listener (Kendon, 1967), also suggesting that they are engaged in more intense processing of some sort at these points in time.<sup>7</sup> Likewise, equating for distance in words, nouns in the same clause are more likely to lead to subject–verb agreement errors than nouns in different clauses (Bock & Cutting, 1992), suggesting that items within the same proposition in a message are more available than those from different propositions.

Although speakers seem to know a lot about the message content of an utterance before they begin to speak, they often do not know all of the words they will use to express the message before they begin to articulate the utterance. Evidence from a wide variety of experimental tasks suggests that speakers often select the nouns that follow verbs while articulating the first words of their sentences (e.g., Griffin & Bock, 2000; Kempen & Huijbers, 1983; Lindsley, 1975; Smith & Wheeldon, 1999). However, variations in instructions or other aspects of a situation lead speakers to alter the number of words they prepare prior to speaking the first one (e.g., Griffin & Bock, 2000; Wheeldon & Lahiri, 1997). So, the question is what factors determine when speakers prepare and articulate their words?

There is a tradition in psycholinguistics of searching for units in which planning is incremented. With respect to the minimum amount of planning a speaker must complete

---

<sup>6</sup> A clause contains a verb and roughly corresponds to a simple sentence.

<sup>7</sup> Shifting topics in a conversation is also associated with decreased fluency (e.g., Butterworth, 1975), suggesting that speakers engage in resource-demanding macroplanning spanning multiple clauses.

before beginning a fluent utterance, the primary units proposed have been based on prosody or syntax. In the psycholinguistic literature, a *phonological* or *prosodic word* is typically defined as a single content word along with any adjacent, unstressed function words as in [*beer's a*] [*good*] [*thing*] (Ferreira, 1993; Wheeldon & Lahiri, 1997). Latency data indicate that speakers prepare at least one phonological word prior to initiating an utterance, with more complex phonological words delaying speech onset (Wheeldon & Lahiri, 1997). Other studies suggest that speakers will prepare more than one phonological word prior to speech if it does not form a whole lexical word (i.e., it is half of a compound; Wheeldon & Lahiri, 2002); if the second phonological word is part of the first noun phrase (Costa & Caramazza, 2002); and if the first word will not take long to articulate and speakers try to avoid pausing (Griffin, 2003). Strengthening the case for considering the phonological word an important unit at some level is the observation that the latency to begin articulating pre-planned speech is a function of the number of phonological words the pre-planned utterance contains (Sternberg et al., 1988; Wheeldon & Lahiri, 1997).

Other researchers have argued for phrase-wise word planning (e.g., Martin, Miller, & Vu, 2004). Certainly in many languages (e.g., Dutch, German, and Spanish), grammatical dependencies between nouns (e.g., *beer*, *ale*) and the adjectives (e.g., *hoppy*, *amber*) and determiners (e.g., *a*, *some*) that modify them make it necessary to retrieve information about the noun to determine the correct form of the adjective or determiner. Not surprisingly, picture–word interference studies suggest that nouns are selected before the onset of the determiner when speakers produce gender-marked determiner + adjective + noun phrases in languages such as Dutch and German (Schriefers, 1992; Schriefers, de Ruiter, & Steigerwald, 1999). Other work points to phrase-wise planning even in English speakers in the absence of strong grammatical dependencies. For example, English-speaking patients who, because of brain damage, have difficulty maintaining lexical–semantic information had greater difficulty producing utterances in which adjectives appeared in the same phrase as the noun they modified (e.g., *the long, brown hair*) than utterances in which the adjectives appeared in a different phrase (e.g., *the hair was long and brown*; Martin & Freedman, 2001). Tellingly, patients with impaired memory for phonological information did not show this difference and could produce these utterances as readily as unimpaired speakers.

There is mixed evidence for pre-speech planning of multiple nouns when they occur in a conjoined noun phrase such as *monkey and chair*. Support for phrasal planning comes from finding of semantic interference effects on speaking latencies for both objects within a conjoined noun phrase (Meyer, 1996; but see Meyer, 1997) and when the nouns in the conjoined phrase name semantically related objects (Freedman, Martin, & Biegler, 2004). All else being equal, timing experiments indicate that speakers take about 70 ms longer to initiate sentences with two nouns in a conjoined subject noun phrase than sentences with a single noun (Martin et al., 2004; Smith & Wheeldon, 1999). Such observations have been used to argue that the contents of a noun phrase are processed in parallel (with a small cost) and that articulation of a sentence-initial conjoined noun phrase is not initiated until both nouns are prepared. In contrast, eye-tracking



experiments suggest that under similar circumstances, speakers prepare nouns one at a time, shortly before uttering them, even in complex subject noun phrases or conjunctions (Griffin, 2001; Meyer, Sleiderink, & Levelt, 1998; but see Morgan & Meyer, 2005).

An unanswered question is whether verbs (or other predicates) play a special role in the preparation of utterances. Based on the constraining properties of verbs, some theorists have suggested that verb selection must normally take place early in sentence formulation (e.g., Bock, 1987; Ferreira, 2000; Jarvella, 1977; MacWhinney, 1987). When not required to select verbs in an utterance, speakers begin speaking earlier than they otherwise do (Kempen & Huijbers, 1983; Lindsley, 1975). Such results have been used to argue that verb selection precedes subject selection and therefore often speech onset (e.g., Bock & Levelt, 1994; Ferreira, 2000). However, these same experiments (Kempen & Huijbers, 1983; Lindsley, 1975) are also consistent with a desire to have a full or partially specified message planned before speech onset without verb selection, assuming that messages that include an action or other predicate take more time to compose, all else being equal, than those with only a topic. Similarly, the relationship between ear-to-mouth lag and verb position in translation input (Goldman-Eisler, 1972) supports the idea that a verb is selected before translated production begins, but also the more conservative possibility that production processes wait for a predicate to be included in the message. Further complicating matters is the possibility that speakers may only need to prepare verbs prior to speech onset whenever verbs occur soon after sentence onset (e.g., after short subject noun phrases in English) simply because there would not be time to prepare them while articulating the subject noun phrase (Griffin, 2003).

In addition to semantic and linguistic units and dependencies, time also appears to be important in timing speech. Longer words by definition take more time to articulate than shorter words do and slower speakers take more time to articulate their words than faster speakers do. Both of these aspects of timing have been shown to influence when speakers begin preparing words (Griffin, 2003). That is, speakers may attempt to minimize their buffering of prepared words by estimating how long words will take to prepare and how long it will take to articulate already prepared speech. Speakers are sensitive enough to the timing of articulation and word preparation that they will insert optional words such as *that is* *The coach knew that you missed practice* in response to variations in the availability of the following word (Ferreira & Dell, 2000). Also suggesting sensitivity to the time needed to prepare upcoming speech, speakers are more likely to say *uh* than *um* before shorter delays in speaking (Clark & Fox Tree, 2002).

## 2. SUMMARY

This chapter has described 15 basic properties of spoken language production. These properties characterize word production as consisting of a word-selection stage followed by a sound-processing stage (#1). Selecting a content word such as a noun or verb

involves activating (#2) and then competitively selecting (#3) from a family of meaning-related words in a grammatically constrained (#4) but meaning-sensitive (#5) fashion. This word-selection process may require an intention-to-name to have it commence (#6), and it manifests a long-term learning component (#7). Nonetheless, word production can fail partway through (#8). Function words may undergo a somewhat different selection process than content words do (#9). Sound processing in turn is characterized as assembling sequences of sounds (#10), a process that is affected by speakers' experience (#11), and proceeds from word start to end (#12). Phonological similarity has complex effects on production, attesting to the fact that it probably affects multiple subprocesses (#13). Although only one word may ultimately be spoken to produce a meaning, multiple meaning-related candidates can affect the availability of sound information (#14). Finally, speakers plan messages further in advance than they retrieve sounds, showing a tendency to prepare words for about a noun phrase at a time, due to message-level, syntactic, prosodic, and/or timing constraints or preferences (#15).

In focusing on spoken language and the production of words in particular, we have left untouched the literature on written language production (see e.g., Bonin et al., 2002; Kellogg, 2003), the production of sign languages (e.g., Thompson, Emmorey, & Gollan, 2005), and the complications of knowing words in multiple languages (e.g., Costa, Miozzo, & Caramazza, 1999; Gollan & Acenas, 2004; Kroll & Sunderman, 2003). Within spoken word production, this chapter has not addressed work on how speakers produce morphologically complex words (e.g., Badecker, 2001; Melinger, 2003; Roelofs, 1996; Wheeldon & Lahiri, 2002; for discussion, see Waksler, 2000) such as *morphology*, *litterbox*, or *ko-tätaste* (a Swedish word meaning “most tightly packed with cows”) or idioms such as *It's Greek to me* and *to put one's foot in one's mouth* (see e.g., Cutting & Bock, 1997; Levelt et al., 1999). We have hardly touched on the production of prosody and the role of intonation in spoken language (for discussion, see Ferreira, 1993; Wheeldon, 2000). Nor have we discussed under what circumstances and how speakers may or may not tailor their language to suit their audiences (Barr & Keysar, this volume; Ferreira & Dell, 2000; Ferreira, Slevc, & Rogers, 2005; Horton & Gerrig, 2005; Kraljic & Brennan, 2005; Lockridge & Brennan, 2002). These are active and important areas of research in language production.

Most of the properties we have reviewed are sufficiently basic that they are virtually certain to characterize how production works, at least to some level of approximation. A few of them, however, are more controversial and are likely to be explicated and revised by future research (e.g., whether the intention to name is critically involved in word activation [6], seriality in phonological encoding [12], and origins of phonological similarity effects in production [13]). Nonetheless, in all, these properties represent a tribute to the progress that the field of language production has made, as they represent true gains in our understanding of how speakers produce words. At the same time, they pose challenges to current and future models of production, as such models pursue their goal of transforming these descriptions of how production works into explanations of why it works the way it does.

## REFERENCES

- Astell, A. J., & Harley, T. A. (1996). Tip-of-the-tongue states and lexical access in dementia. *Brain and Language*, *54*, 196–215.
- Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *14*, 382–391.
- Bachoud-Levi, A. C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, *39*, 331–346.
- Badecker, W. (2001). Lexical composition and the production of compounds: Evidence from errors in naming. *Language and Cognitive Processes*, *16*, 337–366.
- Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, *57*, 193–216.
- Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 827–845.
- Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, *6*, 325–356.
- Bates, E., D'Amico, S., Jacobsen, T., Szkely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C.C., Pechmann, T., Plh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin and Review*, *10*, 344–380.
- Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Cross-linguistic research in aphasia: An overview. *Brain and Language*, *41*, 123–148.
- Beattie, G. W. (1979). Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics*, *17*, 61–78.
- Berg, T. (1991). Phonological processing in a syllable-timed language with pre-final stress: Evidence from Spanish speech error data. *Language and Cognitive Processes*, *6*, 265–301.
- Berg, T. (1998). *Linguistic structure and change: An explanation from language processing*. Oxford: Oxford University Press.
- Biedermann, B., Blanken, G., & Nickels, L. (2002). The representation of homophones: Evidence from remediation. *Aphasiology*, *16*, 1115–1136.
- Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition*, *42*, 23–60.
- Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*, *48*, 468–488.

- Bloem, I., Van Den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, *51*, 307–323.
- Bock, J. K. (1987). Coordinating words and syntax in speech plans. In: A. Ellis (Ed.), *Progress in the Psychology of Language* (Vol. 3, pp. 337–390). London: Erlbaum.
- Bock, J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*, 99–127.
- Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In: M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego: Academic Press.
- Bonin, P., Chalard, M., Meot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*, 89–114.
- Boomer, D. S. (1978). The phonemic clause: Speech unit in human communication. In: A. W. Seigman & S. Feldstein (Eds.), *Nonverbal behavior and communication* (pp. 245–262). Hillsdale, NJ: Erlbaum.
- Brédart, S., & Valentine, T. (1992). From Monroe to Moreau: An analysis of face naming errors. *Cognition*, *45*, 187–223.
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337.
- Brysbart, M., & Ghyselinck, M. (2006). The effect of age-of-acquisition: Partly frequency-related, partly frequency-independent. *Visual Cognition*, *13*, 992–1011.
- Burke, D. M., Locantore, J. K., Austin, A. A., & Chae, B. (2004). Cherry pit primes Brad Pitt: Homophone priming effects on young and older adults’ production of proper names. *Psychological Science*, *15*, 164.
- Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*, 542–579.
- Butterworth, B. (1982). Speech errors: Old data in search of new theories. In: A. Cutler (Ed.), *Slips of the Tongue and Language Production* (pp. 73–108). Berlin: Mouton.
- Butterworth, B. (1989). Lexical access in speech production. In: W. D. Marslen-Wilson (Ed.), *Lexical Representation and Process* (pp. 108–135). Cambridge, MA: MIT Press.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.
- Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1430–1450.
- Cattell, J. M. (1889). Mental association investigated by experiment. *Mind*, *14*, 230–250.

- Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, 8, 322–325.
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26, 609–651.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234–272.
- Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29, 217–229.
- Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46, 751–781.
- Cholin, J., Levelt, W. J. M., & Schiller, N. (2006). Effects of syllable frequency in speech production. *Cognition*, 99, 205–235.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84, 73–111.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37, 201–242.
- Cooper, W. E., Soares, C., & Reagan, R. T. (1985). Planning speech: A picture's words worth. *Acta Psychologica*, 58, 107–114.
- Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, 46, 178–198.
- Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Level of categorization effect: A novel effect in the picture–word interference paradigm. *Language and Cognitive Processes*, 18, 205–233.
- Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41, 365–397.
- Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of *chipmunk*, *cherry*, *chisel*, *cheese*, and *cello* (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132, 163–201.
- Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, 25, 57–71.
- Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 318–344.

- Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture–word interference tasks. *Psychonomic Bulletin and Review*, *10*, 111–117.
- Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345–361.
- De Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 824–845.
- De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1001–1018.
- Deese, J. (1984). *Thought into speech: The psychology of a language*. Englewood Cliffs, NJ: Prentice-Hall.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*, 313–349.
- Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In: N. O. Schiller & A. S. Meyer (Eds), *Phonetics and phonology in language comprehension and production* (pp. 9–38). Berlin: Mouton de Gruyter.
- Dell, G. S., & O’Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.
- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611–629.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149–195.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*, 1355–1367.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in normal and aphasic speakers. *Psychological Review*, *104*, 801–838.
- Eikmeyer, H.-J., Schade, U., Kupietz, M., & Laubenstein, U. (1999). A connectionist view of language production. In: R. Klabunde & C. V. Stutterheim (Eds), *Representations and Processes in language production* (pp. 205–236). Wiesbaden: Deutscher Universitaets Verlag.
- Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, *8*, 505–520.

- Ferreira, F. (1993). The creation of prosody during sentence production. *Psychological Review*, *100*, 233–253.
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33*, 715–736.
- Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In: L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). London: Psychology Press.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.
- Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*, 263–284.
- Fodor, J. A. (1975). *The language of thought*. New York, NY: Cromwell Co.
- Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In: J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 797–827). Cambridge, MA: MIT Press.
- Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, *6*, 35–53.
- Freedman, M. L., Martin, R. C., & Biegler, K. (2004). Semantic relatedness effects in conjoined noun phrase production: Implications for the role of short-term memory. *Cognitive Neuropsychology*, *21*, 245–265.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52.
- Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (1997). The origins of formal paraphasias in aphasics picture naming. *Brain and Language*, *59*, 450–472.
- Garrett, M. F. (1975). The analysis of sentence production. In: G. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.
- Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In: A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London: Academic Press.
- Glaser, W. R., & Döngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640–654.
- Goldman-Eisler, F. (1958a). The predictability of words in context and the length of pauses in speech. *Language and Speech*, *1*, 226–231.
- Goldman-Eisler, F. (1958b). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology*, *10*, 96–106.

- Goldman-Eisler, F. (1972). Segmentation of input in simultaneous translation. *Journal of Psycholinguistic Research*, 1, 127–140.
- Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, 51, 586–603.
- Gollan, T. H., & Acenas, L. A. R. (2004). What is a tot? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and Tagalog–English bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30, 246–269.
- Gordon, J. K., & Dell, G. S. (2002). Learning to divide the labor between syntax and semantics: A connectionist account of deficits in light and heavy verb production. *Brain and Cognition*, 48, 376–381.
- Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, 27, 1–40.
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82, B1–B14.
- Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review*, 10, 603–609.
- Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science*, 15, 814–821.
- Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language*, 38, 313–338.
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279.
- Griffin, Z. M., & Oppenheimer, D. M. (2003). *Looking and Lying: Speakers' gazes reflect locus of attention rather than speech content*. Paper presented at the 12th European Conference on Eye Movements, Dundee, Scotland, August.
- Harley, T. A. (1984). A critique of top–down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8, 191–219.
- Harley, T. A. (1990). Environmental contamination of normal speech. *Applied Psycholinguistics*, 11, 45–72.
- Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, 8, 291–309.
- Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighborhood effects in speech production. *British Journal of Psychology*, 89, 151–174.
- Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74–95.



- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142.
- Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: “category-specific” neuropsychological deficits. *Behavioral & Brain Sciences*, *24*, 453–509.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248.
- James, L. E., & Burke, D. M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1378–1391.
- Janssen, N., & Caramazza, A. (2003). The selection of closed-class words in noun phrase production: The case of Dutch determiners. *Journal of Memory and Language*, *48*, 635–652.
- Jarvella, R. J. (1977). From verbs to sentences: Some experimental studies of predication. In: S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 275–305). Hillsdale, NJ: Erlbaum.
- Jescheniak, J. D., & Schriefers, H. (1997). Lexical access in speech production: Serial or cascaded processing? *Language and Cognitive Processes*, *12*, 847–852.
- Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1256–1274.
- Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 432–438.
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.
- Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, *39*, 170–210.
- Kellogg, R. (2003). Written language production. In: T. Hermann & J. Grabowski (Eds), *Encyklopedie der Psychologie, Themenbereich c: Theorie und forschung, Serie III: Sprache bd. 1*. Goettingen: Hogrefe.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.
- Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*, 185–209.

- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63.
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50, 194–231.
- Kroll, J. F., & Sunderman, G. (2003). Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representations. In: C. Doughty & M. Long (Eds.), *Handbook of second language acquisition*. Cambridge, MA: Blackwell.
- Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, 99, 199–208.
- Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior*, 13, 613–625.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1–22.
- Levelt, W. J. M., & Wheeldon, L. R. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science*, 22, 1–45.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 142–163.
- Lindsay, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, 7, 1–19.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review*, 9, 550–557.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Lupker, S. J. (1979). The semantic nature of response competition in the picture–word interference task. *Memory and Cognition*, 7, 485–495.
- MacKay, A. J., Connor, L. T., Albert, M. L., & Opler, L. K. (2002). Noun and verb retrieval in healthy aging. *Journal of the International Neuropsychological Society*, 8, 764–770.
- MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210–227.

- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, 89, 483–506.
- MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer.
- MacWhinney, B. (1987). The competition model. In: B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
- Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory–verbal STM capacity and lexical errors in repetition. *Brain and Language*, 52, 83–113.
- Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language*, 28, 462–485.
- Martin, R. C., & Freedman, M. L. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory*, 9, 261–280.
- Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41, 3–29.
- Martin, R. C., Miller, M., & Vu, H. (2004). Lexical–semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, 21, 625–644.
- McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.
- Melinger, A. (2003). Morphological structure in the lexical representation of prefixed words: Evidence from speech errors. *Language and Cognitive Processes*, 18, 335–362.
- Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29, 524–545.
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30, 69–89.
- Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language*, 35, 477–496.
- Meyer, A. S. (1997). Conceptual influences on grammatical planning units. *Language and Cognitive Processes*, 12, 859–863.
- Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory and Cognition*, 20, 715–726.

- Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146–1160.
- Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*, 131–147.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33.
- Miozzo, M., & Caramazza, A. (1997a). On knowing the auxiliary of a verb that cannot be named: Evidence for the grammatical and phonological aspects of lexical knowledge. *Journal of Cognitive Neuroscience*, *9*, 160–166.
- Miozzo, M., & Caramazza, A. (1997b). Retrieval of lexical-syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1410–1423.
- Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 428–442.
- Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559.
- Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 555–563.
- Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*, 578–594.
- Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*, 359–377.
- Nooteboom, S. G. (1973). The tongue slips into patterns. In: V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 144–156). The Hague: Mouton.
- Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature*, *202*, 1031–1032.
- O’Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1–17.
- Osgood, C. E. (1963). On understanding and creating sentences. *American Psychologist*, *18*, 735–751.

- Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539–557.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.
- Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*, 59–87.
- Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*, 854–876.
- Roelofs, A. (1997). The weaver model of word-form encoding in speech production. *Cognition*, *64*, 249–284.
- Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*, 173–200.
- Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 922–939.
- Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 246–251.
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production 2. The simulational perspective. *Journal of Psycholinguistic Research*, *21*, 435–462.
- Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, *45*, 33–54.
- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 841–850.
- Schriefers, H., de Ruiter, J. P., & Steigerwald, M. (1999). Parallelism in the production of noun phrases: Experiments and reaction time models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 702–720.

- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.
- Schriefers, H., Teruel, E., & Meinhausen, R. M. (1998). Producing simple sentence: Results from picture-word interference experiments. *Journal of Memory and Language*, *39*, 609–632.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.
- Sevold, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*, 91–127.
- Sevold, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, *34*, 807–820.
- Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In: E. Keller & M. Gopnik (Eds), *Motor and sensory processes of language* (pp. 17–51). Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*, 41–55.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In: J. Gumperz & S. C. Levinson (Eds), *Rethinking linguistic relativity: Studies in the social and cultural foundations of language* (pp. 70–96). New York: Cambridge University Press.
- Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*, 205–246.
- Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, and Computers*, *28*, 516–536.
- Starreveld, P. A. (2000). On the interpretation of onsets of auditory context effects in word production. *Journal of Memory & Language*, *42*, 497–525.
- Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 686–698.
- Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 896–918.
- Stemberger, J. P. (1985). An interactive activation model of language production. In: A. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). London: Erlbaum.

Stemberger, J. P., & MacWhinney, B. (1986). Form-oriented inflectional errors in language processing. *Cognitive Psychology*, *18*, 329–354.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, *45*, 175–197.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1140–1154.

Tabossi, P., Collina, S., & Sanz, M. (2002). The retrieval of syntactic and semantic information in the production of verbs. *Brain and Language*, *81*, 264–275.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Thompson, R., Emmorey, K., and Gollan, T. (2005). Tip-of-the-fingers experiences by ASL signers: Insights into the organization of a sign-based lexicon. *Psychological Science*, *16*, 856–860.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*, 442–472.

Vigliocco, G., Martin, R. C., & Garrett, M. F. (1999). Is “count” and “mass” information available when the noun is not? An investigation of tip of the tongue states and anomia. *Journal of Memory and Language*, *40*, 534–558.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. J. M. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*, B61–B69.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech*, *40*, 211.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 735–747.

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 664–680.

Vitkovitch, M., & Tyrrell, L. (1999). The effects of distractor words on naming pictures at the subordinate level. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *52*, 905–926.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*, 101–175.

Waksler, R. (2000). Morphological systems and structure in language production. In L. R. Wheeldon (Ed.), *Aspects of Language Production* (pp. 227–248). Philadelphia, PA: Psychology Press.

Wells, R. (1951/1973). Predicting slips of the tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 82–87). The Hague: Mouton.

Wheeldon, L. (2000). Generating prosodic structures. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 249–274). Philadelphia, PA: Psychology Press.

Wheeldon, L. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, 18, 81–109.

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37, 356–381.

Wheeldon, L. R., & Lahiri, A. (2002). The minimal unit of phonological encoding: Prosodic or lexical word. *Cognition*, 85, B31–B41.



This page intentionally left blank

## Chapter 3

### Syntax and Production

*Fernanda Ferreira and Paul E. Engelhardt*

#### 1. INTRODUCTION

Our capacity for language enables us to communicate virtually any thought or idea, and this is in large part because the language faculty evolved a syntactic component. Syntax allows words to be combined to create unique combinations of meaning. And although all human languages share some universal syntactic properties – for example, all allow the creation of some type of plural – the constraints on how constituents may be generated vary substantially. Thus, an English speaker knows that verbs usually come before their objects, whereas Japanese speakers learn the opposite setting of this parameter for arranging words and arguments.

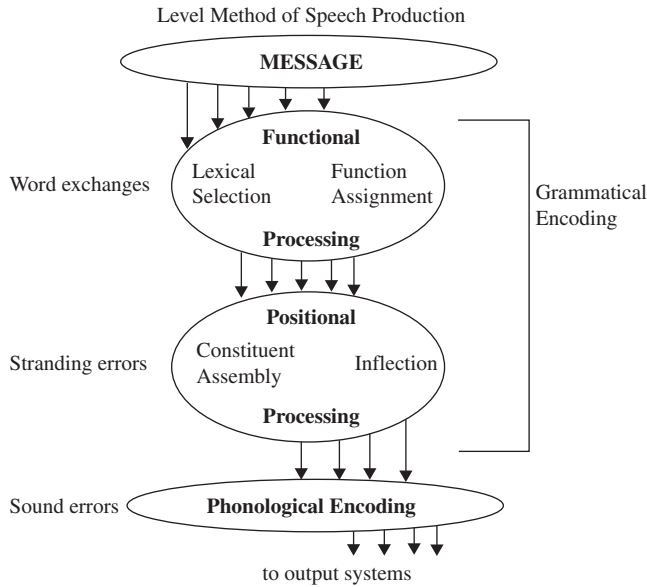
But even though the grammar of any particular language constrains the way words may be put together, all languages also give the speaker some freedom of choice. Even English, which is often characterized as a language that offers little in the way of word ordering options (MacWhinney, Bates, & Reinhold, 1984), allows the speaker to choose from among at least a few different forms to express the same essential idea. A proposition involving a cat, a dog, and a state of fear may be grammatically encoded as *my cat terrifies the dog next door*, or *the dog next door is terrified of my cat*, or *it's my cat that terrifies the dog next door*, and so on. An important insight that has emerged from work on language production is that these syntactic options are used not just to convey different information structures (e.g., that the cat and not the dog is given information, and therefore the cat should take the subject position of the sentence), but also to take advantage of the states of activation within the language and cognitive systems at the moments when speakers make their syntactic decisions. Thus, any model of language production must explain how speakers create utterances optimally given constraints on processing capacity, while at the same time respecting grammatical constraints, at least most of the time.

This chapter is a review of syntax in language production, and focuses on three critical issues. The first addresses what a syntactic representation for production looks like. We consider to what extent it is hierarchical and ordered. Then we discuss whether it contains empty categories such as traces of constituent movement, as would be expected according to theories of syntax which assume that questions, relative clauses, passives, and other complex structures are formed via movement. We also evaluate whether grammatical information is represented and accessed in a way that is purely form-based, or whether structures are actually small trees associated with specific lexical items, which get combined to create utterances (F. Ferreira, 2000). In the second major section, we ask whether grammatical encoding is an automatic or a resource-demanding process, and we consider whether certain structures are inherently more difficult to produce than others. We also try to determine whether the system generates structure incrementally, or whether production involves planning units larger than a single word (e.g., clauses). In the third section, we now review the quite extensive literature on syntactic choice in production, examining how people choose between syntactic forms such as active and passive, and how they decide whether to include optional words like complementizers (e.g., *Mary knows (that) her cat terrifies the dog*). In the final section, we briefly consider the question of how syntax affects prosody during production. We conclude that syntactic structure is created in two stages, using lexically stored syntactic templates that include representations of empty categories. This process of grammatical encoding is computationally demanding, which is why the system makes optimal use of scarce resources by being moderately incremental. And the prosody of a sentence reflects its syntactic organization, both because syntax affects prosodic constituency, and because difficulties associated with syntactic planning can affect phonological phrasing.

## 2. SYNTACTIC REPRESENTATIONS IN PRODUCTION

### 2.1. Two-Stage Models of Grammatical Encoding

As has been argued since the earliest days of generative grammar, syntax is an interface between meaning and sound (articulation/phonetic form). A word such as *cat* has a particular meaning, but the expressive power of language is enhanced immeasurably by our ability to create meanings compositionally, by putting words together – for example, our ability to say *not a cat* or *that's my cat*. Models of production instantiate this basic architecture fairly, transparently. Consider the Bock and Levelt (1994) model, which was described in the previous *Handbook of Psycholinguistics*. The model, henceforth referred to as Bock–Levelt (BL), is shown below (reprinted from the 1994 chapter):



The process of speaking begins with a message-level representation, which captures the idea the speaker wishes to convey. This message becomes sound at the other end of the model, at a stage called phonological encoding. Linking the message and phonological levels are two stages of syntactic processing (or grammatical encoding, as it is called in the model), one called functional processing, and the other positional processing. Notice that the basic linguistic architecture in which syntax mediates between meaning and form is replicated in the BL model of production. Yet an important difference is that syntactic operations are factored into two components. This two-stage architecture originated with Garrett (1975), who argued from speech error data that the production system first creates the global, syntactically functional structure for a sentence, and in a separate stage determines phrasal details such as serial order.<sup>1</sup>

In BL, grammatical encoding begins with functional-level processing. Abstract lexical entries termed *lemmas*, which contain information about a word’s meaning and its syntactic requirements but do not represent its phonology, are retrieved and assigned grammatical functions such as subject and object.<sup>2</sup> For example, for the utterance *my cat*

<sup>1</sup> The Fromkin model of speech errors, which is a forerunner of BL and of Garrett’s work, included more than two stages of syntactic processing.

<sup>2</sup> Here the models deviate from generative approaches to syntax in assuming that the notions ‘subject’ and ‘object’ are syntactic primitives, whereas since the earliest days of generative grammar, grammatical relations have been computed from the geometry of a phrase structure representation. For example, the subject is the Noun Phrase (or Determiner Phrase) immediately dominated by the sentence node. Other approaches to formal syntax such as Lexical-Functional Grammar (LFG) and Relational Grammar do treat notions such as subject and object as primitives.

*terrifies the dog next door*, the lemmas for CAT, DOG, NEXT, DOOR, and TERRIFY would be retrieved, and CAT would be assigned the role of subject and DOG (modified by NEXT DOOR) the role of object. At this point, then, the speaker has committed to some type of active structure in which CAT will be the subject; a passive structure is ruled out, because in any type of passive, DOG would be the subject. But notice that a structure such as *the dog next door, my cat terrifies him* is still possible, because in this form CAT is the grammatical subject and DOG is the object (*dog* is the object in both the preposed position and in the pronominal form *him*). The difference between the regular active and this left-dislocated construction is a matter of constituent ordering, which is left undecided at this stage of grammatical encoding.

The second component of syntactic processing takes place at the positional level, which operates on the functional-level representation. At this point, serial order is imposed on the utterance. Beginning with the initial constituent, each grammatical function created earlier (e.g., subject, object, modifier) is translated into a linearized constituent. The grammatical encoder retrieves a prestored phrasal frame, which contains slots for all the elements of that phrase – the determiner *my* and the noun *cat*, in the current example. Inflectional affixes are represented as an intrinsic part of the frame, so that if the subject were plural, the plural morpheme would already be in place and would therefore not have to be separately retrieved and inserted.

Because the language production system is assumed to be incremental (see Section 3.3. for a more thorough discussion of incremental production), the order in which lemmas are ‘worked on’ determines the overall order of the phrases in the utterance (F. Ferreira, 2000). So if the lemma for DOG were processed before the one for CAT, then the resulting structure might be the left-dislocation form given above or perhaps a topicalized form such as *Mary my cat loves*. (Although this construction is disfavored in most dialects of English, it can be acceptable given the appropriate context). Thus, positional-level processing determines both the serial order of phrases and the order of elements within any given phrase, and all inflectional processing takes place at this stage of processing as well. For many ordering decisions, the processor simply obeys grammatical constraints such as the requirement that determiners initiate an NP, that adjectives precede nouns, and that verbs precede objects but follow subjects (for English). But because languages give speakers some ordering options, there may be decisions about order that still need to be made, particularly at the within-phrase level. One good example is the sequencing of conjuncts, illustrated in *the dog and cat slept soundly*. The other order of *dog* and *cat* is equally grammatical, and so the choice about how to sequence the conjuncts within the overall NP subject must be based on extra-grammatical considerations (Pinker & Birdsong, 1979; Bock, 1987).

## 2.2. Evidence for Two-Stage Models

The evidence for this two-stage architecture separating functional and positional-level processing comes from two sources: speech error analyses and data from experiments designed to shed light on how structure is created. The argument from speech errors is as

follows. First, speakers sometimes make semantic substitutions, as in *my cat terrifies the boy next door* when *the girl next door* was intended. These errors almost invariably respect a form-class constraint: nouns substitute for nouns, verbs for verbs, and so on. Speakers also sometimes make word exchange errors, illustrated by *my boy terrifies the cat next door*. The interacting elements in these errors usually come from different phrases, and the words tend to be of the same form class. Semantic substitutions and word exchange errors indicate that there is a level of processing at which grammatical category is relevant and at which the roles for lemmas are decided. In contrast, errors such as phonological substitutions and stranding occur as well, but they have quite different characteristics. In a phonological substitution, a word with a similar sound is incorrectly assembled and made part of the utterance. In stranding errors, content morphemes end up misordered but inflectional material does not, as in *I went to get my park trucked* (Garrett, 1980). Notice that the morphemes *truck* and *park* swapped places, but the suffix *-ed* is in its correct location. Phonological substitutions and stranding errors indicate that there is a level of processing at which sound and serial order are decided, and stranding suggests that the inflectional morpheme is an intrinsic part of the phrasal frame.

The experimental evidence for distinguishing functional and positional level processing comes from priming studies, both lexical and syntactic. Let us begin with lexical priming. Notice that the two-stage architecture divides lexical processing so that word *meanings* become available at the functional level, but word *phonology* only gets generated after (and probably after most positional-level processing takes place as well; F. Ferreira, 1993). This is because the sounds of words are (arguably) not relevant for deciding grammatical functions such as subject and object, but (again, arguably) phonology may help the system decide how to sequence words, as suggested by the finding that, in conjuncts, short words tend to precede longer ones (Cooper & Ross, 1975; Bock, 1987), for example. Experiments in which words are either semantically or phonologically primed have demonstrated that making a lemma available (i.e., semantically priming a concept) causes the constituent containing that lemma to be the subject of the sentence. In contrast, phonological priming has either weak effects or leads to late positioning of the constituent containing the word (Bock, 1987; cf. Cleland & Pickering, 2003). This pattern is typically taken to support a division of labor between functional and positional level processing, because the idea is that only a manipulation, which affects the lemma can influence processes hypothesized to be taking place at the functional level. Of course, this interpretation is somewhat compromised by the finding that phonological priming sometimes leads to late constituent placement, but the effect is much smaller and has been argued to reflect a late stage in production where an utterance is evaluated and then changed if it is judged to be deficient before it is articulated (Levelt, 1989).

Another source of evidence for the two-stage architecture comes from syntactic priming. If a speaker produces or even simply hears an utterance with a particular structural form, he or she is likely to mimic that structure in a subsequent utterance. The classic demonstration (Bock, 1986b) involves both the active/passive and the prepositional/double-object dative alternations. Speakers will tend to produce a passive sentence after

hearing or producing one themselves (Levelt & Kelter, 1982; Schenkein, 1980); the same goes for the prepositional dative (*the driver showed the overalls to the mechanic*) and the double-object dative construction (*the driver showed the mechanic the overalls*). (It is not clear that the active can be primed, possibly because of ceiling effects due to its high frequency.) Hartsuiker and Westenberg (2000) discovered using Dutch that a very low-level ordering decision (the sequencing of an auxiliary and a main verb at the end of a sentence) can be primed, leading them to argue for a two-stage model of syntactic processing where a ‘dominance-only’ representation (i.e., one that is not linearized) can prime a representation that is ordered.

Further evidence for the two-stage architecture comes from the way speakers compute agreement relations during sentence production. Consider the fragment *the spokesman who defended the actions*. If this fragment is the subject of a sentence, then it must agree with the main verb. In English, this agreement process is visible mainly on forms of *to be* and *to have* (particularly in the past tense), but in other languages agreement is overt on a wide range of verbs and other words. Carefully designed experiments have revealed that agreement errors occasionally happen, particularly in examples like *the spokesman who defended the actions*, in which the head noun *spokesman* is singular but there is another noun in the subject (*actions*) that is plural (Bock & Eberhard, 1993; Bock & Miller, 1991). Agreement errors turn out to be just as likely in yes/no questions as in declaratives, suggesting that agreement is computed on a representation that specifies dominance but not linear relations (Vigliocco & Nicol, 1998). This argument can be seen in contrast between *the helicopter for the flights are safe* and *Are the helicopter for the flights safe*, where the linear positions of the head noun are different but the likelihood of an agreement error is the same. This result suggests that agreement relations are computed from a syntactic representation created before linearization takes place. Notice that this particular finding is consistent not only with a two-stage view of syntactic processing, but also some version of a transformational/derivational account of grammar, because the linearization process at issue here is the one that moves the verbal material to the front of the sentence to create an interrogative construction. This general idea will be discussed in Section 2.4. when we consider the question whether syntactic representations created during production show evidence of processing attributable to constituent movement. Additional evidence for the idea that hierarchical position but not linear order is critical for computing agreement can be found in a study of complex NPs such as *the computer with the program(s) of the experiment(s)* (Franck, Vigliocco, & Nicol, 2002). Agreement errors were found to be more likely when the medial noun *program* was plural compared to the more proximate noun *experiment*, indicating that position in a hierarchical structure has more effect on agreement than does linear position.

Thus, evidence from speech errors, from syntactic and lexical priming, and from the process of computing subject–verb agreement seem to converge on the idea that syntactic structure is generated in two distinct stages during production. Nevertheless, this architecture has been challenged, and we turn now to evidence that is argued to support a single-stage model of grammatical encoding.

### 2.3. Evidence Challenging Two-Stage Models

First, recall the effects of semantic and phonological priming on grammatical form. If it had turned out that only semantic primes could affect the establishment of grammatical relations, then an architecture separating syntactic generation into a stage that uses only lemma information to assign roles such as subject and object, and a separate stage that uses sound to determine linear order, would have been supported. But recall that phonological primes do have a small but significant effect (Bock, 1987). For example, if participants encountered the word *trump* and then a picture of a truck towing a car, they were likely to say *the car is being towed by a truck*, because the phonological relationship between *trump* and *truck* leads to some type of inhibition. Thus, the effect of a phonological prime appears to be opposite from one that is semantic, but the important point is that according to the classic two-stage architecture, it should have no effect at all. Therefore, it may be argued that this finding undermines two-stage models.

However, there are two problems with this argument. The first was briefly mentioned earlier: It is possible that this effect of the phonological prime occurs not during grammatical encoding but during a stage at which the utterance is checked for overall acceptability (the so-called monitor; see Hartsuiker, Corley, & Martensen, 2005, for a recent discussion of its properties). The second problem with this argument is that the inhibitory effect of phonological primes only challenges the assumptions regarding lexical processing during grammatical encoding – specifically, the idea that lexical retrieval occurs in two stages, with only the second including access of phonology. It is possible that semantic, syntactic, and phonological information about words is all retrieved simultaneously, but that dominance and linear relations are nonetheless computed separately. An important question too is why a phonological prime should be inhibitory rather than facilitatory. Bock (1987) suggested that the effect could be due to lateral inhibition among phonological competitors, but some studies of lexical processing have shown that phonological primes facilitate processing (Grainger & Ferrand, 1996; Tanenhaus, Flanigan, & Seidenberg, 1980). The Bock (1987) finding clearly should be pursued further, and indeed, it has not yet even been replicated.

The second finding that has been argued to undermine the two-stage model of syntactic processing concerns priming in the dative structure (Pickering, Branigan, & McLean, 2002). Consider once again the prepositional-dative, illustrated in *the driver showed the overalls with the stains to the mechanic*. Another grammatical alternative is the shifted form in which the prepositional phrase (PP) precedes the object (*the driver showed to the mechanic the overalls with the stains*), a structure that is more likely to be generated the longer and heavier object (Wasow, 1997). Shifted and non-shifted prepositional datives share the same hierarchical or dominance relations but differ in how the NP and the PP are ordered. Thus, on a two-stage view in which dominance relations are computed separately, the shifted version should prime the non-shifted version. However, such priming does not occur. Based on these results, Pickering et al. (2002) argued for a single-stage model in which dominance and linear relations are computed simultaneously. But the results could be attributed to the peculiarities of the shifted dative form, which is not only



fairly rare (even in the Pickering et al. experiments in which measures were taken to elicit them) but also seems to require fairly strict discourse conditions to be felicitous (Hawkins, 1994). These properties of the shifted prepositional dative might compromise its ability to prime any other construction. It would be very useful to see whether this result can be found using a less marked structure. Exploring this possibility might require consideration of languages that allow more flexibility in constituent ordering than English does.

#### 2.4. Do Syntactic Structures Contain Evidence of Constituent Movement?

Perhaps the most distinctive characteristic of generative grammar compared to other approaches to syntax is its assumption that syntactic structures are generated by movement. Anyone who has taken even an undergraduate course in cognitive psychology knows that in the earliest versions of this theory, noncanonical structures such as passives were created by rearranging the basic active structure (Chomsky, 1965). Somewhat less well known is the transition to the Government and Binding (GB) theory, which assumed that syntactic representations contain evidence of movement. For instance, a passive such as *the dog next door<sub>i</sub> was bitten t<sub>i</sub> by my cat* requires movement of the NP *the dog next door* from the post-verbal position to the subject position, but the starting position of the phrase is marked in the representation with a trace (indicated with the *t*). The trace allows the phrase to be interpreted as the object of the verb *bite* even though it is no longer in object position in the surface structure. The same holds for structures such as wh-questions and relative clauses: A sentence such as *which dog<sub>i</sub> did my cat bite t<sub>i</sub>?* is created by moving the wh-NP (*which dog*) to the top of the tree, again leaving behind a trace so that *dog* can be interpreted as the object of *bite*. It has been common in the psycholinguistic literature to refer to traces as gaps and to moved constituents as fillers (J. D. Fodor, 1978, 1989, 1991), and so we will follow this convention for the rest of our chapter.

The question we now turn to is, do the syntactic structures that people create when they talk contain any evidence of constituent movement? It is widely believed that they do not. For example, it has been argued that one way to conceptualize the two-stage architecture for grammatical encoding is to assume that the first stage creates a ‘deep-structure’ representation and that the second creates a ‘surface structure’ representation. It is important to note, however, that the concept of a ‘deep structure’ has really not been part of generative grammar for the last 25 years, and so it would be surprising to find any evidence for it in language production. And, not unexpectedly, we do not.

Bock, Loebell, and Morey (1992) used the syntactic priming technique to distinguish between the direct and mediated (first a deep structure is computed, then a surface structure) approaches to syntactic generation. Participants heard sentences and then repeated them, and then they had to describe an unrelated picture of a simple transitive event. The critical feature of the study was that if the pictures were described in the active voice, the subject would be inanimate and the object animate (e.g., *the clock woke up the boy*). The prime for the picture description (i.e., the heard and repeated sentence) was

either active or passive, and it had either an animate or an inanimate subject. The assumptions behind the design were that both structural form and animate placement would be mimicked in the picture descriptions. Bock et al. reasoned that if the ‘deep structure then surface structure’ hypothesis is right, then both an active with an inanimate object AND a passive with an inanimate surface subject would prime the active picture description, because the passive actually has an inanimate object at deep structure. However, this pattern was not observed; instead it was the surface placement of the animate entity that determined the degree of animacy priming. Hence, Bock et al. concluded that the direct mapping account is correct.

But these results might simply indicate that the old-fashioned transformational model in which passives are generated by modifying an active kernel sentence is indeed no longer viable. More problematic for the view that constituent movement is psychologically real in production is the finding that a passive such as *the plane was landed by the pilot* can be primed with a non-passive such as *the plane was landing by the tower* (Bock & Loebell, 1990). This result suggests that the newer GB version of generative grammar is also wrong for production, because the presence of the gap in the passive sentence does not seem to matter for the priming effect. And this lack of evidence for gaps is consistent with the BL model of production, which is based not on a generative grammar but rather on Lexical Functional Grammar (LFG) (Bresnan & Kaplan, 1985), since LFG does not assume the existence of movement operations or gaps in syntactic representations.

At the same time, there is also some compelling evidence for the presence of gaps in the syntactic representations that speakers generate. One type of evidence is admittedly intuitive, but we believe the effect is so strong, an experiment is hardly necessary (but see F. Ferreira, 1988, for experimental evidence). To begin, consider the way we normally pronounce a sequence such as *to the party* – notice that the preposition *to* undergoes vowel reduction so that it ends up sounding more like ‘t-schwa’ than the citation form /too/. This phonological process is essentially mandatory; it would be odd to pronounce the sentence as ‘too-the-party,’ especially at normal rates of speech. But now consider a sentence involving wh-movement, such as *John knows who<sub>i</sub> Mary talked to t<sub>i</sub> at the party*. The wh-word *who* is semantically and lexically the object of the preposition *to*, which according to GB theory means that a gap must occur in that position if *who* moves (as it must in English). But now notice how the word *to* is pronounced: it is not reduced but rather lengthened, so it ends up sounding more or less like its citation form /too/ (F. Ferreira, 1988). This means that the trace was represented in some way in the syntactic structure guiding the creation of prosodic structure, which allowed it to block the normal process of vowel reduction. The same seems to hold for NP-gaps as well, as in *The boy<sub>i</sub> was spoken to t<sub>i</sub> by his teacher*.

Recent evidence from errors of subject-verb agreement also support the idea that traces are mentally represented during grammatical encoding (Franck, Lassi, Frauenfelder, & Rizzi, in press). An experiment designed to elicit such errors from French speakers showed that displaced direct objects in a cleft construction (*It's the deputy that the senators welcome t*) determine whether errors of agreement occur, even though the object

does not intervene between the head and the verb in the surface word sequence. Franck et al. argue that their results can only be accounted for if we assume not just a single transformational process that turns a set of lexical items into a surface structure, but rather a grammar consistent with the Minimalist Program (Chomsky, 1995) in which syntactic structures are generated through a series of operations termed MERGE, MOVE, and AGREE (only the latter two are relevant for our purposes; MERGE simply refers to the combining of lexical items). According to Minimalism, elements in a syntactic structure cyclically move up the tree until they are in the appropriate position to allow an agreement relation to be checked. Each movement, even intermediate ones, leaves behind a trace. Franck et al. argue that the pattern of agreement errors they observe in French and Italian can only be explained if it is assumed that gaps of intermediate movement interfere with the process of computing agreement. These results are some of the strongest evidence to date that gaps are generated as part of the normal process of creating a syntactic structure for a sentence.

What should be said, then, about the work suggesting that gaps do not exist? At this point, the most unbiased assessment of the state of our knowledge is that this entire issue needs to be examined in much more detail. In fact, it is worth noting that no experiment has ever been conducted to test directly whether gaps are psychologically real in language production (although the Franck et al. paper comes close). One potentially useful observation about our current state of knowledge is that the evidence for and against traces comes from different sources – the evidence against the reality of gaps is based largely on results from syntactic priming, and the evidence for them has come from studies of subject-verb agreement as well as the process of translating syntactic structures to prosodic constituents.

Of course, the traces of wh- and NP-movement are not the only types of empty categories that have been proposed in generative grammar; another important phonetically null category results from ellipsis, as in *Mary can tie her shoes and Natalie can too*. Comprehension studies have shown that people reconstruct the missing material, eventually obtaining the interpretation that Natalie can tie her own shoes (perhaps by first entertaining but then rejecting the so-called ‘strict’ reading on which Natalie ties Mary’s shoes; Shapiro, Hestvik, Lesan, & Garcia, 2003). The production of ellipsis has not been studied at all, so the extent to which the omitted or deleted material is mentally represented is not known. Consider our shoe-tying example. It is clear that lexical forms corresponding to the second verb phrase *tie her shoes* are not retrieved. But at the message level, the speaker almost certainly generates the idea that Natalie is capable of tying her own shoes. The question those interested in grammatical encoding might ask is, what about the levels in-between? Is a VP generated in the second clause and then left unpronounced because the lemmas do not point to any word-forms, as Levelt (1989) suggests? Another intriguing question is what leads speakers to choose one type of ellipsis over another; for example, an alternative to the form above is *Mary can tie her shoes and so can Natalie*. Clearly, many important issues concerning the grammatical encoding of empty categories remain to be even formulated in the field of language production.

## 2.5. Are Syntactic Structures Lexically Anchored?

Another question that has been of interest both in formal linguistics and in psycholinguistics is whether syntactic structures are linked to words – specifically, lemmas. The classic work by Levelt (1989) argued for lexical generation of syntax. In contrast, the BL model of production assumes a non-lexical view of syntactic structure. In BL, trees are conceptualized as ‘control hierarchies,’ which contain no lexical content but instead coordinate the insertion of lexical material which is retrieved and assembled separately (BL, pp. 947–948). This conceptualization is in part based on Bock’s earlier findings suggesting that lexical overlap does not enhance syntactic priming (Bock, 1989; Bock & Loebell, 1990). For example, the amount of priming for a sentence such as *the girl handed the paintbrush to the man* is the same given a prime like *the secretary baked a cake for her boss* and *the secretary gave a cake to her boss*, even though in the latter case the PPs share the same prepositional head (*to*).

However, more recent work using the syntactic priming paradigm suggests that verb identity does increase the magnitude of priming (Pickering & Branigan, 1998, 1999; Cleland & Pickering, 2003). One motivation for examining this issue in careful detail is that many formal theories of syntactic structure assume that words and the syntactic environments in which they may occur are lexically linked. In the earliest versions of transformational grammar, for example, verbs specified the syntactic environments in which they could occur via their subcategorization frames (Chomsky, 1965). A verb such as *put* would be represented as requiring an object and a locative PP. The theory of GB essentially dissolved phrase structure rules altogether in favor of lexical storage of constituent structure, so that all words were represented in the lexicon with their associated arguments (Chomsky, 1981; Stowell, 1981). Retrieval of a word would then automatically bring along its associated structures. This elimination of phrase structure rules was a logical extension of X-bar theory (Jackendoff, 1977), which described a universal format for all phrases regardless of their type. Other models of syntax such as LFG (Bresnan & Kaplan, 1985), Categorical Grammar (Moorgat, 1988; Steedman, 2000), and Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, & Sag, 1985) also connect words and their syntactic environments. The same trend is evident in Tree-Adjoining Grammar (TAG) (Joshi, Levy, & Takahashi, 1975; Joshi, 1985). In TAG, the primitive objects of the grammar are treelets, which consist of a word (a lexical head such as a verb) and the arguments the head licenses.

Given these theoretical perspectives on the representation of words and syntactic structures, and in particular verbs and arguments, it makes sense to expect that syntactic priming would be greater when the main verb in the prime and target sentences overlap. This issue was investigated in a study designed to assess whether priming would be observed in simple dialogue situations (Branigan, Pickering, & Cleland, 2000), and which manipulated verb identity. A confederate and a genuine experimental subject described pictures to one another, and it was found that the naïve participant tended to use the same construction as the confederate. In addition, priming was greater when the verb in the prime and target sentences was the same – the effect was about twice as large. The model of

syntactic generation offered by Pickering and Branigan (1998) assumes that words such as verbs are linked to the phrases with which they may combine (termed *combinatory nodes*). Cleland and Pickering (2003) demonstrated remarkably similar priming effects for noun phrase structure, including enhancement when the head of the noun phrase (the noun) was shared, indicating that this form of representation and the process for creating structures are similar both for clauses and phrases.

These results, then, tell us that the syntactic representations used for language production are ones in which structures may be generated directly from lemmas rather than through the accessing of contentless phrasal templates (Levelt, 1989). F. Ferreira (2000) presented a model for human language production, which uses TAG as the database for creating structures through lemmas (see also F. Ferreira, Lau, & Bailey, 2004). All heads, including verbs, nouns, prepositions, and adjectives, are represented with the arguments that they license. These elementary trees consisting of a head and its licensed arguments are combined to form utterances (see F. Ferreira, 2000 for a description of the operations that combine elementary trees). The model proposed by Pickering and his colleagues is somewhat different, but it shares the basic insight that words and syntactic structures are representationally linked. Thus, although it may be possible for grammatical encoding to take place using control structures that have no lexical content, perhaps via extraction of some type of general schema for forming particular construction types, in general it appears that the syntactic structures used for production are lexically anchored.

### 3. PROCESSING RESOURCES FOR GRAMMATICAL ENCODING

Thus far, we have considered mainly representational issues, focusing particularly on the properties of syntactic structures and the format in which syntactic information is stored. The question we turn to in this section is how these structures are formed, and in particular, how the process of grammatical encoding draws on processing resources, and how computational load is managed. This will lead us to consider the degree to which grammatical encoding is incremental.

#### 3.1. Is Grammatical Encoding Automatic?

Talking generally feels effortless, but even the most fluent speakers occasionally experience some difficulty formulating their utterances. The classic work of Goldman-Eisler (1968) demonstrated that almost half of most people's speaking time is devoted to pauses and disfluencies such as *um* and *er*. Ford (1982) measured spontaneous speech and observed that about 20% of all clauses are preceded by a pause of about one second in duration. This finding suggests that the process of creating syntactic structure is resource-demanding, but the result is not definitive; clauses are both major syntactic and semantic junctures, and it might be that grammatical encoding is automatic but semantic processing requires planning and can therefore be resource demanding. This view was articulated by Levelt (1989). He adopts Kempen and Hoenkamp's (1982, 1987) model of grammatical encoding which assumes that syntactic procedures are modular, and thus

have the characteristics that J. A. Fodor (1983) views as typical of cognitive modules: grammatical encoding processes consult a proprietary vocabulary, they operate whenever they recognize their standard input, and they operate automatically. Thus, it is to message level planning that resources are devoted during production; syntactic decisions are made automatically, and measures of processing load such as reaction time and pausing will not be affected by the complexity of syntactic operations.

Work conducted since the publication of Levelt's book, however, does appear to suggest that syntactic planning demands computational resources, especially as structural complexity increases. F. Ferreira (1991) had people memorize declarative sentences, which they then had to produce upon receipt of a visual cue; latency to begin speaking was measured. The variable that was manipulated was the syntactic complexity of the sentential subject. It was either short (*the river*) or long, and in the long conditions, it was either of low (*the large and raging river*), medium (*the river near their city*), or high (*the river that stopped flooding*) syntactic complexity. Complexity was defined in terms of a node count, so that the more syntactic nodes the subject needed in its representation, the greater its complexity. Ferreira found that as complexity increased, so did production latencies. Interestingly, memorization times were not affected by this variable, suggesting that the effect was particular to the task of speaking. In a second experiment, she orthogonally varied the syntactic complexity of the subject and object in subject-verb-object sentences. Once again, latencies to begin speaking increased with the complexity of the subject, and the object's characteristics had no effect. However, when both the subject and the object were syntactically complex, speakers tended to pause within the sentence, and the preferred pause location was the subject-verb phrase boundary.

Because speakers were not required to generate any of the sentences' content, these effects of syntactic complexity cannot be attributed to any semantic complexity that might be correlated with the syntactic manipulation. Moreover, as speakers had not chosen the syntactic forms themselves either, the effect cannot be attributed to the need to make syntactic decisions. Instead, it appears that simply saying a sentence with a complex structure takes up processing resources. The results also demonstrate that if both the subject and object are complex, speakers divide the utterance into two processing units, one consisting of the subject, and the other consisting of the verb phrase. Notice that this division into processing units respects the syntactic structure of the sentence, and indeed Ferreira observed almost no cases in which participants paused after the main verb rather than before it (see also F. Ferreira, 1993 for further discussion of these issues). This finding that the processing units are syntactic constituents is consistent with the assumption that the difficulty in processing is localized to the syntactic level.

In another study, Smith and Wheeldon (2001) tested whether speakers plan their utterances before saying them. Participants were asked to describe pictures, and they were primed with sentences such as *The spoon and the car move up*. When participants uttered sentences that were syntactically similar to the prime sentence, a reliable 50 ms advantage to begin speaking was observed. Smith and Wheeldon also tested the scope of this effect and found that it held only for the first phrase of an utterance, consistent with

F. Ferreira's (1991) second experiment demonstrating that only the complexity of the subject affects utterance initiation times. The advantage that Smith and Wheeldon found for sentences that had been syntactically primed suggests that we should revisit the phenomenon of syntactic priming in light of this question concerning processing resources. Recall that speakers are more likely to produce a particular construction when they have just heard it or produced it themselves. Based on the finding that syntactic priming is particularly robust in dialogue, Pickering and colleagues have suggested that syntactic priming is used by the production system as a tool for "reducing the load associated with syntactic processing" (Pickering & Branigan, 1999, p. 136). What this idea assumes, of course, is that syntactic generation is a resource-demanding process, so much so that speakers try to find ways of managing and reducing the computational burden.

It should be noted that studies predating the recent psycholinguistic era also demonstrate that syntactic processing is computationally demanding. Johnson (1966) compared the generation of sentences such as *The person who jumped over there is good* and *The person over there who jumped is good*. Because the structure of the second sentence is right-branching, it is less complex according to the Yngve (1960) complexity metric, and so Johnson predicted it would take less time to initiate. This prediction was confirmed. Second, although Rochester and Gill (1973) did not find any effects of what they termed "syntactic complexity" on speech hesitations and disruptions, they did find that such disruptions in speech varied along with the type of nominal modifier people produced. Specifically, speakers were more likely to show speech disruptions before a noun phrase complement (e.g., "The fact that the woman was aggressive threatened the professors") than before a relative clause (e.g., "The book that was written by Millet was lauded by all"). Goldman-Eisler (1968), who like Rochester and Gill (1973) failed to find effects of syntactic *complexity* on hesitation, also found hesitation differences before different types of syntactic forms. If disruptions in speech are a measure of mental load, and more disruptions occur before one particular ordering of words than another, then one structure must have required the use of more mental resources than the other. We turn next to a more detailed consideration of this question concerning the inherent difficulty of certain forms.

### 3.2. Are Some Constructions Difficult to Generate?

In this section, we ask a question that has received surprisingly little attention from experimental psycholinguists. Are some syntactic constructions inherently difficult to produce, or does difficulty arise only when a structure must be generated in an infelicitous discourse context? To see what is the issue here, consider the passive construction, which is often viewed as more complex than the active, and is certainly more difficult to understand (F. Ferreira, 2003). The passive may be harder to produce than the active because it has a noncanonical structure, because it is less frequent, or because it is more complex, in the sense of requiring more syntactic nodes in its phrase-structure representation. Alternatively, it has been argued that the passive may be the \*right\* construction for particular discourse situations. For example, Tomlin (1983) observed that passives are very common in hockey broadcasts, because what the commentator tries to do is make the player in possession of the puck the subject of the sentence. If that player is affected

in some way (e.g., gets checked), then the sentence form the commentator will use is a passive (*Gretzky was checked by his opponent*), because that is the form that allows the topic to be maintained as subject, even when it is not an agent. But Tomlin's study did not examine whether passives are harder to produce even when they are licensed by the discourse. To answer this question, it is necessary to measure processing load rather than just frequency of occurrence.

Tannenbaum and Williams (1968) conducted one relevant study. Speakers first read a story that was either about trains, cars, or a topic that was relatively neutral. They then saw a picture of a train hitting a car, and their task was to produce either an active or a passive sentence (cued by a letter superimposed on the picture). They found that latencies to produce the active were fastest in the subject-focus condition and about equally long in the object-focus and neutral conditions. Passives were produced fastest in the object-focus condition, next fastest in the neutral condition, and most slowly in the subject-focus condition. This finding would appear to suggest that as long as a construction occurs in the appropriate context, it is easy to produce. However, a closer examination of their data suggest a different conclusion. Although this pattern was observed, it was also found that actives in the "wrong" discourse were produced as quickly as passives in the "right" discourse; indeed, in no condition were passives initiated faster than actives. The picture that emerges, then, is that noncanonical structures can be inherently hard to say, even in proper contexts. A construction that is rare or that is syntactically complex (or both, as these two characteristics tend to co-occur) requires a specific sort of context but is still difficult to generate, perhaps because more syntactic nodes take more processing resources to create, or because the production system has less experience generating forms such as the passive.

This finding is compatible with a study that investigated whether certain verbs license passives more easily than others, as might be expected on a lexicalist view of grammatical encoding (F. Ferreira, 1994). Participants were asked to generate sentences out of three visually presented words – e.g., LAYOFFS MANAGER WORRIED. The verb either had a theme-experiencer argument structure (as in *worried*) or a more conventional agent-patient structure (LAYOFFS MANAGER ORDERED). Speakers produced passives more often when the verb was theme-experiencer, which was predicted based on the idea that speakers attempt to place the more prominent thematic role in the subject position of the sentence, and experiencers are more prominent than themes (Grimshaw, 1990). Nevertheless, passives took longer to formulate than actives, suggesting that even though certain lexical conditions might license them, they still seem to take more time to grammatically encode.

Clearly, however, this issue needs to be examined in more detail, particularly now that there is such intense interest in the idea that the frequency of exposure to a syntactic construction affects how easily it can be comprehended (MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell & Cuetos, 1991). As Race and MacDonald (2003) have pointed out, these distributional patterns come from speakers – they reflect the choices speakers make in different circumstances. The Race and MacDonald approach to processing assumes that comprehension and production must be examined together, and they



predict that the forms that are hard for people to produce are also the ones that are hard to understand. This parity is based on speakers' tendency to avoid difficult structures, thus creating distributional patterns. But this interesting research program is predicated on the idea that some forms are inherently difficult to produce – for example, an object relative such as *The story the quiet boy read was long*, which does not contain the relative pronoun *that* – and that is why it is less likely to be said. The empirical question that arises is whether these sentences are harder to say when they are generated, or whether the discourse conditions, which obtained at the time the sentence was said in fact made the structure easy to encode and articulate.

### 3.3. Incremental Production

Incremental production may be viewed as a way to reduce the processing resources required for production. The idea is that at particular points in time certain concepts may be more available to a speaker than others, and the grammatical encoder tends to begin with those accessible lemmas. Incrementality is viewed as optimizing the use of processing resources, because it allows the system to begin with the 'easy bits', so to speak, and to deal with the more difficult portions of the utterance during articulation (F. Ferreira & Henderson, 1998). Of course, as a reviewer of an earlier version of this chapter pointed out, incrementality might create a situation in which an accessible constituent forces a syntactic structure that is computationally demanding (e.g., the passive). But the reason incrementality will generally still lead to efficient processing is that the difficulty of making a passive can be 'spread out' over the entire utterance rather than being localized entirely to the point of its initiation. As a result, there need not be any hesitation or disfluency before utterance production, and the demands of managing the rest of the structure can be distributed over the remaining constituents, with planning going on in parallel with articulation.

Recent work suggests that the degree to which the system is incremental is under strategic control, as would be expected if incrementality is a way for the production system to manage its resources (F. Ferreira & Swets, 2002). Participants were asked to calculate the answers to arithmetic problems and to provide the answer in the form of a sentence (*The answer is 58*). The problems always included at least one two-digit addend (e.g.,  $53+5$ ), so participants were unlikely to be able to retrieve the sum. In the first experiment, participants were allowed to begin to speak whenever they felt ready, and the data provided no evidence for incremental production. Initiation times were longer the more difficult the entire problem, but durations were unaffected. This pattern indicates that the entire utterance was planned before articulation. In the second experiment, speakers were required to begin to speak before a deadline (indicated by a punishment 'beep'). This manipulation dramatically reduced initiation times overall, from over 2 s in the first experiment to about 700 ms in the experiment with the deadline (interestingly, accuracy was not compromised). Nevertheless, initiation times still reflected the difficulty of computing the sum. At the same time, the duration of the earlier part of the utterance was also affected by problem difficulty, suggesting that speakers postponed some planning of the sum until they were actually speaking. This study suggests that the degree to which the

system is incremental depends on the speaker's strategy. If a premium is placed on beginning to speak quickly, then the production system does indeed become more incremental; but if speakers have the opportunity to plan, they seem to prefer to do so. Moreover, the system engages in some planning even under conditions most conducive to incremental production – that is, when there is a premium on initiating speech quickly.

On some views of incrementality, constructions are chosen indirectly; they emerge from the speaker's attempt to place a highly accessible concept in the most prominent syntactic position. If that concept happens to be a theme or patient, in a relatively fixed word order language such as English a passive structure will need to be produced to accommodate a thematic patient in subject position. Under this view of human sentence production, the lemmas associated with the most accessible concepts automatically grab the earliest positions in utterances. However, Bock (1986b) questioned this radical version of incrementality: "Typically, speakers do not simply produce words in the order in which they come to mind . . . . Rather, the syntactic forms of sentences seem to be changed so as to accommodate word order variations without altering the intended meaning" (p. 359). But a radically incremental model is assumed, for example, by van Nice and Dietrich (2003), who interpret their German data as supporting the view that "the first-conceptualized referent will continue onward as the first-lexicalized and, ultimately, as the first in word order" (pp. 829). This view, they point out, is also held by Kempen and Hoenkamp (1987) as well as de Smedt (1996).

Christianson and F. Ferreira (2005) attempted to resolve this controversy by examining production in Odawa, an Algonquin language which allows constituents to be ordered freely (i.e., any arrangement of subject, verb, and object is grammatically licensed). Speakers were asked questions about a pictured event. The questions topicalized either the agent, the patient, or neither (this latter question was simply, *What happened?*). Even though Odawa speakers have access to any word order arrangement of subject, verb, and object, their descriptions were similar to those observed for English speakers. Given the no-topic and agent-topicalizing questions, actives were the forms most commonly produced; in the patient-topicalizing question condition, passives were preferred. Thus, even though speakers of Odawa could have produced active sentences with the patient in the first position (i.e., OSV or OVS sentences) when the patient was topicalized, they in fact chose to produce passives, which not only are about as rare in Odawa as they are in English, but also require the omission of the agent argument altogether (because passives in Odawa do not permit any type of by-phrase). Thus, a highly available constituent primes a particular syntactic form, and if that constituent is a patient, the form that will be generated is a passive.

These findings are inconsistent with extreme versions of incremental production, and instead support V. Ferreira and Dell (2000), who argued that the lexically driven picture of production – in which the most accessible lexical item wins a figurative "race" out of the mouth – might not be sufficient to accurately describe their results. Instead, they proposed that speakers choose a syntactic structure without necessarily first deciding between alternative lexical items. The structure, then, is what is really primed by pictures,

sentences, and questions. On this view, incrementality applies to the filling of available NP nodes in the primed structure (F. Ferreira, 2000). Incrementality encourages the selection of a syntactic structure that allows accessible material to be mentioned sooner (V. Ferreira & Dell refer to this as ‘lexical-syntactic interactionism’).

Another way to think about incrementality in production is to ask what sorts of planning units the system uses. The most extreme versions of incrementality assume that there is little or no look-ahead, and so predict that planning units will be essentially non-existent (i.e., utterances are planned more or less word-by-word). On the other hand, non-incremental views assume that the system does engage in look-ahead over some multi-word domain. Clauses have been classically assumed to serve as planning units for grammatical encoding (Boomer, 1965; Ford & Holmes, 1978; Garrett, 1975; Lashley, 1951). The idea is that the system organizes an entire clause (i.e., a verb and its arguments) before engaging in any phonological encoding.

One way that this issue has been addressed is by examining how speakers compute grammatical agreement between complex subjects and inflected verbs. A variety of studies have demonstrated that a distractor noun in the subject can be an attractive lure for agreement, especially if it is plural (Bock & Eberhard, 1993; Bock & Miller, 1991). Bock and Cutting (1992) used this phenomenon to determine whether the unit of grammatical encoding is the clause. They varied whether the constituent that intervened between the head noun of the subject and the main verb was a PP modifier or a relative clause (e.g., *the editor of the history books* versus *the editor who rejected the books*). They reasoned that if clauses are planning units for grammatical encoding, then agreement errors (plural inflections on a form of *to be*) should be less common in the relative clause condition. This prediction follows because the relative clause would be planned separately, and thus the head noun and the verb would be more closely linked during processing. This prediction was confirmed: Agreement errors were more common when a relative clause came between the head noun of the subject and the main verb, consistent with the classic idea that the unit of syntactic planning is the clause. This finding is inconsistent with radical incrementality or any type of production system which generates utterances on a word-by-word basis, but it can be reconciled with more limited incrementality (Christianson & F. Ferreira, in press; F. Ferreira, 2000).

Finally, there is evidence that the production system operates more efficiently when it has syntactic options that allow potentially different states of activation to be taken into account during grammatical encoding. V. Ferreira (1996) compared the production of sentences headed by a verb such as *give*, which alternates between a double-object and a prepositional dative form, and verbs such as *donate*, which only allow the prepositional dative (e.g., *\*The widow donated the library her entire collection*). He found that sentences with syntactically flexible verbs such as *give* were generated more quickly and more fluently than sentences with more restrictive verbs. He argued that flexibility allows the system to accommodate lemmas’ potentially different states of activation over time. For example, if a speaker has said *The widow gave* and then finds that direct object hard to retrieve, he or she can continue processing by working on the indirect object instead,

because the verb *give* permits this flexibility. Thus, one benefit of syntactic freedom of choice is that it enhances the efficiency of language production. In the next section, we focus specifically on the issue of how speakers make syntactic choices.

## 4. SYNTACTIC CHOICE

### 4.1. Choice of Syntactic Construction

As already mentioned, work by Carroll (1958), Bock (1986a, b), Bock and Warren (1985), and others has shown that, in English, when a noun phrase is made accessible by showing someone a picture of a semantically related item, asking a focusing question, or establishing a context, speakers tend to begin their sentences with that primed NP. Bock and Warren's (1985) work on the production of passives and dative structures in English indicates that the most accessible entity claims not only an early position in the string, but also the most prominent syntactic function (i.e., subject or non-oblique dative in ditransitive structures). A similar finding is that passives tend to occur with theme-experiencer verbs, because the passive allows the experiencer to be placed in subject position (F. Ferreira, 1994). This effect is larger when the experiencer is human and the theme is not, indicating that an animacy contrast perhaps helped to distinguish the conceptual prominence of the two entities even more than just their thematic role status. Spanish speakers also are sensitive to accessibility when they choose syntactic constructions (Prat-Sala & Branigan, 2000). Spanish syntax includes a dislocated active structure (OSV, along the lines of 'Cheese I love to eat'), which allows the effects of inherent accessibility (animacy) to be distinguished from those of derived accessibility (discourse prominence). Spanish speakers tended to place the more salient entities in higher syntactic positions, making use of both passives and the dislocated active structure. In general, then, syntactic forms are chosen to allow speakers to line up conceptual and syntactic prominence.

### 4.2. Inclusion of Optional Functional Elements

Thus far, we have considered how speakers decide on a syntactic form – active versus passive, double-object versus prepositional dative, and so on. Now we ask a slightly different question: How do speakers decide whether to include an optional function word such as the complementizer *that* in a sentence like *The weary traveler claimed (that) his luggage had been stolen?* If the complementizer is omitted, an ambiguity about the status of the noun phrase *the luggage* is created for the comprehender: *the luggage* could be either the direct object of *claimed* or the subject of a complement clause. The presence of the complementizer essentially disambiguates the structure, making it clear that the noun phrase is a subject. (It is possible for *that* to be a determiner, as in *that luggage*, not *your luggage*, but Roland, Elman, and V. Ferreira (in press) have demonstrated that post-verbally, the word *that* is almost always a complementizer, and the parser is likely tuned to this distributional information.) If speakers attempt to produce utterances that are easy for their listeners to understand, one might predict that, the greater the chance of a misinterpretation, the greater the likelihood that speakers will include the complementizer.

For example, if the verb preceding the ambiguous noun phrase subcategorized for only clausal complements, the *that* is unnecessary, and so it might be omitted; but if the verb takes both direct objects and clausal complements, the *that* would help the listener avoid making a parsing error.

The evidence suggests that speakers' needs motivate complementizer inclusion. This has been shown in a variety of experiments by V. Ferreira (2003; V. Ferreira & Dell, 2000), which demonstrate that *that* is more likely to be included in complement and relative clause structures when the speaker is having difficulty retrieving the word that would follow *that*. Two different mechanisms can be proposed to account for this relationship: Alleviation and Signaling (Jaeger, 2005). According to the Alleviation hypothesis, speakers include *that* to give themselves time to plan (Race & MacDonald, 2003), making the complementizer essentially like a filler term such as *uh*. The alternative hypothesis, Signaling, assumes that the complementizer is a signal or at least a symptom of upcoming difficulty. The two hypotheses make opposite predictions about the distribution of complementizers and filler disfluencies. If Alleviation is right, then the presence of a complementizer should reduce the likelihood of a filler. If Signaling is correct, then *thats* and fillers should be positively correlated. Jaeger (2005) and V. Ferreira and Firato (2002) found results consistent with the second pattern, which supports the Signaling hypothesis. It is important to note, however, that the data are compatible with the idea that complementizer inclusion is merely a symptom of difficulty – that is, the same factors that lead to disfluencies lead to the inclusion of a complementizer as well.

This pattern has emerged in other studies as well—speakers in dialogue tasks fail to make use of either optional words or disambiguating prosody to avoid ambiguity (Allbritton, McKoon, & Ratcliff, 1996; Arnold, Wasow, Asudeh, & Alrenga, 2004; Kraljic & Brennan, 2005). One exception is a recent study reported by Haywood, Pickering, and Branigan (2005), who found that speakers did provide more disambiguating *that's* when they were describing objects to a conversational partner. However, as V. Ferreira, Slevc, and Rogers (2005) argue, the effect may be due to the visual properties of the situation the interlocutors were presented with. The situation which led to ambiguity in the Haywood et al. study was one in which there was more than one object of the same type, thus inviting the use of a disambiguating modifier (e.g., *the penguin THAT'S in the cup on the star*), and it is in these situations that the word *that* tended to be included. Ferreira et al. argue that perhaps speakers were simply sensitive to the existence of more than one token of the same type and in those cases produced more explicit utterances. At the same time, it must be acknowledged that even if this interpretation is correct, it still appears that Haywood et al. have indeed observed the altruistic rather than egocentric use of optional functional elements. Moreover, as highly skilled speakers are probably better able to avoid ambiguity than those who are less practiced, it is clear that some mechanisms must exist to allow speakers to monitor their speech and include optional elements in just those situations when they might be helpful to listeners. What we do not know is how this process, which is potentially quite resource-demanding, is coordinated with the other tasks performed by the production system.

## 5. SYNTAX AND PROSODY IN PRODUCTION

Utterances have regular rhythmic and tonal properties, and these prosodic features of a sentence are influenced directly and indirectly by syntactic structure. Rhythm results from the way syllable stress and duration change over the course of an utterance, and intonation is the result of changes in fundamental frequency (F0) or tone (see F. Ferreira, 1993, for a description of what constitutes prosody). Consider the sentence *If she goes, I go too*. It would likely be spoken as two intonational phrases separated at the boundary indicated with a comma. The word *goes* would tend to have a long duration and would receive greater stress than it would in a non-clause-final position, and the same effects would be observed on *too*, perhaps even more markedly, because *too* ends not only as a clause but also an entire sentence.

It is clear that utterance stress and timing have something to do with a sentence's syntactic structure. Phonologists have debated whether the correct characterization of these effects appeals directly to syntactic constituents, or instead makes reference to prosodic entities such as phonological words, phrases, and intonational phrases. According to the syntactic view (Cooper & Paccia-Cooper, 1980; Odden, 1990; Selkirk, 1984; Wagner, 2005), the amount of lengthening and stress assigned to a given word can be directly related to syntax. For example, the more syntactic right brackets that terminate on a word, the longer and more stressed it will tend to be. On the prosodic constituency view, syntax is used to create prosodic constituents, but then it is features of prosodic constituency that determine timing and stress (Gee & Grosjean, 1983; Inkelas & Zec, 1990; Levelt, 1989; Selkirk, 1986). Disentangling these two approaches to rhythm can be challenging, because prosodic and syntactic constituency are highly correlated (F. Ferreira, 1993), but one important theoretical difference between them is that prosodic structure is generally viewed as flatter and less articulated than syntactic structure, because prosodic constituency is generally thought not to permit recursion (Selkirk, 1986; cf. Gee & Grosjean, 1983; Ladd, 1986; Wagner, 2005). The idea is that, in syntax, a clause may have another clause inside it (for example), but in prosody, such self-embedding is forbidden. As a result, prosodic structures are flatter than syntactic ones, allowing prosody to serve as an interface between hierarchical and recursive syntactic/semantic representations and the sequential speech channel through which articulation must take place. Another important difference between the two types of structures is that prosodic representations pay attention to the distinction between function and content words. Therefore, a phrase consisting of just a pronoun, for instance, would typically not behave the same way as a full lexical NP.

Intonation is related to syntax too, but again, it has long been known that the intonational phrasing of a sentence may not be isomorphic to its syntactic constituency. One famous example is *This is the cat that chased the rat that swallowed the cheese...*, which tends to be phrased as *(this is the cat) (that chased the rat) (that swallowed the cheese)*, even though the major syntactic boundary is between *is* and *the cat*. Other more realistic examples include *(Mary left)(after the party)* and *(Mary gave the book) (to her brother who lives in Ohio)*. In both these cases, the major intonational boundary comes not between the subject and verb phrase, but after the verb. To account for these cases, Selkirk

(1984) proposed the Sense Unit Condition, which states that the constituents making up a single intonational phrase must be in either a head-argument or head-modifier relation. The Sense Unit Condition rules out apparently malformed examples such as (*Ten mathematicians*) (*in ten derive a lemma*) (Steedman, 2000), because *in ten* is a modifier of *mathematicians* and is not a head, argument, or modifier of *derive a lemma*. Steedman (2000) argues that a grammar such as Combinatory Categorical Grammar, which allows a wider range of syntactic constituents than other approaches captures these sorts of facts and eliminates the need for a separate and stipulative Sense Unit Condition. The important point for our purposes, however, is that even though the intonational phrasing of a sentence might ultimately deviate from its syntactic structure, the well-formedness of the intonational phrasing appeals to syntactic concepts such as head, modifier, and argument.

So far we have considered only aspects of prosody that can be directly related to linguistic structures, either prosodic or syntactic. But syntax affects the sound properties of a sentence in another way, which we will roughly characterize as having to do with performance effects. For example, hesitations and pauses due to planning difficulty tend to cluster at clause boundaries (Ford, 1982; Goldman-Eisler, 1968). In addition, it has been argued that the most probable location for pauses and prosodic breaks can be predicted from algorithms, which assume that break points are jointly determined by the complexity of material to the left and to the right of the boundary (Gee & Grosjean, 1983; Watson & Gibson, 2004). F. Ferreira (1991) demonstrated that the syntactic complexity of upcoming material affected pause duration, and she argued that the effects were due to the difficulty of planning upcoming material. Thus, the sound pattern of a sentence has at least two possibly distinct sources: One is the syntactic and prosodic representation which might mandate breaks in particular locations, and the other is the speaker's need for more time to plan upcoming material.

An important question for future research on the syntax-prosody interface in language production is whether these two sources are indeed distinct, or whether prosody and performance phenomena can be reduced to the same underlying causes. Another critical issue is whether it is necessary to postulate a distinct level of prosodic constituency to account for phenomena related to rhythm and intonational phrasing, or whether syntactic structure is sufficient to explain prosodic patterns in spoken sentences. One limitation of work that has been conducted up to this point is that virtually all studies investigating prosody in production have used simple reading or repetition tasks to elicit utterances. The reason is that in order to test prosodic and syntactic hypotheses adequately, it is necessary to precisely control what the speaker says. But unless more naturalistic tasks are used that allow speakers to talk relatively normally, it will be impossible to assess to what extent the need to plan affects the sound features of a sentence, and to evaluate how incrementality in production affects the distribution of hesitations, pauses, and even intonational boundaries across an utterance.

## 6. CONCLUSIONS

What is currently known about the process of grammatical encoding indicates that the syntactic structures used in language production have the following characteristics.

First, they are generated in two separate stages, the first one creating a representation that represents hierarchical relations but not necessarily linear order, and a second stage in which linearization within phrases takes place. Second, the structures for creating both the global form of the entire utterance and the form of the individual phrases are generated from trees anchored to specific lexical heads. Third, there is some evidence that syntactic representations contain gaps or traces. Admittedly this is a point on which there is little consensus and almost no data, but recent evidence about the computation of subject-verb agreement (Franck et al., in press) as well as data concerning the blocking of function word reduction following a gap suggests that gaps are in fact mentally represented at some stage in production. Moreover, all of these features of grammatical encoding can be captured using TAG as the representational format for syntactic information (F. Ferreira, 2000), which again assumes lexical generation of structure. The main verb of an utterance provides an overall clausal template constrained by the verb's phrase-taking properties, and then each specific phrase is fleshed out and attached as its head (e.g., a noun for a noun phrase) is accessed. TAG represents gaps not via movement but as part of the treelet anchored to the lexical item, thus explaining phenomena such as the blocking of function word reduction in the hypothesized vicinity of a gap.

In addition, although the concept of processing resources is somewhat vague (as MacDonald & Christiansen, 2002 argue), we can resort to an operational definition and say that processing resources are what measures such as initiation time and pause probability/duration reflect. With this assumption, we conclude that grammatical encoding requires resources, and that some constructions appear to be difficult to generate even in felicitous contexts. In addition, the bulk of the evidence indicates that production is incremental in the sense that the most accessible concept will tend to capture the syntactically most prominent position in a functional level structure. This tendency toward moderate incrementality reduces the computational burden on the grammatical encoder because the system can begin with what is already accessible and wait for other elements to become available as processing unfolds. In addition, if it indeed turns out that syntactic priming is particularly robust in dialogue because it makes the task of generating a syntactic structure easier (Pickering & Garrod, 2004; Pickering & Branigan, 1999), then we have further evidence that grammatical encoding requires significant processing resources.

Recall that the original argument against this idea was that syntactic processing was assumed to be modular (Levelt, 1989), and one of the characteristics of a module is that it operates automatically (J. A. Fodor, 1983). Do our conclusions undermine this assumption? Not necessarily. Fodor's conception of automaticity appears to have more to do with whether a person's conscious goals and intentions can influence processing than with whether the process is computationally costly. Moreover, it is clear that many specialized systems call upon working memory resources, and one point of debate has been whether the working memory that is involved is domain-general or entirely devoted to just that module. Thus, a system might be modular but still draw on working memory, and the resource pool that is used could itself be modular, in the sense that it is dedicated to processing in that one domain. These are topics for further investigation.



We also conclude that syntactic choices are made largely for the benefit of the speaker. The decision about what syntactic construction to use is at least in part based on the accessibility of the lemmas that will comprise the utterances. Optional function words such as complementizers are left out when the speaker has time to retrieve the immediately following word but included when retrieval is slow and difficult. There is also evidence that speakers obey the Gricean Maxim of Quantity only to a limited extent (Grice, 1975), in part because they have a tendency to describe objects in whatever way is salient to them, neglecting to take into account the effect the description might have on the comprehender (Engelhardt, Bailey, & Ferreira, in press). For example, even though the relevant discourse might include just a single hat, studies show that more than 25% of the time, the hat will be described as *the red hat* or *the hat with the feather*. These over-descriptions likely occur because, from the speaker's point of view, the object IS a red hat or a hat with a feather. Because those features of the object are salient, they have a good chance of making it into the message-level representation. In these situations, it would require extra effort for the speaker to produce concise descriptions, because he or she would have to remove content to make sure that information did not get grammatically encoded.

Finally, the syntactic structure of a sentence affects the way it is spoken. For example, the presence of a gap in surface structure affects whether a preceding word is reduced or lengthened. More generally, syntax has profound effects on all aspects of prosody, including the duration and stress level of words, the location and duration of pauses, and the intonational tune and phrasing of the sentence. An unresolved question is whether syntax is directly responsible for these effects, or whether they are mediated through prosodic constituency. Another is how linguistic structure and performance limitations play off of each other to help establish a sentence's overall prosodic form. In addition, it is still not clear how lemma retrieval, word-form activation, and functional and positional level processing are coordinated with the tasks of creating prosodic constituents, generating intonational contours, and implementing a phonetic plan (F. Ferreira, 1993). Moreover, essentially the same questions can be asked about prosody that we considered with respect to syntax in the present review. What sorts of computational resources does the process of creating prosodic representation draw upon, and how do speakers manage and even take advantage of optionality in prosody (for some discussion, see Steedman, 2000; Watson & Gibson, 2004). Unfortunately, although we can ask these questions, there is still not enough evidence to allow us to provide even tentative answers. Ultimately, a complete understanding of syntax in production will require consideration of issues relating to prosody. We therefore hope that the next decade will see an integration of research on syntax, prosody, and language production.

## REFERENCES

- Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 714–735.

- Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*, 55–70.
- Bock, J. K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575–586.
- Bock, J. K. (1986b). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.
- Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, *26*, 119–137.
- Bock, J. K. (1989). Closed-class immanence in sentence production. *Cognition*, *31*, 163–186.
- Bock, J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*, 99–127.
- Bock, J. K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, *8*, 57–99.
- Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In: M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.
- Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*, 1–39.
- Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*, 150–171.
- Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.
- Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47–67.
- Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148–158.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–B25.
- Bresnan, J., & Kaplan, R. (1985). Introduction: Grammars as mental representations of language. In: J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 17–52). Cambridge, MA: MIT Press.
- Carroll, J. B. (1958). Communication theory, linguistics, and psycholinguistics. *Review of Educational Research*, *28*, 79–88.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, 98, 105–135.
- Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49, 214–230.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Cooper, W. E., & Ross, J. R. (1975). World Order. In: R. E. Grossman, L. J. San, & T. J. Vance (Eds), *Papers from the parasession on functionalism* (pp. 63–111). Chicago: Linguistic Society.
- de Smedt, K. (1996). Computational models of incremental grammatical encoding. In: A. Dijkstra, & K. de Smedt (Eds), *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 279–307). London: Taylor & Francis.
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity. *Journal of Memory and Language*, 54, 554–573.
- Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion*. Unpublished Doctoral Dissertation, University of Massachusetts, Amherst, MA.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210–233.
- Ferreira, F. (1993). The creation of prosody during sentence processing. *Psychological Review*, 100, 233–253.
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33, 715–736.
- Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In: L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). London: Psychology Press.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentence. *Cognitive Psychology*, 47, 164–203.
- Ferreira, F., & Henderson, J. M. (1998). Linearization strategies during language production. *Memory & Cognition*, 26, 88–96.
- Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, 28, 721–749.
- Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57–84.

- Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35, 724–755.
- Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying “that” is not saying “that” at all. *Journal of Memory and Language*, 48, 379–398.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40, 296–340.
- Ferreira, V. S., & Firato, C.E. (2002). Proactive interference effects on sentence production. *Psychonomic Bulletin & Review*, 9, 795–800.
- Ferreira, V. S., Slevc, L. R., & Rogers, E. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96, 263–284.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 11, 759–770.
- Fodor, J.D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes*, 4, 155–209.
- Fodor, J. D. (1991). Sentence processing and the mental grammar. In: P. Sells, S. Sheiber, & T. Wasow, (Eds), *Foundational issues in natural language processing* (pp. 83–114). Cambridge, MA: MIT Press.
- Ford, M. (1982). Sentence planning units: Implications for the speaker’s representation of meaningful relations underlying sentences. In: J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 798–827). Cambridge, MA: MIT Press.
- Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, 6, 35–53.
- Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, 101, 173–216.
- Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject–verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, 17, 371–404.
- Garrett, M. F. (1975). The analysis of sentence production. In: G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.
- Garrett, M. F. (1980). Levels of processing in sentence production. In: B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–220). London: Academic Press.
- Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411–458.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, 35, 623–647.

Grice, P. (1975). Logic and conversation. In: P. Cole, & J. Morgan (Eds), *Syntax and semantics: Speech acts* (Vol. III, pp. 41–58). New York: Academic Press.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975). *Journal of Memory and Language*, 52, 58–70.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, 75, B27–B39.

Hawkins, J. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, 16, 362–366.

Inkelas, S., & Zec, D. (1990). *The phonology-syntax connection*. Chicago: University of Chicago Press.

Jackendoff, R. (1977). *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Jaeger, T. F. (2005). Optional *that* indicates production difficulty: Evidence from disfluencies. *Proceedings of DiSS'05, disfluency in spontaneous speech workshop. 10–12 September 2005, Aix-en-Provence, France* (pp. 103–109).

Johnson, N. F. (1966). On the relationship between sentence structure and the latency in generating the sentence. *Journal of Verbal Learning and Verbal Behavior*, 5, 375–380.

Joshi, A. K. (1985). How much context-sensitivity is required to provide reasonable structural descriptions: Tree adjoining grammars. In: D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 206–250). Cambridge: Cambridge University Press.

Joshi, A. K., Levy, L., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, 10, 136–163.

Kempen, G., & Hoenkamp, E. (1982). Incremental sentence generation: Implications for the structure of a syntactic processor. In: J. Horecky (Ed.), *Proceedings of the 9th international conference on computational linguistics*. Amsterdam, North-Holland.

- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.
- Kraljic, T., & Brennan S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee. *Cognitive Psychology*, *50*, 194–231.
- Ladd, D. R. (1986). Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook*, *3*, 311–340.
- Lashley, K. S. (1951). The problem of serial order in behavior. In: L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*, 78–106.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review*, *101*, 676–703.
- MacWhinney, B., Bates, E., & Reinhold, K. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.
- Mitchell, D. C., & Cuertos, F. (1991). The origins of parsing strategies. In: C. Smith (Ed.), *Current issues in natural language processing* (pp. 1–12). Austin, TX: University of Texas, Center for Cognitive Science.
- Moortgat, M. (1988). *Categorial investigations: Logical and linguistic aspects of the Lambek Calculus*. Dordrecht: Foris.
- Odden, D. (1990). Syntax, lexical rules and postlexical rules in Kimatummbi. In: S. Inkelas, & D. Zec (Eds), *The phonology-syntax connection* (pp. 259–278). Chicago: University of Chicago Press.
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*, 633–651.
- Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*, 136–141.
- Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory and Language*, *46*, 586–605.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic theory of dialogue. *Behavioral & Brain Sciences*, *27*, 169–226.
- Pinker, S., & Birdsong, D. (1979). Speaker's sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior*, *18*, 497–508.

- Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*, 168–182.
- Race, D. S., & MacDonald, M. C. (2003). The use of “that” in the production and comprehension of object relative clauses. *Proceedings of the 25th annual meeting of the cognitive science society*. Boston, Massachusetts.
- Rochester, S. R., & Gill, J. (1973). Production of complex sentences in monologues and dialogues. *Journal of Verbal Learning and Verbal Behavior*, *12*, 203–210.
- Roland, D., Elman, J. L., & Ferreira, V. S. (2005). Why is *that*? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*, *98*, 245–272.
- Schenkein, J. (1980). A taxonomy for repeating action sequences in natural conversation. In: B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 21–47). London: Academic Press.
- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Selkirk, E. O. (1986). On the derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405.
- Shapiro, L. P., Hestvik, A., Lesan, L., & Garcia, A. R. (2003). Charting the time-course of VP-ellipsis sentence comprehension: Evidence for an initial and independent structural analysis. *Journal of Memory and Language*, *49*, 1–19.
- Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production – an online study. *Cognition*, *78*, 123–164.
- Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.
- Stowell, T. (1981). *Origins of phrase structure*. Unpublished doctoral dissertation, MIT.
- Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition*, *8*, 513–520.
- Tannenbaum, P. H., & Williams, F. (1968). Generation of active and passive sentences as a function of subject or object focus. *Journal of Verbal Learning and Verbal Behavior*, *7*, 246–250.
- Tomlin, R. S. (1983). On the interaction of syntactic subject, thematic information, and agent in English. *Journal of Pragmatics*, *7*, 411–432.
- van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture description. *Linguistics*, *41*, 825–849.
- Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, *68*, 13–29.

Wagner, M. (2005). Long distance effects on prosody. Paper presented at the 18th annual meeting of the CUNY conference on human sentence processing. Tucson, Arizona.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9, 81–105.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 713–755.

Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104, 444–466.



This page intentionally left blank

## Chapter 4

### Speech Disorders

Gary Weismer

#### 1. INTRODUCTION

A treatment of speech disorders in a handbook of psycholinguistics implies the relevance of the latter discipline to an understanding of the disorders. Although this relevance may seem self-evident to some readers – after all, psycholinguistics concerns the processes underlying the production, perception, and comprehension of language, of which speech is an uncontroversial component – there have been substantial controversies concerning the *explanatory* role of psycholinguistics in speech disorders. By *explanation* we mean, for this essay, an account of the root or cause of a speech disorder. For example, an English-speaking, post-stroke patient who on perceptual analysis produces an apparent speech error of replacing a voiceless stop consonant (such as p, t, or k) with its voiced cognate (b, d, or g) and whose voice-onset time (VOT) for the error is in the short-lag range (say, less than 20 ms) rather than the target-appropriate, long-lag range (greater than 30 ms) might be said to produce a phonemic, not a phonetic error. On this view, the patient is making an error of selecting the wrong phonological unit, not of misarticulating a correctly selected unit.

This explanation depends on assumptions concerning psycholinguistic processes that link phonological units to their phonetic implementation: stop consonants specified as – voice (i.e., phonologically voiceless) are implemented phonetically with long VOTs, and stop consonants specified as +voice are implemented with short VOTs. When a short VOT is produced for a – voice stop, an inversion of the assumed psycholinguistic process for this phonetic implementation rule leads to the explanation of an incorrectly selected phonological unit.

A potential problem with this line of reasoning – an old one in scientific discourse – is that of confusing description with explanation (see McReynolds, 1988, pp. 420–421 for relevant discussion). This is an especially dangerous, potential trap when dealing with

linguistic formalisms and their use in older-style, black box models of psychological processes or newer style, neural network concepts. Simply because a model can be made to mimic behaviors consistent with a particular theory does not prove the theory or any of its individual, explanatory axioms. Other box models or computational methods consistent with opposing theories may be able to mimic the same behaviors; *ad hoc* criteria are then required to choose between the competing theories, never a felicitous state of affairs for scientific advancement. The box and/or computational models are only descriptions, which may at some later point in time and with converging sources of evidence become part of an explanation.

This issue has direct relevance to psycholinguistic accounts of speech disorders, because so much of the discussion concerning the causes of specific speech disorders depends on descriptions derived from linguistic formalisms, and descriptive devices (such as the box or computational models alluded to above) concerning the processes that manipulate linguistic symbols within the formalisms. In this chapter we will review some psycholinguistic approaches to understanding speech disorders, with a selected focus on developmental speech delay and speech production in persons with neurological disease. Other speech disorders (such as speech of persons with cleft palate, hearing impairment, and fluency problems, as well-developmental apraxia of speech) have also been conceptualized in psycholinguistic terms, but the two disorders reviewed here provide excellent case studies of the benefits and dangers of the psycholinguistic perspective.

## 2. SPEECH DELAY

If there is a speech disorder that has been most influenced by psycholinguistic models, it is certainly the group of developmental disorders referred to as speech delay. Simply put, a child with speech delay is one who produces segmental errors that are not age appropriate, but are not unusual in the typical progress of speech development. Thus, an otherwise typically developing child aged six who produces a profile of speech sound errors consistent with a typically developing child aged three and a half years would be said to have speech delay. A child who produces errors considered to be atypical for the normal sequence of speech development – for example, a child who consistently omits word-initial obstruents – would not fall into the category of speech delay, at least as conceptualized by most clinicians and scientists working in the area.<sup>1</sup>

We refer above to a *group* of disorders falling under the rubric of speech delay because, depending on the classification system, different types of delay qualify for the classification. For example, the well-known case of an older child who has difficulty with a small set of sounds typically mastered late in languages that use them contrastively would clearly be categorized as having speech delay, specifically one involving so-called

---

<sup>1</sup> This kind of child, one who produces errors not typically seen in the course of normal speech sound development, may be the type who would be diagnosed with developmental apraxia of speech. See Hall, Jordan, and Robin (1993) for a complete review of the criteria for this diagnosis.

*residual errors*. Relatively consistent errors for sounds such as /r/, /l/, and /s/, singly or as a group, by children who are older than about six years of age would be considered a speech delay characterized by residual errors.<sup>2</sup>

Residual errors are usually thought to be the consequence of a delay in speech motor maturation. The sounds mentioned above qualify for the residual error label because they are assumed to require the most complex articulatory skill (hence their relatively late developmental age). The speech motor development explanation for residual errors has, for the most part, ruled out any serious attempts to develop psycholinguistic accounts of this form of speech delay.<sup>3</sup>

There are other kinds of speech sound error observed in the course of typical speech sound development but which are somewhat less amenable to a speech motor explanation when they are part of a speech delay profile. Such errors are therefore not categorized as residual errors. For example, during the course of typical speech sound development many children will produce stop consonants requiring a relatively back place of articulation in the vocal tract (such as /g/ and /k/) with a more forward place (such as /d/ and /t/). Thus a word like *kitty* may be produced as /tɪri/, a word like *go* as /do/. The opposite pattern may occur as well, but perhaps with less frequency: *tickle* may be /kɪkl/, *dog*, /gog/. An explanation of speech motor immaturity does not seem to make as much sense here because the sounds in question – stop consonants – are mastered fairly early in typical development, probably not much later than three years of age. These kinds of errors *have* been the subject of psycholinguistic theories, experiments, and analyses, as described below.

## 2.1. From Functional Misarticulation to Optimality Theory: A Brief Tour

Before terminology and perspectives from the discipline of psycholinguistics found their way into thinking about speech sound development problems, children with a broad range of articulation problems but no obvious, underlying causal basis for the delay were called *functional misarticulators* (see McReynolds, 1988, for an outstanding analysis of the use of this diagnostic term; see also Barlow & Gierut, 1999). The delay was viewed

---

<sup>2</sup> There are many sources for typical speech sound development, which differ in certain details depending on sampling method, language, and so forth; excellent reviews of this work and references to the extensive older literature on developmental norms can be found in Grunwell (1981, 1987), Smit, Hand, Frelinger, Bernthal & Bird (1990), Smit (1993), and Dodd (1995), among others.

<sup>3</sup> Again, there is a qualification to this claim that will not be pursued in detail in this chapter, but it should at least be mentioned that efforts to explain developmental speech sound errors on the basis of immature or faulty *perceptual* mechanisms have a long history in speech-language pathology (see McReynolds, 1988, pp. 422-424), but not one with much empirical success (see, for example, Locke, 1980a,b; Williams & McReynolds, 1975; Nittrouer, 1999). The psycholinguistic link is the possibility of immature perceptual mechanisms affecting a child's ability to form proper category features for the contrastive sounds of her or his language. In fact, at least one theoretical strand in the literature on normal speech sound acquisition (Best, 1994; Kuhl, 1993) works from this perspective, so it is easy to see how it could be extended to children with speech sound production errors: they produce errors because they have a bad representation of the sound category as a result of deficient perceptual and/or categorization skills. As noted here, efforts to link specific production error types with specific perceptual problems have not been successful.

as functional in the sense of a poorly learned or willfully immature sound system, as opposed to neurological disease, hearing impairment, or structural abnormality (such as cleft palate) causing the errors. *Functional misarticulation* was clearly a wastebasket term (much like the use of the term *idiopathic* in the literature on certain diseases, as in Idiopathic Parkinson's Disease), but it served the purpose of distinguishing speech-delayed children without obvious physical problems from those with a clear basis for their speech production problems. Most speech-language pathologists and scientists with interest in children labeled as functional misarticulators were uneasy, however, with the broadness of the label and its lack of prognostic guidance.

In particular, there was concern with variability in type and severity of symptoms of children diagnosed with functional misarticulation. Children with speech delay might present in the clinic with just a few errors, or many errors, and some of these errors may occur on some words but not others and at some times but not others for the same words (see below). In the early history of speech delay, there was a tendency for these errors to be regarded as separate entities; theoretical and practical links between the different kinds of errors and their conditions of occurrence were not recognized. Eventually there was recognition that developmental and clinical errors often grouped into sound classes (e.g., fricative errors), and as phonological theory advanced and accounted for more regularities and systematicities in the sound structure of languages, so did the possibility of applying these notions to an explanation of speech sound errors. Moreover, as phonological theory and cognitive psychology began to see their reflections in each others' gaze, psycholinguistics entered the realm of functional misarticulation where it has played a major role, in different ways, at least since the late 1960's and early 1970's. This history is nicely reviewed by Gierut and Morrisette (2005).

One of the primary interests in the literature on speech delay has been the relationship between an observed error and the form and/or processes underlying its production. With respect to underlying forms (or the 'input units' of optimality theory (OT): see below), most investigators and clinicians have assumed them to be no different from those in the normal, ambient phonology. Thus the child who says [tIti] for *kitty* or [kul] for *school* is generating her or his word productions from the representations /kIti/ and /skul/. This assumption is largely one of common sense, but also may have been thought to enjoy indirect empirical support from the lack of apparent perceptual problems – a near-normal ability to discriminate between phonemes – among children with speech delay (see footnote 3, and discussion in Hewlett & Waters, 2004). In the typically developing infant such phoneme-discrimination abilities have been demonstrated for specific cases as early as one month of age and for a broad range of phonetic events before six months of age (see Houston, 2005, for a review), so the presence of an intact perceptual phonology in children with speech delay may lead to a very simple conclusion: the majority of sound errors produced by children with speech delay are phonetic in nature, probably a result of articulatory immaturity and therefore best described using phonetic, not phonological methods (see a form of this argument in Hewlett & Waters, 2004, and discussion of phonetic pressures on phonological description in Locke, 1983). However, claims including a lack of isomorphism between the perceptual and productive phonologies (Menn, 1980;

Straight, 1980) as well as the success of certain phonological analyses in accounting for regularities in *groupings* of sound errors, called this assumption into question.

Evidence of possible differences between the receptive and productive phonologies is typically drawn from examples of specific production errors that are rejected by the children themselves when an adult produces the aberrant form. Berko and Brown (1960) described a famous example of a child who produced [fɪs] for *fish* but labeled as erroneous an adult's production of [fɪʃ] for *fish*. A similar example, one with more direct relevance to the application of phonological process analysis to delayed speech, was reported by Weismer, Dinnsen, and Elbert (1981) who described three children who regularly omitted final stop consonants from words like *dog* and *cab* but who did not like adults' productions of these words without the word-final stops; these children also produced evidence of having the correct, underlying forms when asked to inflect the stems (e.g., children who said [dɒ] for *dog* nevertheless produced [dɒgi] when asked to give the name for a little dog).

Whatever the assumption concerning underlying forms, an observation made by many investigators was that speech sound errors often occurred in groups tied together by an underlying theme. The child who omitted the word-final /g/ in *dog*, for example, was also likely to omit other, and perhaps most word-final obstruents. Similarly, the child who produced /kʊl/ for *school* most likely produced other s+ stop sequences (as in words like *spot* and *stop*) by reducing the cluster to a singleton consonant. This observation led to a reformulation of the 'functional misarticulation' notion, specifically by implicating *phonological processes* as a key player in normal speech sound development as well as in many cases of speech delay. A phonological process is a psycholinguistic operation applied to the underlying representation, the output of which is a surface form. When a child produces open-syllable forms for words with final consonants (e.g., [dɒ] for *dog*, [kæ] for *cat*, [mɪ] for *miss*), a process of final consonant deletion is said to intervene between the (correct) underlying form and the surface representation. Interestingly, it is known that at various stages in the learning of the ambient phonology typically developing children often omit word-final consonants, produce only one consonant instead of two or three in target clusters, and solve the apparent difficulty of fricatives by producing them as stop consonants. Whereas previous descriptions of speech sound development, and treatment plans in cases of speech delay, treated individual sound errors as more or less independent entities or paid attention to classes of speech sounds (such as fricatives), the phonological processes of final consonant deletion, cluster reduction, and stopping of fricatives, among others, were believed to capture the organized nature of developing and delayed sound systems in a way that transcended sound classes and reflected the cognitive reality of phonological structure. Phonological processes were therefore thought to provide a more elegant account of speech learning behavior.

Phonological processes, as reviewed by Ingram (1976), Locke (1983), Shriberg and Kwiatkowski (1980), and Elbert, Dinnsen, and Weismer (1984), signaled in at least two important ways a specific paradigm shift in the view of children with speech delay. First, the idea that psycholinguistic processes mediated between the underlying, phonological

forms and their phonetic realizations signaled a subtle shift away from the belief that speech motor maturity played a central role in the sequence of typical speech sound development, and in speech delay. True, many phonological processes in typical and delayed development resulted in presumptive articulatory simplifications as inferred from markedness conventions (see Hyman, 1975, Chapter 5 for an excellent review of markedness and its relation to articulatory simplification and ‘natural’ phonological processes). Thus consonant–vowel–consonant (CVC) syllables reduced to CV by a process of final consonant deletion was consistent with articulatory simplification because CV is the unmarked form. The same can be said of consonant singletons relative to clusters, and stop consonants relative to fricatives. The shift from speech motor maturity to phonological processes as explanations for speech sound errors is in some ways a subtle move, in light of the strong influence of the markedness concept on the phonological processes noted frequently in typically developing phonologies and cases of speech delay. Clearly, investigators conceived of the shift as one from an articulatory-motor to linguistic perspective, but the linguistic apparatus seemed to be based on articulatory simplicity. If phonological processes are merely a new descriptive device whose motivation is much the same as speech motor maturity, why submit to the paradigm shift?

The answer was in the second aspect of the paradigm shift. Phonological processes suggested a mechanism for the joint mastery of multiple sounds, often cutting across sound classes, in typical development or as a result of treatment in cases of speech delay. In treatment from a speech motor maturity perspective, children would often be given articulatory drill as a form of speech motor practice for specific speech sounds; the repetition was assumed to *automatize*<sup>4</sup> the articulatory behavior required for correct production of the sound. Treatment from the perspective of phonological processes viewed the child’s production problem as one of rule-learning (or unlearning, as the case may be: see Stampe, 1973), for which a properly constructed remediation plan would not only exploit at least some of the multiple sounds encompassed by the rule but also result in the normalization of *all* such sounds affected by the process. Thus the child who deleted word-final obstruents would be taught, by use of multiple but not exhaustive examples of syllable-closing consonants, the concept of closed syllables.

More recently, OT has been argued to supersede phonological processes as a putative account of speech sound learning in typical development, and to show great promise in explaining errors and their treatment in speech delay. Briefly, OT assumes a set of constraints operating between underlying forms – termed *inputs* – and output forms which are the surface (phonetic) representations. The constraints, thought to be universal, mediate the relationship between input and output forms by virtue of their ranking, which

---

<sup>4</sup> *Automatized* is a term from the older speech pathology literature (see, for example, Johnson, Shelton, Ruscello, & Arndt, 1979), signifying a process whereby a motor behavior becomes routinized or ‘burned in’ to neural circuitry by virtue of multiple repetitions. It obviously shares much with notions of motor skill acquisition in sports and musical performance. The term still shows up in the professional literature on reading and even speech perception (Johnson & Ralston, 1994).

is highly flexible. Operating under a basic goal of obtaining the same number of segments in the input and output representations (the faithfulness constraints), the best match is sought between input and output forms but not at the expense of violating highly ranked constraints which dominate lower-ranked constraints in a sort of tournament model. A very good introduction to OT and its application to sound development and delay is found in Barlow and Gierut (1999).

OT has been claimed to be a better paradigm than phonological processes for understanding developmental speech sound errors – typical or delayed – because processes are seen as relatively clumsy in their sequential application (as compared to the parallel processing in OT) and are ill-adapted to handling the common intra- and inter-child variability in speech sound production (see below). Markedness, however, is a prominent concept in both paradigms. Early in typical development, and presumably for children with speech delay, markedness constraints in OT are ranked very high and therefore result in mismatches between input and output forms. For example, syllable shape has been proposed as a highly ranked markedness constraint in the early stages of typical speech sound development (Barlow & Gierut, 1999). By virtue of this high ranking, input syllables having CVC form will be realized in the output as CVs because the syllable shape markedness constraint dominates lower-ranked and more easily violated faithfulness constraints of matching the number of input and output segments. Similarly, markedness constraints of ‘no consonant clusters’ and ‘no fricatives’ rank very highly in early phonological development and therefore result in cluster reduction and stopping of fricatives, respectively. Gradually, with the progression of typical development or as a result of speech therapy, markedness constraints lose their high ranking – in the language of OT, they are *demoted* – and better matches are obtained between input and output forms. Importantly, the constraints do not disappear, like certain phonological processes that are undone as development proceeds, but they are reorganized. The constraints can be thought of as a variably programmed filter, the soft assembly of which resides in their flexible ranking. At different times throughout development, this filter will transform the input in different ways until the proper filter configuration – yielding the adult output – is achieved.

Why would a universal set of constraints get reranked over the course of normal development or speech therapy? It appears this question has not yet been treated in the kind of theoretical depth required for OT to emerge as a truly compelling account of speech sound errors. Some clues as to how this aspect of OT will be developed have been summarized by Boersma and Levelt (2003), who argue that reranking of constraints is driven by production errors and their comparison by the learner to the ambient phonology – that is, the targets used by adults. Presumably, the learner develops over time some sort of quantitative model of mismatches between her output and adult forms, the cumulative effect of which is a reranking of constraints to better match her output to the adult forms. This notion carries with it the assumption for the child learner of adult-like, phonetic perceptual skills, as well as the related OT assumption of adult-like underlying forms. If these assumptions are not operative, of course, the markedness filter of the early constraint rankings (what Boersma & Levelt refer to as the ‘emergence of the unmarked’) makes little sense.



These considerations again raise the question, posed earlier, for both OT and phonological processes as they apply to typical speech sound development and speech delay: If both depend so much for their explanatory substance on markedness, a concept with much allegiance to articulatory complexity (and by inference speech motor maturity), what advantage is gained by the reformulation of speech sound development and delay in psycholinguistic terms? Considering phonological processes first, even if they are motivated partially or largely in the same way as speech motor maturity, the potential for rule induction moves past the presumptive neural nuts and bolts of articulatory practice, either as a part of typical development or as a remediation strategy (see above). Similarly, OT would view speech sound errors as the result of a constraint ranking that differs from the ranking adhered to by speakers with 'normal' phonologies. Barlow & Gierut (1999, p. 1491) clearly summarize these issues for the case of typical development.

Extending this to phonological acquisition, these constraints that are present at initial stages of development are also assumed to be the very same constraints present at later points in time. The longitudinal course of developmental change is characterized by reranking constraints. The likelihood is that the markedness constraints in particular will be demoted in children's systems. This follows from the observation that markedness constraints outrank faithfulness constraints in development, which is just the reverse pattern in fully developed systems.... Notice that this view of change is different from prior derivational accounts involving phonological processes or rules that are lost, suppressed, or eliminated from the grammar over time.

## **2.2. Consistency and Modifiability of Errors: Does the Theoretical Viewpoint Matter?**

Are the various, evolved views of normal and disordered phonological development merely academic playthings, or do they mean something in the real world? Recent work on treatment outcomes based on approaches developed with psycholinguistic theory in mind suggests that it does matter.

The clinical plan for a child who has speech delay characterized by multiple sound errors will be to normalize those errors in the most efficient way, but exactly how one goes about organizing such a plan has been the subject of a great deal of debate. As pointed out by Gierut (1998), the treatment plan for a given child is very much dependent on the diagnostic procedures used to identify the details of a child's speech delay; the diagnostic procedures used will depend on the examiner's theoretical orientation. The primary issue in matching a treatment plan to a diagnostic profile is that of *generalization*. When speech-language pathologists use this term they may mean several things, but the most basic is the spread of treatment effects from a treated sound to other, untreated sounds. For a diagnostic analysis based on sound classes, one might hope that treatment of a single misarticulated fricative will generalize to other misarticulated fricatives, even if the latter are not directly addressed in therapy. Similarly, a phonological process analysis might prompt treatment of selected cases of closed syllables to generalize to the entire class of closed syllables. And finally, an OT perspective in the diagnostic workup would

suggest a therapeutic focus on the highest-ranking constraint indicated by the child's sound error pattern, the assumption being that its demotion will, by implication, 'pull along' lower-ranked constraints to lower levels thus eliminating the mismatches between input and output forms that suggested speech delay and prompted treatment.

It would seem as if a good test of these different phonological theories would be found in the results of different forms of treatment and the diagnostic procedures from which they are derived. Unfortunately, as in most kinds of human behavior, it is not quite so simple. First, *most* kinds of therapy for children with speech delay produce positive effects (Gierut, 1998), making it difficult to choose between theories underlying different treatment strategies. The most desirable kinds of experiment to address competing theories would be to organize a group of children of the same age and with the same profile of phonological delay, with random assignment of each child to one of the treatments based on different theories. The ability to conduct such an experiment is obviously very limited, for the obvious and many logistical reasons known to investigators who do treatment research. More basically, this experiment presupposes clear-cut predictions between the different theories. OT, for example, makes certain predictions concerning which constraints should produce the greatest effect throughout a delayed phonology, but it seems as if the same kind of predictions might be made (at least in certain cases) by a phonological process approach. This makes sense because both theories are, as discussed above, dominated by markedness considerations.

With these caveats in mind, here is what seems to be known about the relationship between theory and treatment planning. There is a general sense, perhaps counterintuitive at first blush, that treatment with the greatest potential for generalization should always focus on sounds having some form of greater, as compared to lesser complexity within an ambient sound system. For example, experimental evidence reviewed in Gierut (1998) and Gierut and Morissette (2005) suggests greater generalization with treatment of more, as compared to less phonetically complex sounds (e.g., fricatives vs. stops), later as compared to earlier developing sounds (e.g., certain glides vs. stops), and non-stimulable as compared to stimulable sounds. This latter finding requires some additional comment. When a child presents in the clinic with multiple sound errors, the clinician wants to determine which of these error sounds can be produced correctly by the child under optimal diagnostic conditions, and then increasingly more complex conditions. 'Optimal diagnostic conditions' implies sound production with minimal influences of phonetic context and lexical identity, as well as of pressures to formulate language and convey an intelligible message. When a child can produce one of her error sounds correctly, under optimal conditions, the sound is said to be 'stimulable'; the child is said to be 'non-stimulable' for those sounds remaining in error even under the best of production conditions.<sup>5</sup> It is worth repeating the finding

---

<sup>5</sup> The precise notion of stimulability is not codified in the speech-language pathology literature, and is best thought of as a relative notion. For example, some children may be identified as stimulable for sounds in isolation, others for sounds in isolated words; or, the notion of stimulability may depend on how a correct production is elicited (auditory stimulus, placement of articulators by the clinician, and so forth). See Bauman-Waengler (2000, pp. 132–133) for a review of stimulability and clinical diagnosis.

reported by Gierut (1998): when a child has multiple sound errors, treatment of the non-stimulable sounds seems to generate more widespread change throughout the sound system as compared to treatment of stimulable sounds. The overall message seems to be, the more complex the work, the more generalization will occur.

This provisional notion of greatest generalization being achieved when treatment ‘targets’ are chosen from the high end of the phonetic complexity continuum may have a unifying expression in the hierarchical constraint structure of OT. Recall that one view of normal phonological development entails a reranking over time of a fixed set of constraints, with the typical progression of markedness constraints dominating faithfulness constraints early in development, but gradually demoted below faithfulness constraints to allow the emergence of adult forms. It has been hypothesized that within the set of markedness constraints, certain *fixed* rank orders may not be violated, even when one constraint within a set is demoted. This important aspect of the potential explanatory power of OT is worth a fairly detailed consideration, using an example from the literature.

Dinnsen and O’Connor (2001a) have presented a theoretical analysis of two common errors in typical phonological development, *consonant harmony* and *gliding*. The former error can be characterized as “... an assimilatory process that copies or spreads place or manner features to either a consonant or glide elsewhere in the word” (Dinnsen & O’Connor, 2001a, p. 599). An example of this error is the target word *won* (/wʌn/) produced as *none* ([nʌn]). The latter error, gliding, involves a change in consonantal /r/ or /l/ to [w], as in error *won* ([wʌn]) for target *run* (/rʌn/). Dinnsen and O’Connor (2001a) note the co-occurrence of these errors in the developing phonologies of several children and ask the logical question, why should these errors be seen together (see also Dinnsen & O’Connor, 2001b; Gierut & Morissette, 2005, for additional examples of co-occurring error types)? A derivational approach to phonological errors – for example, explanations of the errors involving phonological processes – cannot provide a reason for these kinds of co-occurrence and in fact treats them as chance (or at least unprincipled, with respect to theory) pairings. Dinnsen and O’Connor (2001a) show for several children and by appeal to a data base of phonological errors, however, that not only do consonant harmony and gliding co-occur but they do so with a unidirectional, implicational relationship. Specifically, if consonant harmony errors are present gliding errors will be as well (i.e., errors of consonant harmony imply errors of gliding), but gliding errors can exist alone and hence do not imply consonant harmony errors. If derivational accounts cannot provide principled accounts for the co-occurrence of errors, they certainly cannot explain a one-way, implicational pattern such as the one just described.

The advantage of OT in understanding phonological development and disorders is best exemplified in the way the theoretical apparatus can explain error co-occurrences and their implicational relations. This explanation, in turn, points to a principled selection of treatment targets having the greatest potential for generalization to untreated sounds. In the current example, Dinnsen and O’Connor (2001a) argue for a universal constraint ranking in which the constraint for avoidance of [r] always outranks the constraint for matching

the place or manner features of consonants within a word.<sup>6</sup> Thus if the latter constraint filters the input (the underlying representation), so must the former because all higher-ranked constraints must apply; but the highest-ranked constraint in this example – avoid [r] – can filter the input and result in gliding without requiring the presence of consonant harmony, which is produced by a lower-ranking constraint. The implicational, one-way relation is captured by the constraint ranking. Clinically, the logic of OT holds that in a child with consonant harmony and gliding errors, a treatment focus that targets and successfully demotes the ‘avoid [r]’ constraint will pull along the ‘consonant harmony’ constraint as well, because of their fixed and universal ranking. On the other hand, targeting the harmony constraint, even if successful, will only change those errors, leaving the higher ranked constraint to filter the input as before and thus produce gliding errors. If this concept is correct, the most efficient and productive use of a clinician’s time in treating a child with these co-occurring error patterns would clearly be to focus on gliding.

Gierut and Morrisette (2005) describe some clinical data for a different pair of co-occurring errors, consistent with the concept of fixed constraint ranking and implicational relations. Stopping (the substitution of stops for fricatives, as in [tʌn] for *sun* (/sʌn/), very common in typical, developing phonologies and in children with speech delay) and liquid gliding are said to co-occur as a result of the following, fixed constraint ranking: ‘avoid liquids’ must outrank ‘avoid fricatives’. This ranking *requires* gliding to co-occur when stopping is present, but allows gliding to exist when fricatives are produced correctly (or at least without stopping; there are other kinds of fricative error, discussion of which is beyond the scope of this chapter). Gierut and Morrisette (2005) cite treatment results of Tyler and Figurski (1994), who identified two children with this pair of errors and showed that the child treated for demotion of ‘avoid fricatives’ did not show generalization to liquids, but the child treated for demotion of ‘avoid liquids’ showed generalization to fricatives. Once again, a treatment effect wherein elimination of one member of a pair of co-occurring error patterns generalizes to the other member, but only in one direction; the effect seems to be ‘explained’ by the constraint rankings of OT and the implicational error patterns implied by them.

Can these treatment effects be viewed through the lens of complexity, as discussed above? The hypothesized, fixed universal constraint rankings, using the non-technical language adopted here for the two examples of co-occurring errors, are shown below.

Error pairs →	[w] for /r/ & consonant harmony	[w] for /r/ & stopping
Constraints →	Avoid [r] Match consonant manner	Avoid [r] Avoid fricatives

<sup>6</sup> The jargon and assumptions of optimality theory are being simplified for this discussion. OT would call the first constraint ‘\*R’ ( ‘avoid [r]’ (actually, avoid the co-occurrence of [+ consonantal] and [approximant]), and the second constraint ALIGN (manner features must be aligned with the left edge of prosodic domain).

These are all markedness constraints, by definition having the function of simplifying the output relative to the input (underlying forms). For any one of the constraints, there is probably little argument about their simplifying nature, but a more pertinent question is why one markedness constraint gains supremacy over another, a necessary condition to produce the one-way, implicational error patterns and preferred clinical targets discussed above. Is the production of /r/ more obviously complex than the production of two different manners of consonant articulation with the same word or the production of fricatives? Or, are the kinds of universal, fixed constraint rankings invoked by Dinnsen and O'Connor (2001a, 2001b) and Gierut and Morrisette (2005) more reasonably described as descriptions of error patterns elevated to the status of theory, and made to fit the particulars of the theory? What is needed to move OT forward with respect to speech sound development and delay is an extensive data base and analysis of co-occurring error patterns, as well as additional theoretical development of rankings *within* the set of markedness constraints.

### 2.3. Additional Thoughts on OT and Error Variability

At a coarse level, there are some 'population' patterns of typical phonological development: in most children stops will be mastered before fricatives, nasals before liquids, voiced consonants before voiceless. But, for individual children there will be deviations from these patterns. Similarly, within the group of children diagnosed with speech delay there will be frequently seen errors, but their nature and resolution as a function of time and/or treatment is highly variable. Theories of phonological development and disorders have always struggled with this variability; an appealing aspect of OT is the fixed set of constraints that can be reshuffled in rank to produce different patterns of errors at different times within the same child, and at the same time across several children. It is not clear, however, how this flexibility of constraint ranking can handle some of the well-known variability phenomena in typically developing and disordered phonologies. For example, individual children may produce a sound correctly in some word positions, but not others, and even this pattern may be inconsistent. Moreover, children may be inconsistent in sound production in a word-specific way, with the sound produced correctly in some words, incorrectly in others, even for the same position-in-word (see Betz & Stoel-Gammon, 2005, for a review). One approach to dealing with this kind of very particular inconsistency is to invoke new constraints, as Gierut and Morrisette (2005) do to explain the apparent generalization benefit of training error sounds in high, as compared to low frequency words (see Morrisette & Gierut, 2002).

OT is a theory owing much of its structure and apparatus to modern connectionist approaches to cognition, and therefore would seem to have compelling psycholinguistic cachet. In theory, OT could 'explain' any pattern of sound errors, simply by invoking new constraints. This makes the theory excessively powerful, however, and detracts from its appeal. The *ad hoc* addition of constraints as they are needed to fit the theory to data is, as noted in the Introduction, precisely the case of confusing explanation with description.

## 2.4. Summary

Psycholinguistics plays an important role in research and clinical practice among children with speech delay. Because cases of speech delay are thought to mirror the patterns of speech errors in typical development, it makes sense to have a theory that applies broadly to typical and disordered speech sound development (see Bernstein & Weismer, 2000). The review presented above provides compelling evidence for the potential utility of phonological analyses – whether in the form of generative or OT theories – in explaining error patterns and selecting a therapy plan for remediating errors in children with speech delay. These theories move past a simple speech motor perspective of speech sound learning and errors, but do not discard the importance of physiological factors in their explanatory apparatus. In particular, the concept of markedness – an important component of which is different degrees of articulatory complexity – plays a critical role in the generative and OT approaches discussed here. As suggested above, large-scale studies of error patterns in typical and delayed sound development are needed to evaluate precisely how much theoretical weight should be given to the factor of articulatory complexity and, by implication, to markedness. Especially in OT, where constraints are thought to be universal but reordered in rank for different languages, children, and even possibly within children as function of time, markedness must be understood completely to grasp its reach and limitations in understanding speech sound errors. This is especially urgent in light of recent evidence that articulatory complexity explains rather different degrees of variance in typical speech sound development for different languages (Stokes & Surendran, 2005). At this point in time, only an *ad hoc* addendum to the theoretical structure of OT would seem capable of accounting for this cross-linguistic finding.

## 3. VOT AND THE VOICING DISTINCTION IN SPEECH DISORDERS

In the history of speech production research on both normal speakers and speakers with disorders probably no other measure has been explored so consistently as VOT. VOT, a measure typically obtained from the speech acoustic signal, is defined as the time interval between the release of a stop consonant and the onset of vocal fold vibration for the following vowel. In spectrographic displays of the speech signal, the release is indicated by the burst, and the onset of vocal fold vibration by the first vertical striation following an interval of aperiodic energy; the interval is illustrated for word-initial, English /t/ and /d/ in Figure 1.

Lisker and Abramson (1964) published the seminal paper on VOT, showing for several different languages how VOT mapped onto phonemic voicing distinctions. English, for example, has two voicing categories for stops, the voiceless, aspirated group (p,t,k) and the voiced group (b,d,g). As described by Lisker and Abramson (1964) and a legion of scientists following them, voiceless stops in English are produced with VOTs in excess of about 30 ms (long-lag VOT) whereas voiced stops have VOTs less than 20 ms (short-lag VOT) or voicing lead (negative VOT, wherein the onset of vocal fold vibration precedes the burst). French and Spanish also have two stop voicing categories but differ from

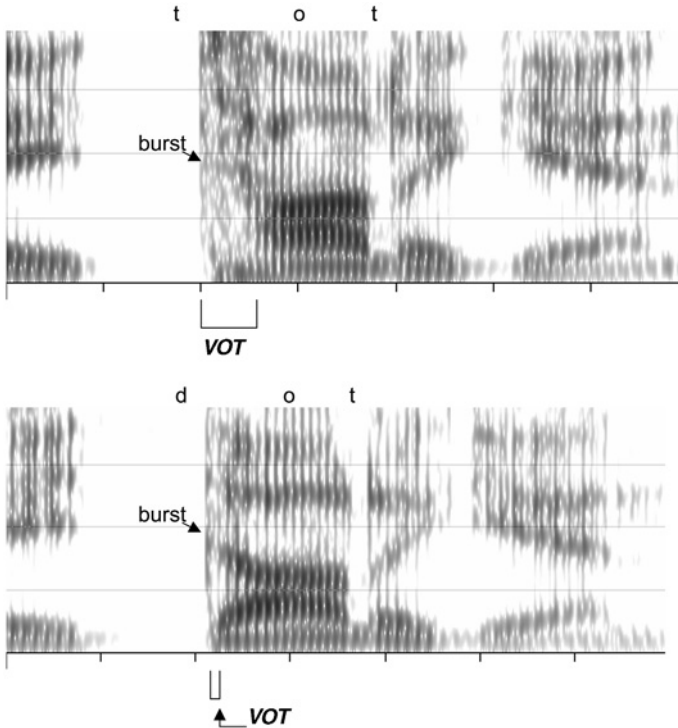


Figure 1. Spectrograms of the words ‘tot’ (top) and ‘dot’ (bottom), showing difference in VOT for voiceless and voiced stops. The left hand, vertical line of each VOT interval is aligned at the stop burst, the right-hand line at the first glottal pulse of the following vowel. Time scales are the same for the two spectrograms.

English in using other parts of the VOT continuum to signal the contrast: voiceless stops are unaspirated and produced with short-lag VOTs, voiced stops with voicing lead. No matter where the contrast is made along its continuum, VOT corresponds consistently well with the phonological voicing distinction in many different languages. In fact, because of Lisker and Abramson’s work and its many sequela VOT is often regarded as a good expression of the voicing status of stop consonants. It is this correspondence – the assumption of a fairly straightforward mapping between the acoustic measure and phonological voicing status – that prompted a literature on voicing errors in neurogenic speech disorders and inferences from acoustic data concerning their origin.<sup>7</sup>

<sup>7</sup> The actual separation of VOT values for different members of a phonological voicing contrast is somewhat messier than implied by this brief description, as a result of factors such as phonetic context, position in word, stress, speaking rate, speaking style, and even speaker age. The discussion here is limited to those cases in which the VOT-voicing status mapping is thought to be clear, for example, stops produced in word-initial, stressed syllables spoken at conversational rates and in non-casual speaking styles.

Speakers who have degenerative or acquired neurological disease are known to produce voicing errors for stop consonants (Blumstein, 1973, 2001; Platt, Andrews, & Howie, 1980; Shewan, Leeper, & Booth, 1984). Among the acquired diseases, stroke has been studied a fair amount for segmental articulatory errors. And among the several different kinds of speech-language disorders associated with stroke, there has been interest in differential effects of more anterior vs. more posterior lesions on speech production performance. Anterior lesions, often associated with Broca's aphasia or symptoms more or less consistent with its classical description (agrammatism, hesitant, dysfluent speech, numerous speech sound errors, and possibly dysprosody: see Rosenbek, Kent, & Lapointe, 1984) is thought to disrupt mechanisms of speech production but typically not the integrity of underlying units of production (e.g., phonological representations). On the other hand, posterior lesions, in Wernicke's area, were thought to leave motor execution areas relatively intact but to have the potential to affect phonological representation. Damage to the insula, sometimes associated with conduction aphasia, also produces errors which have been thought to be phonological in nature (Cantor, Trost, & Burns, 1985; Shewan et al., 1984) or may be associated with speech motor programming difficulties and hence apraxia of speech (Dronkers, 1996).

Several papers were published in the 1970s and 1980s (e.g., Blumstein, Cooper, Zurif, & Caramazza, 1977; Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980; Itoh et al., 1982; Shewan et al., 1984), attempting to show how VOT measures could adjudicate the origin of phonetic voicing errors. The reasoning, at first glance, is quite simple and compelling: If a patient's attempt to produce a voiceless stop consonant results in a short-lag VOT, either the representation of the phonological unit was incorrect (i.e., a voiced consonant rather than a voiceless one) or something went wrong in the selection of the correct phonological unit; the reasoning would apply in the same way to voiced stop targets produced with long-lag VOT. These would be considered *phonemic* errors because the problem was in the underlying segmental unit or how it was selected, not in its implementation by the speech mechanism. On the other hand, an attempt to produce a voiceless stop with a VOT in the region *between* short- and long-lag VOT – say, between roughly +20 and +35 ms – would be regarded as a *phonetic* error, derived from a correct representation and selection process but distorted by faulty and presumably lower-level motor planning and/or production mechanisms.

Blumstein et al. (1977) reported measurements of VOT for the /d/-/t/ contrast from single words produced by persons with a variety of aphasic types, including patients diagnosed with Broca's and Wernicke's aphasia. A patient with Wernicke's aphasia produced non-overlapping distributions of VOT for the /d/-/t/ contrast with no 'in-between' values, but some clearly long-lag VOTs for /d/ targets. These latter errors were interpreted as phonological in nature:

“...the VOT measurements of the posterior aphasics revealed only the presence of phonemic substitutions, i.e., clear-cut shifts in phonetic category from the target phoneme...these errors reflect a deficit in selecting the appropriate phoneme or underlying phonological form, and subsequently programming correctly the articulatory commands for the substituted phoneme” (Blumstein et al., 1977, p. 381).



In contrast, a patient with Broca's aphasia produced mostly overlapping distributions of VOT for /d/ vs. /t/, with values in the ambiguous region and only an occasional value clearly in the 'wrong' category. For Blumstein et al. (1977), this was evidence of an articulatory programming or phonetic disintegration (p. 381) problem, different from the origin of the errors in the Wernicke's aphasic. Similar results and conclusions were published later by Blumstein et al. (1980).

Much of the work following the two papers by Blumstein et al. (1977, 1980) has not supported a differential type of error made by persons with anterior vs. posterior lesions of the left hemisphere. If anything, the results of Itoh et al. (1982) and Shewan et al. (1984) showed a tendency for *all* patients to make stop voicing errors of both sorts, with both 'between-region' (20–35 ms) and 'error-region' VOTs (e.g., long-lag VOT for voiced targets or short-lag VOT for voiceless targets) observed for person's diagnosed with Broca's, Wernicke's, and conduction aphasia. Baum, Blumstein, Naeser, and Palumbo (1990), in a study of VOT plus other temporal measures of speech utterances produced by patients with anterior and posterior lesions, showed that whereas both types of patients made phonetic errors, the anterior patients did so with much greater frequency.

Baum et al. (1990) is an interesting paper because of the explicit nature of the predictions concerning the inference of phonetic vs. phonological errors from temporal characteristics of various speech acoustic events. Although it is beyond the scope of this chapter to examine each of the types of inference considered by Baum et al. (1990, see p. 35 for their reasoning), it is interesting to examine critically the underlying logic of VOT measures as a window to phonological representations and selections. Apparently, even when data have been somewhat or largely ambiguous – as in almost all the studies – this logic has not been questioned (but see the Discussion in Shewan et al., 1984, pp. 214–216). But there are reasons, both physiological and empirical, to question the validity of an inference from VOT measures to the phonetic vs. phonological nature of a stop voicing error.

### 3.1. Psycholinguistics Meets Speech Physiology

If there is a transparent relationship between VOT and phonological voicing status for stop consonants, we should ask if a similar, straightforward mapping exists for the underlying physiology of the voicing distinction. There is no *a priori* reason to prefer a mapping between an acoustic, as compared to physiological characteristic and a phonological category. In fact, for either the so-called translation theories or gesture theories of speech production (for reviews of such models see Löfqvist, 1997; Fowler, 1985) it is easy to formulate an argument for a reasonable mapping of a physiological event or small array of events to at least some of the phonological contrasts of a language. The voicing distinction for stops is just such a case, where the underlying physiology is well understood for normal speakers and can be used to evaluate the logic of the inferences discussed above.

Figure 2 presents a schematic summary of the physiology of the voicing distinction for English stops. The figure is based on work published by Löfqvist (1980) and Yoshioka, Löfqvist, and Hirose (1981), among others. The top two panels show

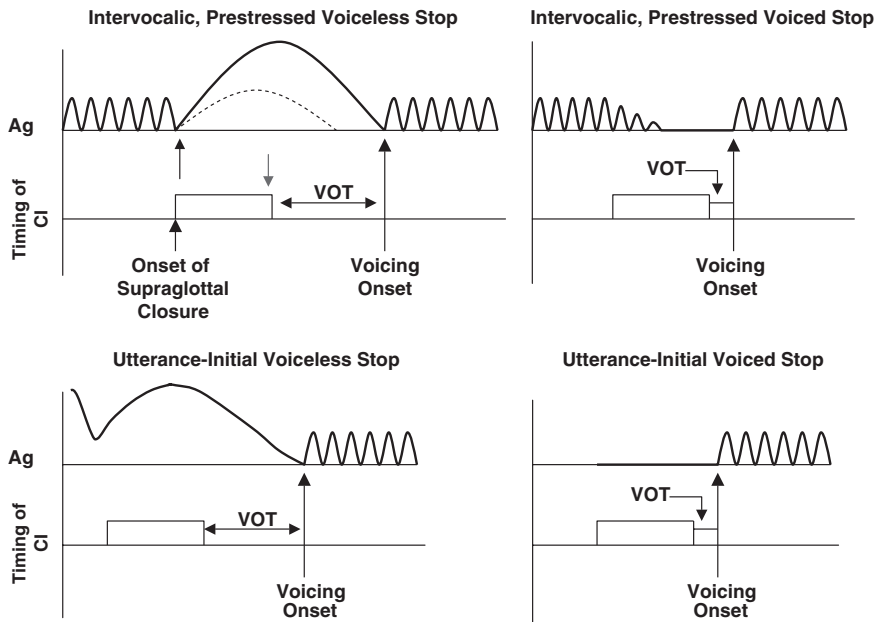


Figure 2. Schematic drawings of laryngeal and supralaryngeal behavior for voiceless and voiced stops in English. Intervocalic, prestressed stops are shown in upper two panels, utterance initial stops in lower two panels. Ag = glottal area function, Timing of CI = onset and offset (stop release) of stop closure interval. See text for details.

laryngeal – supralaryngeal events for intervocalic, prestressed stop consonants, as in common carrier-phrase utterances such as ‘Say *tot* again’ (upper left panel) and ‘Say *dot* again’ (upper right panel). The time history labeled *Ag* shows the opening and closing of the vocal folds throughout the final half of the vocalic /eɪ/ in the word ‘Say’, the closure interval and release phase of the stop consonant (/t/ in the left panel, /d/ in the right), and the vowel /a/ in ‘tot’ (left) and ‘dot’ (right). Observations of the opening and closing of the vocal folds can be made in a number of ways but here we will assume they have been recorded more or less directly, using a fiberscope inserted through the nose with images sampled at a sufficient rate to permit viewing of the very rapid opening and closing phases of vocal fold motion during the voiced segments of speech (e.g., see Mergell, Herzel, & Titze, 2000; Wittenberg, 1997). Immediately below the *Ag* functions are event histories labeled *Timing of CI*, where *CI* = closure interval. Stop consonants are produced, in part, by creating a brief, complete obstruction in the vocal tract to the egressive air stream from the lungs. The interval over which the obstruction is maintained, usually no more than 100 ms in the kinds of carrier-phrase utterances illustrated here, is called the closure interval. In the case of /t/ and /d/ it begins when the tongue tip/blade makes firm contact with the alveolar ridge, and ends when this contact is released. In the schematic diagram the *Timing of CI* functions are shown as raised boxes extending across some unspecified time interval,

where the onset (beginning) and offset (release) of the closure interval are the important events for the current discussion. The *Ag* function, reflecting a laryngeal event, and the *CI* function, reflecting a supralaryngeal event, are shown on the same time scale. Note also the identical durations of the closure interval for the voiceless (upper left) and voiced (upper right) stops, consistent with previously published work on stop closure durations in the intervocalic, prestressed position (Umeda, 1977; Stathopoulos & Weismer, 1983).

Considering the voiceless stop first, the relatively rapid openings and closings of the glottis at the beginning of the *Ag* function are due to the motions of the vibrating vocal folds for the vocalic /eɪ/ preceding the stop closure. These motions, which produce phonation (voicing), are the result of aerodynamic and mechanical forces (see Broad, 1979, for an excellent review) and are essentially periodic at typical rates of approximately 120 Hz (period = 8.33 ms) for adult men and 200 Hz (period = 5 ms) for adult women.<sup>8</sup> In the schematic, after six cycles of these nearly periodic motions there is a relatively long opening and closing gesture of the vocal folds. This is the *laryngeal devoicing gesture* (LDG), and it differs from the opening and closing gestures of phonation in some obvious and not-so-obvious ways. The most obvious difference is in the much-longer duration of the LDG, which is typically between about 100 and 150 ms. A less obvious difference, but one critical to the current discussion, is that the opening and closing motions of the LDG are under *muscular* control (Hirose, 1976), and do not result from aerodynamic and mechanical forces as in the case of phonatory behavior. For this reason the LDG is properly referred to as an *articulatory* gesture of the larynx, suggesting its essential role in the segmental characteristics of not only voiceless stop consonants, but also voiceless fricatives (such as /fəʃ/) and the voiceless affricate (/tʃ/). With some minor differences, the LDG is basically the same for all of the voiceless obstruents. Note also the synchrony, indicated in the figure by upward pointing arrows, of the opening gesture of the LDG with the onset of the supralaryngeal closure for the /t/; this time-locking of the laryngeal and supralaryngeal gestures also holds true for fricatives and affricates. Finally, the release of the closure for /t/ (downward-pointing arrow) occurs well before the LDG is completed, resulting in a relatively substantial interval between the stop closure release and the onset of phonatory behavior – marked in the figure as ‘voicing onset’ – for the following vowel. As marked by the horizontal line ending in arrowheads, this is the VOT interval.

An important aspect of the LDG is its continuous and stereotyped nature. Once initiated, the gesture seems to evolve over time without the kind of conscious control over its individual parts that might characterize many other movements. Löfqvist (1980) referred to the LDG as a *ballistic* gesture, a term used in the speech production literature to suggest an articulatory motion that is not adjusted after its initiation but rather runs a stereotyped course; there are experimental data consistent with this idea, suggesting that

---

<sup>8</sup> Vocal fold vibration is commonly described as “quasiperiodic” to indicate that the motions and their acoustic results are not perfectly periodic; the typical values of vibratory rate given for adult men and women are subject to a host of variables other than gender, such as age, health status, emotional state, type of speech material, and personal history (e.g., history of smoking or taking certain drugs).

speakers do not have much control over the scaling (size) of the LDG, even when given the best opportunity to change it (see Löfqvist, Baer, & Yoshioka, 1981). The ballistic nature of the LDG is an important piece of the argument that its presence signifies the phonological voicelessness of a stop consonant, and its absence the opposite; here the binary nature of the linguistic contrast is captured well by the underlying physiology.

The top right panel of Figure 2 shows the laryngeal and supralaryngeal events for the voiced /d/ in ‘dot’. The phonatory motions for the preceding /eɪ/ are, not surprisingly, the same as those preceding the /t/,<sup>9</sup> but at the onset of the closure interval for /d/ the periodic motions continue, gradually declining in amplitude until they disappear shortly before release of the stop. In some cases of voiced stops such as the /d/ shown here motions of the vocal folds may continue throughout the closure; the presence of those motions during the closure interval, or how rapidly they cease after the onset of closure, depends on the time history of the pressure difference across the vibrating vocal folds.

The VOT interval is clearly much shorter for /d/, as compared to /t/, and these typical short-lag VOTs for voiced stops will occur whether or not voicing continues throughout the closure interval or is terminated before the end of the closure interval, as shown in Figure 2. What is important here, and critical to the issue of inferring phonological status from empirically measured VOT values, is that this underlying physiology can be boiled down to a fairly clear dichotomy: voiceless stops have an LDG, voiced stops do not. When the LDG is present, at least for intervocalic, prestressed stops, VOTs will be in the long-lag range simply because the supralaryngeal closure interval is released about halfway into the LDG, when the vocal folds are maximally open (Löfqvist, 1980). This is so because it takes time to move the vocal folds back to the midline position where phonation can resume, and if the closure interval is released near the middle of the LDG or even slightly later, the VOT must be in the long-lag range. Conversely, when there is no LDG, as is the case for the upper right panel of Figure 2, periodic motions of the vocal folds can resume shortly after the release of the closure interval. Thus, in an English speaker with no neurological disease, the presence of the LDG can be assumed to be part of the voiceless specification for stops, and its absence part of its voiced specification. Presumably, a correct phonological representation of a /t/ will trigger the LDG as part of its specification for articulatory control; a /d/ will not. Note the bottom panels of Figure 2, where the typical physiology is illustrated for /t/ (bottom left) and /d/ (bottom right) in utterance-initial position, as would be the case for isolated words with initial stops. Even though there is no voiced segment preceding the /t/ an LDG is observed. For utterance-initial /d/, especially in English, there are typically no vocal fold vibrations during the closure interval (i.e., utterance-initial voiced stops are typically devoiced in English), but the vocal folds are held in the midline position during the closure and so begin to vibrate shortly following the stop release. The binary physiological opposition

---

<sup>9</sup> Like most generalizations of this sort, there are some qualifications; as vocal fold motions approach the closure interval of a stop consonant, details of the vibration are somewhat different for voiceless vs. voiced stops; see Ní Chasaide and Gobl (1997) for a summary of these effects.

of presence vs. absence of the LDG applies equally to utterance-initial stops, as it does to intervocalic stops, as important point as some experiments use isolated words to study the voicing opposition in aphasia or other neurogenic speech disorders.

So far, this explanation of the physiology of voiceless and voiced stop production and its relationship to the phonological voicing status of stops seems consistent with the inverted logic of inferring voicing status from VOT values. The critical question is, however, can an underlying physiological scenario be imagined wherein the LDG was present but the measured VOT was clearly in the short-lag range? If so, the logic of inferring phonological voicing status from VOT values runs into serious problems, because it makes little sense to label a stop as phonologically voiced when it *could* have an LDG; after all, the LDG is a physiological implementation of voicelessness. In fact, physiological data consistent with the scenario of production of an LDG with measured VOTs in the short-lag range have been described in the literature. For example, in languages having a phonemic distinction between unaspirated and aspirated voiceless stops (e.g., French and Mandarin) or in cases where different degrees of stop aspiration are elicited by varying position-in-word or syllable stress level (e.g., Swedish), fiberoptic observations have shown the presence of the LDG in all cases, but modifications of timing or size (i.e., maximum opening of the glottis) of the LDG and/or of the stop closure duration result in the small amounts of aspiration – that is, a short VOT (see, e.g., Iwata & Hirose, 1976; Benguerel, Hirose, Sawashima, & Ushijima, 1978; Löfqvist, 1980).

More specifically, the underlying physiology of voicelessness for stops clearly indicates how a patient might implement the correct and crucial articulatory feature – the LDG – yet still produce a VOT in the short-lag range. First, variations in duration of the closure interval, with magnitude of the LDG opening and its synchrony with onset of supralaryngeal closure held constant, will result in variations in VOT; longer closure intervals will produce shorter VOTs, and vice versa (see Löfqvist, 1980; Weismer, 1980). Longer stop closure intervals would certainly be consistent with the generally slower speaking rates observed in patients with anterior lesions (e.g., Kent & Rosenbek, 1983; Baum et al., 1990), and could possibly explain the occurrence of some short-lag VOTs for voiceless stop targets as reported by Blumstein et al. (1977, 1980) and Baum et al. (1990). Second, the LDG could be implemented but with smaller-than-normal magnitude – perhaps one of the subtle phonetic deficits hypothesized for patients with posterior lesions by Baum et al. (1990) and Kurowski, Blumstein, and Mathison (1998) – which would also result in shorter VOTs. Third, asynchronies between the onsets of an LDG of normal magnitude, and a supralaryngeal closure interval of normal duration, would also produce variations in VOT. An LDG onset that lags onset of a supralaryngeal closure interval will result in relatively longer VOTs, and vice versa. Coordination problems involving two or more articulators are often cited as a prominent problem in anterior aphasics and persons with apraxia of speech (Baum et al., 1990; Itoh, Sasanuma, & Ushijima, 1979) and could be part of the subtle phonetic deficit proposed for posterior aphasics. It is hardly far-fetched to imagine that such asynchronies might characterize a neurologically based speech production problem (see, e.g., Kent & Adams, 1989; and review in Weismer, Yunusova, & Westbury, 2003).

In the hypothetical cases described here, the presence of the LDG must be interpreted as good evidence of phonological voicelessness. Yet in each of case a VOT in the short-lag region is shown to be a possibility, and that possibility has been interpreted at times as a voiced-for-voiceless error of the phonological kind. These hypothetical cases, however, are physiologically plausible and are clearly *phonetic* anomalies. For these reasons, when a voiceless stop target is produced with a short-lag VOT it seems ill-advised to interpret the event as evidence of a phonological error, at least in the absence of additional evidence. The best form of such evidence would be direct viewing of the larynx during the production of voiceless stop targets, but of course this is not feasible in most clinical settings and indeed, is a specialty type of data even in research venues. Fortunately, acoustic measures and observations can be added to VOT measures to clarify the underlying physiology associated with an apparent voiced-for-voiceless substitution. For example, the possibility of a lengthened closure interval contributing to a short-lag VOT when an LDG is produced can be addressed by combining measures of the *voiceless interval* with measures of closure duration. Figure 3 illustrates the voiceless interval in a sequence like the one shown in Figure 2, where /eɪ/ precedes a voiceless stop, which is then followed by an /a/ (as in the ‘Say tot again’). The voiceless interval is measured from the final glottal pulse preceding the voiceless stop closure (leftmost, solid, upward-pointing arrow) to the first glottal pulse following that closure (rightmost, solid, upward pointing arrow); as such, the

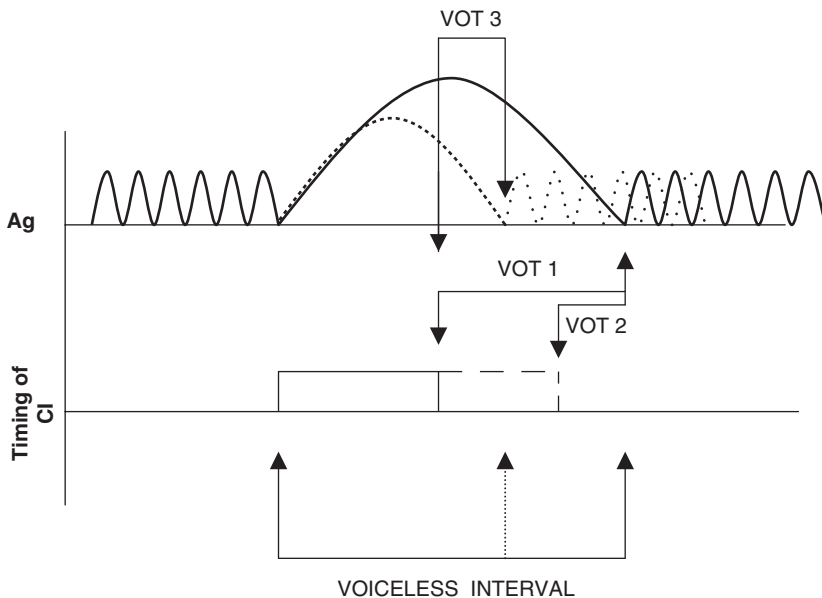


Figure 3. Schematic drawing of laryngeal and supralaryngeal behavior for voiceless stops, showing the effects of variations in stop closure duration and magnitude of the LDG on VOT measurements. All VOT intervals shown on the figure are taken between the release of the stop (the end of the indicated closure interval) and the first glottal pulse of the following vowel. See text for additional detail.

voiceless interval duration is the sum of the closure interval duration plus the VOT. Moreover, the interval should correspond rather closely to the duration of the LDG. If a lengthened closure duration is contributing significantly to the measurement of short-lag VOTs, the voiceless interval duration should be close to normal (somewhere between 100 and 150 ms) and the closure duration longer than normal (>100 ms).<sup>10</sup> Measured intervals like these would likely yield a short-lag VOT but would not be consistent with an interpretation of a phonological error. A comparison of VOT 1 to VOT 2 in Figure 3 shows clearly the effect of a lengthened stop closure interval (indicated by the dashed-line extension of the original closure interval) on a measured VOT, even with a normal LDG.

Figure 3 also shows how acoustic measures could suggest the presence of a smaller-than-normal LDG and its role in yielding a short-lag VOT. The reduced LDG and the vocal fold vibrations following it are illustrated by the dotted-line *A<sub>g</sub>* function; note that the onset of this LDG is still synchronized with the onset of the closure interval. Because, the LDG has a smaller-than-normal magnitude it is completed in a shorter amount of time and vocal fold vibration resumes earlier than in the case of the normal-sized gesture. With a normal closure duration (the solid-line closure interval), a short-lag VOT (VOT 3 in Figure 3) will result, but should not be taken as evidence of a phonological error. Here, the relevant acoustic measures would be the shorter-than-normal voiceless interval (indicated in Figure 3 by the interval between the solid and dotted upward-pointing arrows) and a normal closure duration.

The effects of asynchrony between the onsets of the LDG and closure interval are illustrated in Figure 4. In Example 1, the two events are synchronized, producing the expected long-lag VOT. The asynchrony in Example 2 has the onset of the closure interval lagging the LDG, which could result in a short-lag VOT. This asynchrony should also result in *preaspiration*, a relatively substantial interval of aperiodic energy prior to the stop closure, produced as a result of the vocal folds opening for the LDG while the vocal tract is still open. In normal, young adult speakers, preaspiration of voiceless stops occurs only occasionally, and then rarely for more than 10–15 ms; in speakers with neurogenic speech disorders preaspiration almost certainly occurs more frequently than in normal speakers, and often has durations exceeding 30 ms.<sup>11</sup> The presence and duration of preaspiration, plus measurement of a close-to-normal voiceless interval duration, would suggest that a short-lag VOT had the underlying physiology of voicelessness and therefore should not be considered as evidence of a phonological error. The opposite asynchrony, with the closure interval leading the LDG, is shown as Example 3. In this case the VOT would have an exaggerated, long-lag value; the acoustic evidence of such asynchrony would be the presence of glottal vibrations extending into the closure interval for more than approximately 20 ms, a value rarely exceeded in normal speakers (Weismer, 1997, 2004).

<sup>10</sup> Stop closure durations >100 ms are very unusual in normal speakers, and especially in the so-called ‘connected’ speech samples such as passage reading and spontaneous speech. In more formal types of speech, and especially in citation forms of speech (single words or words in brief carrier phrases), closure durations for bilabial stops may approach 100 ms.

<sup>11</sup> Supporting data have not, to the author’s knowledge, been published; the claim is made based on his experience in measuring acoustic records of speakers with dysarthria and aphasia.

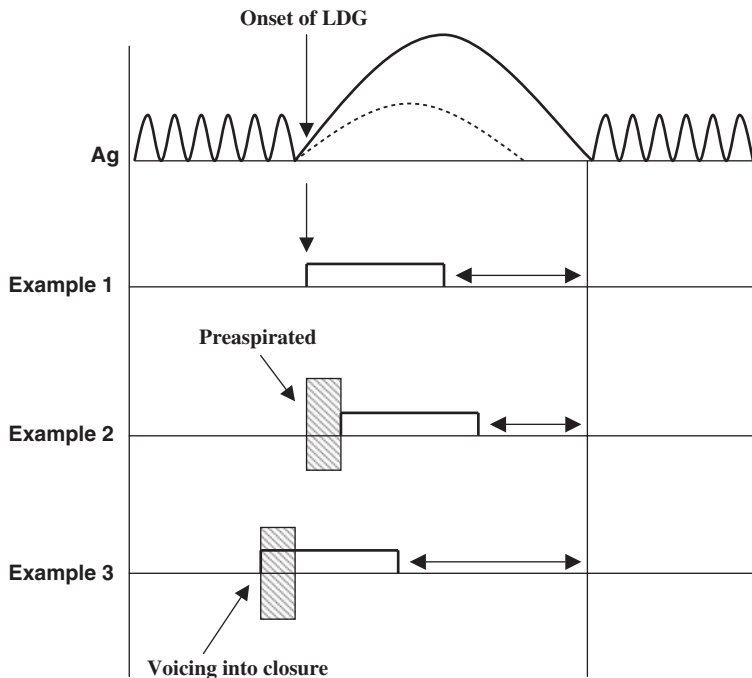


Figure 4. Schematic drawing of laryngeal and supralaryngeal behavior for voiceless stops, showing the effects of asynchronies between the LDG and stop closure interval on VOT measurements. See text for additional details.

To summarize, short-lag VOTs may be measured even when an LDG is present. In the absence of additional acoustic measurements, it does not seem wise to interpret short-lag VOTs as mapping onto phonological voicing status in a straightforward way. In papers such as Blumstein et al. (1977, 1980), and Baum et al. (1990), short-lag VOTs for voiceless stop targets have been regarded as manifestations of such phonological voicing errors; similar reasoning has been used in Ryalls, Provost, and Arsenault (1995) and Gandour and Dardarananda (1984a) for French and Thai speakers with aphasia, although the phonetic details are somewhat different.

In the case of voiced stop targets, it is substantially more difficult to demonstrate how the underlying physiology could be consistent with phonological voicing yet yield a long-lag VOT, at least in English.<sup>12</sup> Examples of such errors in persons with aphasia,

<sup>12</sup> When a two- or three-way opposition involves voiced and voiceless unaspirated stops, it is not at all difficult to envision the physiology of voiced stops resulting in a short-lag VOT and thus the possible interpretation of a phonemic (voiceless for voiced) error. With no LDG but vocal folds that are prevented from vibrating because the pressures below and above the glottis are the same, as often occurs toward the end of voiced stop closure interval, the vocal fold vibration may be delayed following the stop release by as much as 20 ms, resulting in the kind of VOT observed for voiceless unaspirated stops. See Ryalls et al. (1995) for examples from French and Gandour et al. (1992) for example from Thai.



apraxia of speech, and right-hemisphere damage have been reported by Blumstein et al. (1977, 1980), Kent and Rosenbek (1983), Baum et al. (1990), and Kurowski, Blumstein, and Mathison (1998). If the vocal folds are close to, or at the midline when a stop is released, it does not seem possible for VOT to be much more than 20, or at the limit, 30 ms; a voiced target produced with a VOT of 50–60 ms would seem to require an LDG which, as argued above, signifies voicelessness. Perhaps, then, measurement of a long-lag VOT when the target stop is voiced should be taken as good evidence of a phonemic error. There is, however, one more piece of evidence that demonstrates how tenuous this interpretation might be.

Dysarthria is a neurogenic speech disorder in which damage to the central or peripheral nervous system results in a problem with control of some or many of the scores of muscles involved in the production of speech (see Weismer, 1997, for a review). Although the diseases that result in dysarthria may also produce cognitive problems (such as mental retardation in cerebral palsy, dementia and depression in Parkinson disease, aphasia in stroke, to name a few examples), the speech motor control problem has never been thought to be complicated by a potential loss or modification of phonological representations. Stated otherwise, sound segment errors and their acoustic manifestations in dysarthria have always been considered of a phonetic origin, reflecting only the control problem.

Figure 5 shows data derived from conversational speech samples produced by 22 adults with dysarthria (Weismer, 2004); none of these speakers had serious cognitive problems, and all had been seen for research purposes related solely to their speech motor control deficit. This cumulative probability graph shows VOT for voiced (solid function) and voiceless (dotted function) stop consonants produced in the prestressed position. For 97 voiced stops and 112 voiceless stops, median VOTs across speakers and place of articulation were 16 and 49 ms, respectively, with both functions steeper below, as compared to above, the medians.<sup>13</sup> Mean Data from Lisker and Abramson's (1964) sentence production task, marked on the cumulative probability functions with filled squares, show the current data to be roughly comparable to those from normal speakers. Speakers with dysarthria therefore tend to maintain the VOT distinction for voiced stops in much the same way as normal speakers. There are clearly exceptions to this summary statement, however, because the cumulative probability functions show examples of voiced stops with long-lag VOT values (say, above 30 ms), and voiceless stops with short-lag VOT values (below 20 ms). The low frequency of occurrence of these apparently 'incorrect' VOT values is probably a negligible concern for the question of if and how speakers with dysarthria maintain the voicing distinction, but the evaluation of *individual VOT exemplars* has been taken as critical evidence in adjudicating a phonetic vs. phonological origin of stop voicing errors in aphasia. The data presented here are not unusual; VOT values clearly in the category

<sup>13</sup> Note the absence of VOT values in the 'lead' (negative) range for the voiced function. In connected speech, when voiced stops are in the intervocalic position vocal fold vibration can continue throughout the closure interval, with the duration of voicing lead equal to the duration of the closure interval. For this analysis, if voicing occurred at the same time as the stop release (which will be the case, give or take several milliseconds), the VOT was recorded as '0' (zero).

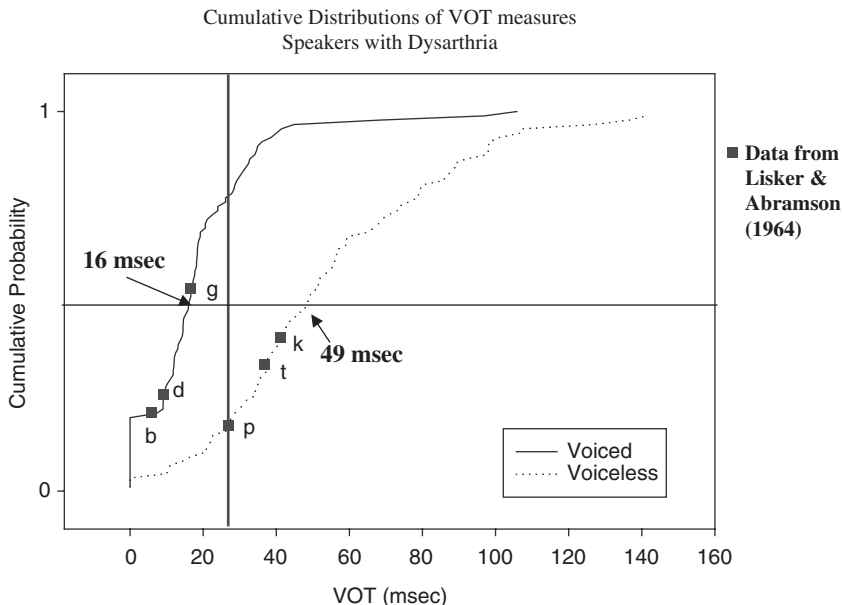


Figure 5. Cumulative probability distributions of voiced (solid function) and voiceless (dotted-line function) stop consonants, produced in the prestressed position by 22 adults with dysarthria. VOT data from Lisker and Abramson’s (1964) sentence productions are marked on the functions by solid squares; the thick vertical line at 25 ms is an approximate boundary value for voiced vs. voiceless stops.

opposed to the target have been reported for speakers with dysarthria by Caruso and Burton (1997), Lieberman et al. (1992), Ryalls Hoffman-Ruddy, Vitels, and Owens (2001), Ackerman, Graber, Hertrich, and Daum (1999), and Bunton and Weismer (2002). The logic of identifying long-lag VOTs for voiced targets and short-lag VOTs for voiceless targets as a result of incorrect phonological representations should not depend on the type of disorder under evaluation. The fact that such large-scale phonetic anomalies also occur in dysarthria, a group of disorders in which phonological representations are assumed to be unaffected by the disease process, highlights another problem in the use of VOT values to make the distinction between phonetic and phonological errors.

#### 4. SUMMARY

There is a long history of using phonetic data to refine and clarify phonological issues. One only need look at the set of volumes in the *Laboratory Phonology* series (see, e.g., Docherty & Ladd, 1992; Gussenhoven & Warner, 2002) to understand how phonetic analysis can advance phonological theory. In the case of speakers with neurogenic speech and language disorders, however, the potential ambiguities in relating phonetic phenomena to phonological representations and/or processes are of much more concern than in normal speakers (see also, Tuller, 1984). VOT is a special case where the underlying

physiology of voicelessness, and its various disruptions in neurological disease, complicate the mapping of phonetic fact to phonological inference. Some have even claimed the same level of inferential complexity for normal speakers' production of VOT (see Cho & Ladefoged, 1999). Although this part of the essay has focused on VOT as the most well-studied phonetic event in normal speakers and speakers with neurogenic disorders, the same caution would apply to other attempts to infer phonological integrity from acoustic measures relevant to, for example, place of articulation for stops (Shinn & Blumstein, 1983; see reaction by Ziegler, 1984), contrastive uses of vowel or consonant duration (Baum et al., 1990; Duffy & Gawle, 1984; Kurowski, Hazen, & Blumstein, 2003) and voicing characteristics of fricatives (Kurowski et al., 2003). To be sure, the kinds of studies reviewed above have great value in understanding the phonetic manifestations of aphasia, regardless of how those manifestations may bear on the origin of any speech production anomalies. One view of models and theories of speech production is that they should be able to account not only for normal behavior, but also for the ways in which that behavior may change in response to disease (Bernstein & Weismer, 2000). Thus studies of speech production in aphasia can be added to those in dysarthria, for example, to form a global view of how neurological damage affects speech motor control. In this regard, it would be very useful to obtain more relevant data in a variety of languages with varying phonological and phonetic characteristics. For example, data reported for Thai, French, German, and Mandarin by Gandour and Dardarananda (1984a, 1984b), Gandour et al. (1992), Ryalls et al. (1995), Ackerman et al. (1999), and Jeng, Weismer, and Kent (2006) have begun to show the range of phonetic effects in both aphasia and dysarthria, and the ways in which these effects either transcend or are constrained by the particular linguistic system under study. When more such data are added to those available from English-speaking individuals, which currently dominate the literature, we should have a better idea of how to construct a comprehensive model of psycholinguistics and speech disorders. This argument is very much parallel to one advanced by Dick et al. (2001): a range of languages must be studied to truly understand the role of grammatical and lexical deficits in aphasia, because different languages afford very different opportunities to explore these two issues. The same can be said for the phonetics–phonology interface.

## ACKNOWLEDGMENTS

The writing of this chapter was supported, in part, by NIH Awards DC00319 and DC03273. Portions of the manuscript were prepared during a leave in Xylokastro, Greece, and I thank Larry Shriberg and Judy Gierut for sending me materials on underlying forms and OT during that time. An earlier version of the VOT analysis was presented at the 146<sup>th</sup> meeting of the Acoustical Society of America, New York, 2004.

## REFERENCES

Ackerman, H., Gräber, S., Hertrich, I., & Daum, I. (1999). Phonemic vowel length contrasts in cerebellar disorders. *Brain and Language*, 67, 95–109.

- Barlow, J. A., & Gierut, J. A. (1999). Optimality theory in phonological acquisition. *Journal of Speech, Language, and Hearing Research*, *42*, 1482–1498.
- Baum, S. R., Blumstein, S. E., Naeser, M. A., & Palumbo, C. L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language*, *39*, 33–56.
- Bauman-Waengler, J. (2000). *Articulatory and phonological impairments: A clinical focus*. Boston: Allyn and Bacon.
- Benguerel, A. -P., Hirose, H., Sawashima, M., & Ushijima, T. (1978). Laryngeal control in French stop production: A fiberoptic, acoustic and electromyographic study. *Folia Phoniatica*, *30*, 175–198.
- Berko, J., & Brown, R. (1960). Psycholinguistic research methods. In: P. Mussen (Ed.), *Handbook of research methods in child development* (pp. 517–557). New York: Wiley.
- Bernstein, L. E., & Weismer, G. (2000). Basic science at the intersection of speech science and communication disorders. *Journal of Phonetics*, *28*, 225–232.
- Best, C. T. (1994). Learning to perceive the sound patterns in English. In: C. Rouvee-Collier, & L. P. Lipsitt (Eds), *Advances in infancy research* (Vol. 9, pp. 217–304). Norwood, NJ: Ablex.
- Betz, S. K., & Stoel-Gammon, C. (2005). Measuring articulatory error consistency in children with developmental apraxia of speech. *Clinical Linguistics and Phonetics*, *19*, 53–66.
- Blumstein, S. E. (1973). *A phonological investigation of aphasic speech*. The Hague: Mouton.
- Blumstein, S. E. (2001). Deficits of speech and speech perception in aphasia. In: R. S. Berndt (Ed.), *Handbook of neuropsychology* (2nd ed., Vol. 3, pp. 95–113). The Netherlands: Elsevier.
- Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain and Language*, *9*, 153–170.
- Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, *15*, 371–383.
- Boersma, P., & Levelt, C. (2003). Optimality theory and phonological acquisition. *Annual Review of Language Acquisition*, *3*, 1–50.
- Broad, D. (1979). The new theories of vocal fold vibration. In: N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 2, pp. 203–256). New York: Academic Press.
- Bunton, K., & Weismer, G. (2002). Segmental level analysis of laryngeal function in persons with motor speech disorders. *Folia Phoniatica et Logopaedica*, *54*, 223–239.
- Canter, G. J., Trost, J. E., & Burns, M. S. (1985). Contrasting speech patterns in apraxia of speech and phonemic paraphasia. *Brain and Language*, *24*, 204–222.
- Caruso, A. J., & Burton, E. K. (1987). Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, *30*, 80–87.

- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27, 207–229.
- Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, 108, 759–788.
- Dinnsen, D. A., & O'Connor, K. M. (2001a). Typological predictions in developmental phonology. *Journal of Child Language*, 28, 597–628.
- Dinnsen, D. A., & O'Connor, K. M. (2001b). Implicationally related error patterns and the selection of treatment targets. *Language, Speech and Hearing Services in the Schools*, 32, 257–270.
- Docherty, G. J., & Ladd, D. R. (Eds), (1992). *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press.
- Dodd, B. (1995). Children's acquisition of phonology. In: B. Dodd (Ed.), *Differential diagnosis and treatment of speech disordered children* (pp. 21–48). London: Whurr.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159–161.
- Duffy, J. R., & Dawle, C. A. (1984). Apraxic speakers' vowel duration in consonant–vowel–consonant syllables. In: J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 167–196). San Diego, CA: College-Hill Press.
- Elbert, M., Dinnsen, D. A., & Weismer, G. (Eds), (1984). Phonological theory and misarticulating child. *ASHA Monographs*, 22, 1–91.
- Fowler, C. A. (1985). Current perspectives on language and speech production: A critical overview. In: R. G. Daniloff (Ed.), *Speech science* (pp. 193–278). San Diego, CA: College-Hill Press.
- Gandour, J., & Dardarananda, R. (1984a). Voice-onset time in aphasia: Thai. II. *Brain and Language*, 18, 389–410.
- Gandour, J., & Dardarananda, R. (1984b). Prosodic disturbances in aphasia: Vowel length in Thai. *Brain and Language*, 23, 177–205.
- Gandour, J., Ponglorpisit, S., Khunadorn, F., Dechongkit, S., Boongird, P., Boonklam, R., & Potisuk, S. (1992). Lexical tones in Thai after unilateral brain damage. *Brain and Language*, 43, 275–307.
- Gierut, J. (1998). Treatment efficiency: Functional phonological disorders in children. *Journal of Speech, Language, and Hearing Research*, 41, S85–S100.
- Gierut, J. A., & Morrisette, M. L. (2005). The clinical significance of optimality theory for phonological disorders. *Topics in Language Acquisition*, 25, 266–279.
- Grunwell, P. (1981). The development of phonology: A descriptive profile. *First Language*, 3, 161–191.
- Grunwell, P. (1987). *Clinical phonology* (2nd Ed.). London: Croom Helm.

- Gussenhoven, C., & Warner, N. (Eds), (2002). *Laboratory phonology 7*. Berlin: Mouton de Gruyter.
- Hall, P. K., Jordan, L. S., & Robin, D. A. (1993). *Developmental apraxia of Speech: Theory and clinical practice*. Austin, TX: Pro-Ed.
- Hewlett, N., & Waters, D. (2004). Gradient change in the acquisition of phonology. *Clinical Linguistics and Phonetics*, 18, 523–533.
- Hirose, H. (1976). Posterior cricoarytenoid as a speech muscle. *Annals of Otology, Rhinology and Laryngology*, 85, 334–342.
- Houston, D. M. (2005). Speech perception in infants. In: D. B. Pisoni, & R. E. Remez (Eds), *The Handbook of Speech Perception* (pp. 417–448). Oxford: Blackwell.
- Hyman, L. M. (1975). *Phonology: Theory and Analysis*. New York: Holt, Rhinehart, and Winston.
- Ingram, D. (1976). *Phonological disability in children*. New York: Elsevier.
- Itoh, M., Sasanuma, S., Tatsumi, I. F., Murakami, S., Fukusaki, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. *Brain and Language*, 17, 193–210.
- Itoh, M., Sasanuma, S., & Ushijima, T. (1979). Velar movements during speech in a patient with apraxia of speech. *Brain and Language*, 7, 227–239.
- Iwata, R., & Hirose, H. (1976). Fiberoptic acoustic studies of Mandarin stops and affricates. *Annual Bulletin Research Institute Logopedics and Phoniatrics*, 10, 47–60.
- Jeng, J.-Y., Weismer, G., & Kent, R. D. (2006). Production and perception of Mandarin tone in adults with cerebral palsy. *Clinical Linguistics and Phonetics*, 20, 67–87.
- Johnson, A. F., Shelton, R. L., Ruscello, D. N., & Arndt, W. B. (1979). A comparison of two articulation treatments: Acquisition and acquisition-automatization. *Human Communication*, 4, 337–348.
- Johnson, K., & Ralston, J. V. (1994). Automaticity in speech perception: Some speech/nonspeech comparisons. *Phonetica*, 51, 195–209.
- Kent, R. D., & Adams, S. G. (1989). The concept and measurement of coordination in speech disorders. In: S. A. Wallace (Ed.), *Perspectives on the coordination of movement* (pp. 415–450). Amsterdam: Elsevier.
- Kent, R. D., & Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research*, 26, 231–249.
- Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In: D. D. Boysson-Bardies, S. D. Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 259–274). Dordrecht: Kluwer.

- Kurowski, K. M., Blumstein, S. E., & Mathison, H. (1998). Consonant and vowel production of right hemisphere patients. *Brain and Language*, *63*, 276–300.
- Kurowski, K. M., Hazen, E., & Blumstein, S. E. (2003). The nature of speech production impairments in anterior aphasics: An acoustic analysis of voicing in fricative consonants. *Brain and Language*, *84*, 353–371.
- Liberman, P., Kako, E., Friedman, J., Tajchman, G., Feldman, L. S., & Jiminez, E. B. (1992). Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language*, *43*, 169–189.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word*, *20*, 384–442.
- Locke, J. L. (1980a). The inference of speech perception in the phonologically disordered child, part I: A rationale, some criteria, the conventional tests. *Journal of Speech and Hearing Disorders*, *45*, 431–434.
- Locke, J. L. (1980b). The inference of speech perception in the phonologically disordered child, part II: Some clinically novel procedures, their use, some findings. *Journal of Speech and Hearing Disorders*, *45*, 435–468.
- Locke, J. (1983). *Phonological Acquisition and Change*. New York: Academic Press.
- Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics*, *8*, 475–490.
- Löfqvist, A. (1997). Theories and models of speech production. In: W. J. Hardcastle, & J. Laver (Eds), *The Handbook of phonetic sciences* (pp. 405–426). Oxford: Blackwell.
- Löfqvist, A., Baer, T., & Yoshioka, Y. (1981). Scaling of glottal opening. *Phonetica*, *38*, 236–251.
- McReynolds, L. V. (1988). Articulation disorders of unknown etiology. In: N. J. Lass, L. V. McReynolds, J. L. Northern, & D.E. Yoder (Eds), *Handbook of speech-language pathology and audiology* (pp. 419–441). Toronto: B.C. Decker.
- Menn, L. (1980). Phonological theory and child phonology. In: G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds), *Child Phonology* (Vol. 1, pp. 23–42). New York: Academic Press.
- Mergell, P., Herzel, H., & Titze, I. R. (2000). Irregular vocal fold vibration: High-speed observation and modeling. *Journal of the Acoustical Society of America*, *108*, 2996–3002.
- Morrisette, M. L., & Gierut, J. A. (2002). Lexical organization and phonological change in treatment. *Journal of Speech, Language, and Hearing Research*, *45*, 143–159.
- Ní Chasiade, A., & Gobl, C. (1997). Voice source variation. In: W. J. Hardcastle, & J. Laver (Eds), *The handbook of phonetic sciences* (pp. 427–461). Oxford: Blackwell.
- Nittrouer, S. (1999). Do temporal processing deficits cause phonological processing problems? *Journal of Speech, Language, and Hearing Research*, *42*, 925–942.

- Platt, L. J., Andrews, G., & Howie, P. M. (1980). Dysarthria of adult cerebral palsy. II. Phonemic analysis of articulatory errors. *Journal of Speech and Hearing Research*, 23, 41–55.
- Rosenbek, J. C., Kent, R. D., & LaPointe, L. L. (1984). Apraxia of speech: An overview and some perspectives. In: J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 1–72). San Diego, CA: College-Hill Press.
- Ryalls, J., Hoffman-Ruddy, B., Vitek, J., & Owens, A. (2001). A preliminary study of the effects of pallidotomy surgery upon voice onset time in persons with Parkinson's disease. *Clinical Linguistics and Phonetics*, 15, 107–111.
- Ryalls, J., Provost, H., & Arseneault, N. (1995). Voice onset time production in French-speaking aphasics. *Journal of Communication Disorders*, 28, 205–215.
- Shewan, C. M., Leeper, H. A., Jr., & Booth, J. C. (1984). An analysis of voice onset time (VOT) in aphasic and normal subjects. In: J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 197–220). San Diego, CA: College-Hill Press.
- Shinn, P., & Blumstein, S. E. (1983). Phonetic disintegration in aphasia: Acoustic analysis of spectral characteristics for place of articulation. *Brain and Language*, 20, 90–114.
- Shriberg, L. D., & Kwiatkowski, J. (1980). *Natural process analysis (NPA)*. New York: Wiley.
- Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research*, 36, 533–547.
- Smit, A. B., Hand, L., Frelinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55, 779–798.
- Stampe, D. (1973). *A dissertation on natural phonology*. Unpublished Ph.D. dissertation, University of Chicago, IL.
- Stathopoulos, E. T., & Weismer, G. (1983). Closure duration of stop consonants. *Journal of Phonetics*, 11, 395–400.
- Stokes, S. F., & Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language, and Hearing Research*, 48, 577–591.
- Straight, H. S. (1980). Auditory versus articulatory phonological processes and their development in children. In: G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds), *Child phonology* (Vol. 2, pp. 43–71). New York: Academic Press.
- Tuller, B. (1984). On categorizing aphasic speech errors. *Neuropsychologia*, 22, 547–557.
- Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, 61, 846–858.
- Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. *Journal of Phonetics*, 8, 427–438.



- Weismer, G. (1997). Motor speech disorders. In: W. J. Hardcastle, & J. Laver (Eds), *The handbook of phonetic sciences* (pp. 191–219). Oxford, UK: Blackwell.
- Weismer, G. (2004). VOT in speech-disordered individuals: History, theory, data, reminiscence. Paper presented at the 147th meeting of the Acoustical Society of America, New York, NY.
- Weismer, G., Dinnsen, D. A., & Elbert, M. (1981). A study of the voicing distinction associated with omitted word-final stops. *Journal of Speech and Hearing Disorders*, 46, 320–327.
- Weismer, G., Yunusova, Y., & Westbury, J. R. (2003). Interarticulator coordination in dysarthria: An X-ray microbeam study. *Journal of Speech, Language, and Hearing Research*, 46, 1247–1261.
- Williams, G., & McReynolds, L. V. (1975). The relationship between discrimination and articulation training in children with misarticulations. *Journal of Speech and Hearing Research*, 18, 401–412.
- Wittenberg, T. (1997). Automatic motion extraction from laryngeal kymograms. In: T. Wittenberg, P. Mergell, M. Tigges, & U. Eysholdt (Eds), *Advances in quantitative laryngoscopy* (pp. 21–28). Gottingen: Verlag Abteilung Phoniatrie.
- Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and germination in American English. *Journal of the Acoustical Society of America*, 70, 1615–1623.
- Ziegler, W. (1984). What can the spectral characteristics of stop consonants tell us about the realization of place of articulation in Broca's aphasia? A reply to Shinn and Blumstein. *Brain and Language*, 23, 167–170.

## Chapter 5

### **Functional Neuroimaging of Speech Production**

*Thomas A. Zeffiro and Jennifer L. Frymiare*

#### **1. INTRODUCTION**

In a relatively short period of time functional brain imaging has proven to be a uniquely effective means to study the role of cortical and subcortical brain structures in the production and understanding of language. By facilitating measurement of brain activity related to the processes underlying communication, it has allowed examination of the regional functional specialization of these phenomena with an accuracy and precision not previously possible with chronometric or electrophysiological techniques. With its ability to non-invasively localize neural activity, it has provided some remarkable opportunities to develop a deeper understanding of language processing in health and disease. Based on a host of elegant experimental designs previously developed by experimental psychologists, the addition of neuroimaging techniques to language research has allowed independent evaluation of theories developed to explain the mechanisms responsible for this most uniquely human activity. Most recently, its combination with various electrophysiological approaches has broadened the scope of inquiry to allow models of language processing that capture not only the spatial distribution of cognitive phenomena, but also their temporal character, providing novel glimpses of the flow of information during communication. The rapid pace of technological advance in functional neuroimaging methods shows no signs of abating, as the limits to our ability to resolve neural activity on ever finer spatial and temporal scales have been more constrained by engineering and financial considerations than physical ones.

During the relatively brief time that systems allowing efficient and non-invasive monitoring of brain activity have been widely available, investigators utilizing functional neuroimaging have relied on a series of related techniques. Most of these methods rely on the tight coupling between neuronal activity and regional cerebral blood flow first noted at the end of the 19th century, based on the observation that focal electrical cortical stimulation resulted in localized changes in cerebral hemodynamics (Roy, 1890). This coupling of brain blood flow and neural activity allows inferences about local neural

activity changes to be derived from sequential measures of regional hemodynamic activity recorded while subjects engage in tasks designed to isolate particular cognitive or perceptual components of interest. Although the basic idea of using brain hemodynamic modulations as a means to study the neural mechanisms responsible for language and communication is relatively simple, it took a number of years to develop practical means to utilize neurovascular coupling mechanisms to probe the distribution and timing of mental operations.

The first successful single photon emission computed tomography (SPECT) imaging system based on these phenomena utilized inhalation of radioactive  $^{133}\text{Xe}$ , which served as a relatively non-invasive marker for cerebral blood flow (Obrist, Thompson, Jr., King, & Wang, 1967). Some of the earliest applications of this technique involved language experiments (Larsen, Skinhoj, & Lassen, 1978; Larsen, Skinhoj, Soh, Endo, & Lassen, 1977) and their publication generated enormous interest among experimental psychologists who immediately saw the great potential this technique held, as even with relatively limited spatial and temporal resolving power it was possible to clearly confirm notions of hemispheric specialization for language that had been based on much more indirect approaches for assessing the localization of cognitive activity. This early work using  $^{133}\text{Xe}$  tracer techniques was quickly followed by experiments that more clearly delineated the spatial character of language processing, with the emergence of positron emission tomography (PET) imaging systems that could use  $\text{H}_2^{15}\text{O}$  as an inert, diffusible tracer that remained in the intravascular compartment of the brain long enough to allow localized measurement of regional cerebral blood flow (rCBF) (Herscovitch, Markham, & Raichle, 1983; Raichle, Martin, Herscovitch, Mintun, & Markham, 1983). The use of  $\text{H}_2^{15}\text{O}$  as a tracer, with its inherent short half-life of 120 seconds, allowed repeated measurements of subjects under multiple conditions, and therefore permitted sensitive within-subject comparisons among tasks involving different combinations of unobserved cognitive components. Language studies using rCBF as a dependent measure began to rapidly appear and many took advantage of the fact that overt responses, including speech, were relatively compatible with this experimental technique (Bookheimer, Zeffiro, Blaxton, Gaillard, & Theodore, 1995, 2000; Frith, Friston, Liddle, & Frackowiak, 1991; Price et al., 1996b). The principal limitation of this method, with respect to cognitive studies, was the limitation in task types that could be studied in each individual participant, due to dosimetry limits on the amount of radioactive tracer that could be used.

The development of functional magnetic resonance imaging (fMRI) systems suitable for psychological studies provided a more convenient and completely non-invasive means to study perception, language and action. In the earliest studies, bolus injections of exogenous magnetic contrast agents were used to identify hemodynamic changes related to visual processing (Belliveau et al., 1991). Soon thereafter, it was discovered that circulating deoxyhemoglobin could act as an endogenous contrast agent. Because of the tight coupling between neural activity and blood oxygenation, blood oxygenation-level dependent (BOLD) contrast became a popular means to estimate hemodynamic change in relation to brain activity (Kwong et al., 1992; Ogawa et al., 1992), rapidly

replacing rCBF PET methods as the most common means to study the neural basis of psychological processes.

A notable exception to this shift in preferred technology to BOLD-contrast was seen among language and communication investigators interested in having participants perform tasks that required understanding or producing speech. Their concerns involved two aspects of the data acquisition process: (1) that the acoustic noise produced by the scanner might cause interference, and (2) that jaw motion associated with speech might result in image artifacts that would obscure the accurate detection of the regional pattern of task-related neural activity. In combination, these properties of fMRI techniques provide a particular challenge for studies in which subjects are required to produce speech responses or discriminate auditory stimuli. While fMRI is an effective and versatile neuroimaging technique, when based on echo-planar imaging sequences, its intrinsic characteristics limit its use in contexts involving speech responses. In particular, echo-planar image quality is exquisitely sensitive to subject's jaw and head movement and these movements may induce both false-positive and false-negative artifacts in statistical maps of task-related activity (Friston, Williams, Howard, Frackowiak, & Turner, 1996). These effects have led some investigators to limit the application of this technology to tasks in which subjects remain mute and otherwise still. As a result, experimental designs that require the subjects to speak inside the magnet have been limited in number, although attempts have been made to compensate for these unwanted effects during data analysis (Barch, et al., 1999; Birn, Cox, & Bandettini, 2004).

In this chapter, we will explore those concerns, review the approaches that have been taken to minimize the possibility that scanner noise or subject motion will result in artifactual results and finally make practical suggestions for performing successful functional imaging studies in the domains of language and communication.

### **1.1. Acoustic Noise Effects in Functional MRI**

Normal operation of an MRI system during a functional imaging experiment is associated with loud repetitive sounds that arise from the interaction of the gradient coils and the system's static magnetic field. This "ringing" of the gradient coils during the read-out phase of the echo-planar imaging sequence results in periodic, narrow frequency band sounds that can reach 138 dB at the isocenter of the magnet (Ravicz, Melcher, & Kiang, 2000). The rate of the noise pulses is determined by the speed at which the slice select gradient is incremented, usually varying between 10 and 15 times per second. Attempts by MRI system manufacturers to develop designs to substantially reduce the gradient noise during echo-planar imaging have met with limited success, involving mechanical isolation of the subject table from the magnet, addition of sound-absorbing materials around the gradients, modification of the system covers to reduce acoustic resonance and development of special pulse sequences less likely to induce gradient vibration. Moreover, active noise-cancellation techniques, so effective in environments contaminated by low-frequency, wider-bandwidth noise (such as jet aircraft and automobile cabins) have proven to be ineffective in reducing the high-frequency, high-amplitude, narrow-band noise associated with fMRI experiments.

However, noise-cancellation techniques are useful in extracting speech signals from background noise in the service of recording subject responses (Jung, Prasad, Qin, & Anderson, 2005). Although the previously mentioned design modifications can reduce the background to around 105 dB above threshold, this level is still loud enough to significantly affect auditory perception. As there has been some concern about damage to the peripheral auditory apparatus with extended exposure at these sound levels (Glover et al., 1995), it is customary to require experimental participants to wear sound attenuation devices, either attenuating earplugs or headphones with good sound isolation properties (Ravicz & Melcher, 2001).

Therefore, while fMRI is arguably a remarkably sensitive method for detecting task-related signal change associated with perceptual and cognitive processes, its use is invariably associated with an unavoidably noisy experimental environment. It follows that detection of auditory stimuli in these studies may be adversely affected in numerous ways. First, studies involving spoken language comprehension or speech sound discrimination are likely to suffer significant stimulus masking such that the subject may be unable to hear the auditory stimuli clearly, a situation likely to result in performance different from that expected if behavioral piloting has been performed in a quiet environment (Barch et al., 1999; Nelles et al., 2003). Second, during a demanding auditory discrimination task, the background noise generated by the scanner may induce additional task difficulty, and these effects may not be consistent across task types, again altering the nature of the task significantly from the investigator's original intent. Third, even in the case of tasks that do not utilize auditory stimuli, it is possible that the loud and intermittent gradient noise may distract the subject from the task at hand, modulating selective attention and performance with unpredictable consequences (Novitski et al., 2001, 2003).

An even more insidious potential confound can arise from the fact that many studies using fMRI to map the functional specialization of human cortex are based on the principle of cognitive insertion. This approach, used in much of functional imaging research, assumes that subtractive logic adheres and that the regional neural activity related to the specific cognitive function can be identified by contrasting two tasks that differ only in the degree to which the cognitive construct of interest is engaged. Although this assumption of linear additivity of auditory perceptual processes may be questioned in general (Talavage & Edmister, 2004), in this particular context the assumption is usually made that the neural activity associated with the gradient noise, present in conditions on both sides of the statistical contrast, will "subtract out" and therefore not confound the identification of the cognitive or perceptual component of interest. Therefore, it is assumed that the physiological changes induced by the acoustic gradient noise are similar enough in both task and comparison conditions that their respective effects will disappear as a result of the subtraction. Should this "subtraction" assumption be incorrect, noise-induced signal change may be misclassified as task-related signal change. Even if the assumption of subtractive logic with respect to cognitive processes is sound, its use for auditory studies could still be problematic. For example, the range of loudness between the resting and activated states in auditory cortex may be artificially decreased in fMRI experiments in

situations in which the contaminating gradient noise reduces the available dynamic range of the auditory system, thereby reducing the “headroom” through saturation of auditory neurons by the gradient noise. Evidence for this effect comes from estimates of the neural activity resulting from echo-planar imaging sequence noise (Bandettini, Jesmanowicz, Van Kylen, Birn, & Hyde, 1998).

Finally, it is possible that sensory stimulation causing activity in one cortical area can be associated with decreases in other cortical areas. During functional imaging experiments using echo-planar imaging, stimulation of the auditory cortical areas may result from the periodic narrow-band noise generated by the gradients. In this regard, significant decreases in rCBF have been detected with PET in auditory cortex during a visual task (Haxby et al., 1994). It is possible that the reduction observed in auditory cortical areas resulted from selective attentional processes, reducing the response to unattended auditory stimuli as attention is focused on the visual stimuli (Shulman et al., 1997). This unexpected auditory cortex modulation is inconsistently observed across cortical areas and among different studies, with the nature of the effects strongly dependent on the specific nature of the task. In the case of gradient noise, unintended auditory stimulation induced by fMRI is likely to promote modulations in areas not thought to be affected by auditory stimuli.

More direct evidence for acoustic contamination by gradient noise has come from fMRI studies of visual and motor tasks in which participants are exposed to stimuli in conditions of variable noise levels (Cho, Chung, Lim, & Wong, 1998). In these studies, gradient noise had opposing effects on sensory and motor cortical areas, enhancing motor cortical activity during a motor task and reducing activity in primary visual cortex during a visual task. These effects have been explained by other investigators as resulting from modulation of early sensory processing by potent “top-down” influences (Fiez et al., 1995; Shulman et al., 1997).

In summary, acoustic noise contamination is an unavoidable consequence of the operation of current MRI devices. Its effects are complex, potentially including stimulus masking, task difficulty modulation, interference with selective attention and reduction in auditory system dynamic range. Insofar as the assumptions of subtractive logic do not pertain to the design of a specific experiment examining the processes responsible for language production or comprehension, the effects of scanner gradient noise may limit the accuracy with which neural activity is estimated in a functional neuroimaging experiment.

## **1.2. Motion Effects in Functional MRI**

Of all the sources of physiological noise in a functional imaging experiment, motion of the head in all its forms is the most significant. As the techniques for its amelioration after the image data are collected are incomplete in their effectiveness, prevention or limitation of excessive motion during the phase of data acquisition is the best way to achieve the highest degree of specificity and sensitivity in functional brain imaging. The methods available to limit or prevent subject head motion in imaging experiments are myriad,

including: (1) insertion of foam cushions between the subject and the head coil, (2) restraint with a custom thermoplastic mask, (3) on-line motion feedback provided to the subject (Yang, Ross, Zhang, Stein, & Yang, 2005), (4) prior subject compliance training (Slifer, Cataldo, Cataldo, Llorente, & Gerson, 1993), and (5) relatively rigid coupling of the subject to the head coil with a bite-bar fashioned from dental acrylic. These approaches are differentially effective and range widely in practicality and comfort. Although all work reasonably well in reducing rigid-body motion of the head, none reduce the susceptibility artifacts arising from articulator motion or the parenchymal brain motion arising from cardiovascular and respiratory effects. As the overall accuracy of fMRI is principally limited by inter-scan head motion artifacts present in statistical maps of task-related brain activity, and the existing post-processing algorithms for image realignment have not been completely successful when applied to time series of echo-planar images, it is obvious that better methods to prevent subject motion of all kinds are needed.

Subject motion relevant to imaging studies involving language and communication can be categorized into three types: (1) rigid-body motion of the head, (2) parenchymal motion of the brain, and (3) motion of the articulators, particularly the jaw.

*Rigid-body motion* refers to the translational and rotational changes the head can make, even when the subject is comfortably supine on the scanner table. Relatively small degrees of translational or rotational motion of the head result in misregistration of sequentially collected brain volumes, resulting in signal intensity changes related to spatially varying partial volume effects. Speech responses are particularly likely to cause head motion, resulting from the mechanical coupling of the jaw and skull such that relatively small jaw movements can result in large translational movements of the skull. Since even small amounts of interscan motion in a time series can result in large artifacts in statistical maps derived from that series, it is best to take precautions to minimize head motion. Although the motion is global, the effects of the motion are regionally specific, being most prominent in regions of variable tissue contrast. Examples include boundaries between gray matter and cerebrospinal fluid. This may result in an easily appreciated “rim” artifact around the edge of the brain in statistical maps generated from time series with excessive interscan motion. If peak-to-peak rigid-body motion exceeds 5–10% of the image voxel width, statistical maps are likely to exhibit obvious motion artifacts. Therefore, assuming constant head motion, statistical map artifact will increase with increasing spatial resolution and decreased voxel width.

In spite of strenuous efforts on the part of the investigators to prevent subject motion, tasks requiring speech responses are inevitably associated with some head movement. This problem is likely to be especially severe in studies requiring articulation. Determination of the magnitude of these unwanted movements provides information useful in deciding whether or not to employ techniques to compensate for the motion. To this end, head motion can be estimated by computing the motion of the center-of-intensity of each image volume. By comparing this location across time points, it is possible to estimate rigid-body motion of the head during the scanning interval. If the

motion detection process reveals head motion greater than 0.10 pixel width in any dimension, then head motion correction is performed using reregistration algorithms (Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998a; Woods, Grafton, Watson, Sicotte, & Mazziotta, 1998b). These algorithms allow rigid-body transformations guided by a least-squares misregistration minimization technique. Having determined that sufficient head motion is present to warrant correction, it is possible to employ these automated realignment procedures to determine the coordinate transformation that will bring the members of the time series back into register (see Figure 1). This is usually the most time consuming and computationally intensive part of the entire analysis procedure. Having determined the appropriate coordinate transformation to reregister the volumes, a resampling algorithm is employed to generate the realigned image volume. For this procedure there is a trade-off between time and accuracy, with the most accurate resampling procedures (sinc interpolation) being significantly slower than the less accurate procedures (nearest neighbor interpolation). Although relatively effective, these techniques correct only for rigid-body motion.

Even in the absence of rigid-body head motion, regionally varying *parenchymal motion* of the brain can produce significant artifacts in statistical parametric maps. This

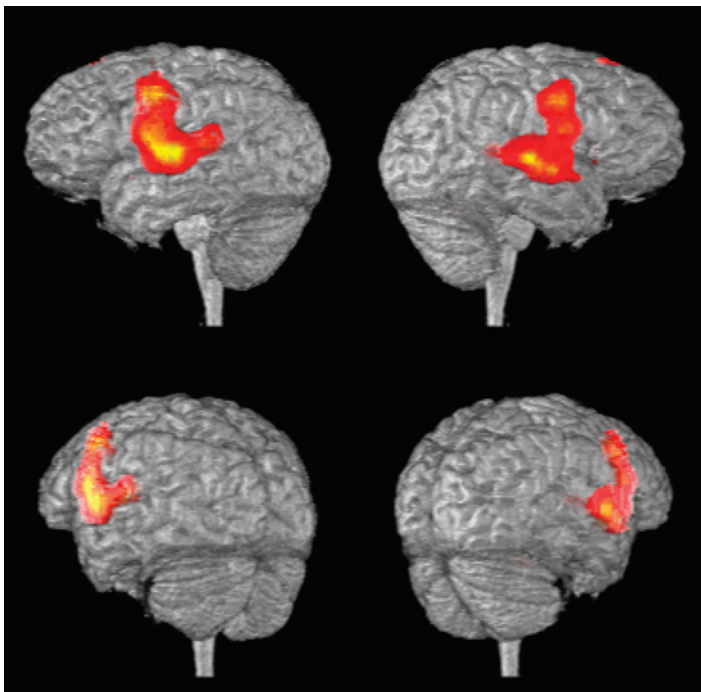


Figure 1. Time plots of rigid-body head motion during a reading task comparing silent and aloud reading. Tasks, such as reading aloud, that associated with jaw and tongue movements invariably are associated with more inter-scan head motion than their silent counterparts.



parenchymal motion results from an interaction between the viscoelastic properties of the brain tissue and local pressure changes induced by arterial and venous pressure modulations of cardiac and respiratory origin (Poncelet, Wedeen, Weisskoff, & Cohen, 1992). These effects result in relatively large MR signal changes at frequencies above the sampling rates customarily employed in fMRI studies (0.25–0.5 Hz) and therefore appear as aliased noise in lower frequency components of the fMRI time series. This aliased noise has most prominent effects in the regions of the derivative statistical maps around the ventricles, where the viscoelastic parenchymal motion is most prominent. Attempts have been made to reduce these effects by designing digital filters that attenuate signal at the appropriate frequencies (Biswal, DeYoe, & Hyde, 1996; Hu, Le, Parrish, & Erhard, 1995; Le & Hu, 1996).

*Motion of the articulators*, particularly those that involve prominent jaw motion can result in image artifacts that are extremely difficult to remove during post-processing. The frequent observation that image artifacts related to motion of the articulatory apparatus may result in prominent false-positive signals led many investigators to conclude, perhaps prematurely, that fMRI was not a suitable technique for studying the neural mechanisms of speech, as experimental designs requiring verbal responses are necessarily associated with jaw movements that can induce noticeable susceptibility artifacts in medial and inferior temporal cortical regions. As these areas are involved in language, it is obvious that jaw motion can generate potentially serious image artifacts in regions that would be expected to show task-related signal change. Other artifacts resulting from the susceptibility changes that are induced by jaw movements include signal loss and image warping in nearby brain regions. These will vary with the position of the articulators responsible for the production of a given word. Spoken responses are also accompanied by tongue movement and swallowing, which provide an additional source of local susceptibility artifact as the shape of the oral cavity is continuously transformed (Birn, Bandettini, Cox, & Shaker, 1999). All of these effects are somewhat paradoxical in that they can occur well out of the acquisition field of view, resulting from the large changes in magnetic susceptibility related to modulation of the shape and volume of the oral cavity during jaw movement.

To summarize, there are a number of sources of motion in functional neuroimaging experiments, each contributing in a specific way to an overall degradation in the efficacy of signal detection and estimation. Although prevention is to be preferred by a wide margin, mitigation of motion effects can sometimes be an effective strategy.

### **1.3. Why Require Overt Responses in Language Experiments?**

Given the difficulties in executing experiments that require spoken language, one might ask why it is advisable to undertake investigations requiring this type of overt response. If the neural mechanisms of overt and covert speech responses are sufficiently similar, or dissociable by subtractive analysis, it might be possible to proceed assuming that the neural correlates of overt speech production are not unlike those of covert speech except for the associated motor and auditory processes. However, several PET and fMRI

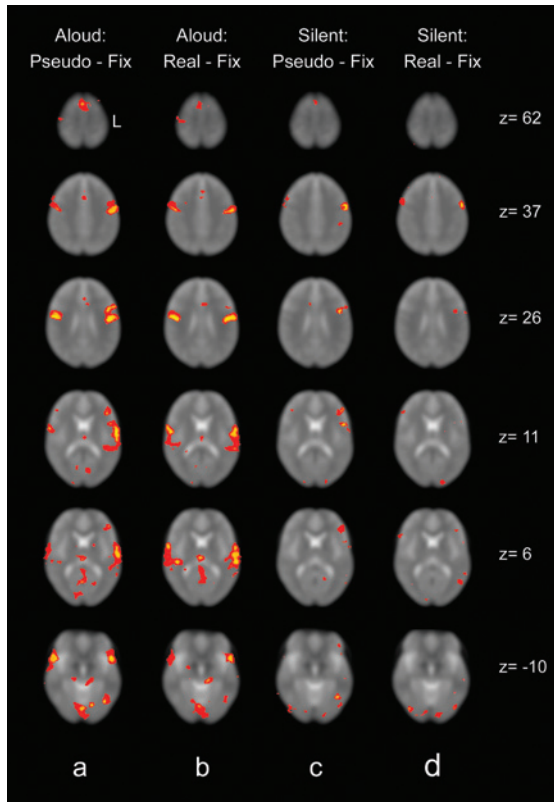


Figure 2. Patterns of activation associated with silent and overt naming. Transverse sections illustrating areas of significant activity for four task conditions relative to a fixation baseline. From left to right: BIG Aloud, BIG Silent, Continuous Aloud, and Continuous Silent. On the right are the  $z$ -coordinates corresponding to the atlas of Talairach and Tournoux (Talairach & Tournoux, 1988). For each of the task runs the activation maps ( $Z > 3.1$ ,  $P < 0.001$ ) for the Time, Volumes, and Stimuli data sets are shown projected onto the same brain template. Note that the regions of activity portray task-related signal changes having a threshold minimum of  $Z > 3.1$ , and do not provide information about the magnitude of the activity above this threshold. Yellow = Time, Blue = Volumes, Red = Stimuli, White = all three.

studies have demonstrated that tasks with speech responses are associated with different patterns of brain activity when compared to their silent counterparts as shown in Figure 2. Consequently, subtractive logic does not seem to be particularly effective and silent responses in language paradigms do not offer a workable substitute for overt responses (Barch et al., 1999; Bookheimer et al., 1995; Dietz, Jones, Gareau, Zeffiro, & Eden, 2005; Huang, Carr, & Cao, 2002; Palmer et al., 2001; Price, Wise, & Frackowiak, 1996a; Rumsey et al., 1997). In order to study the brain mechanisms responsible for speech comprehension and production alternative methods that minimize the effects of gradient noise or subject motion are warranted.

#### 1.4. Alternative Techniques for Functional MRI during Speech

Fortunately the existence of a sizable delay between changes in the recorded BOLD-contrast local signal change and its antecedent neural activity allow experimental strategies in which the auditory stimulus can be presented, or the speech response recorded, after the gradients have fallen silent. While usually said to limit temporal resolution in functional neuroimaging studies, it is the existence of hemodynamic delay and dispersion that allow the interleaving of data acquisition and behavior, permitting task performance under relatively quiet experimental conditions with little acoustic interference. Because the hemodynamic response builds as the neuronal activity associated with task performance continues, eight second task performance periods are sufficiently long to yield excellent single-subject activity maps. Many acquisition types are based on this lag between stimulus presentation or task performance and the associated hemodynamic response that is the source of the BOLD-contrast signal (Buckner et al., 1996; Kwong et al., 1992). The onset of the BOLD-contrast response occurs approximately 2–5 s after stimulus presentation, peaks around 5–6 s, and returns to baseline about 10–12 s after stimulus offset (Belliveau et al., 1991). Several approaches to data acquisition take advantage of the delay in BOLD-contrast response to avoid contaminating head motion and acoustic noise: clustered volume acquisition (Edmister, Talavage, Ledden, & Weisskoff, 1999), sparse temporal sampling (Hall et al., 1999), event-related (single-trial) designs (Birn et al., 1999; Huang et al., 2002), and the Behavior Interleaved Gradients (BIG) technique (Eden, Joseph, Brown, Brown, & Zeffiro, 1999). All of these techniques employ temporal separation of task performance and image acquisition and rely on the fact that the neural activity evoked by acoustic gradient noise is separable from that related to perception of the intended stimulus (Talavage, Edmister, Ledden, & Weisskoff, 1999). On the other hand, these methods also have their drawbacks. The primary disadvantage being that, when compared to the more conventional continuous acquisition methods, fewer images are acquired per unit time, which, all other conditions being equal, means reduced statistical power.

Recently, Birn et al. (2004) systematically examined how acquisition of functional imaging data can be optimized by taking into account the temporal delay between the immediate motion-induced signal change and the more slowly generated hemodynamic signal change. Results from their simulations and experiments involving reading aloud demonstrate some advantages in discarding the images acquired during speech in brain regions prone to exhibit the effects of jaw motion artifact. However, removing segments of the time series in this way also results in an overall decrease in detection sensitivity over the entire brain volume, so the respective contributions of these competing effects must be kept in mind when selecting the details of the analysis procedure.

A different approach involves acquisition techniques in which task performance and image acquisition are interleaved in time. In order to mitigate the effects of gradient noise, these techniques employ an approach in which the gradients are switched off during periods of task execution and then immediately switched back on to allow sampling of the resulting hemodynamic activity. Detection of the task-related signal change relies

on the presence of a neurovascular coupling lag of 4–8 s between stimulus onset and the resulting BOLD-contrast response. In response to neuronal activity, there is an increase of blood flow and oxygen delivery (Kwong et al., 1992; Ogawa et al., 1992). Because the oxygen utilization increase is far less than the blood flow increase in relation to neural activity, there is a net deoxyhemoglobin decrease, resulting in a decrease in local spin dephasing and therefore an increase in the MRI signal, the BOLD-contrast response. However, this signal modulation is shifted in time by 4–8 s, with the MRI signal modulation exhibiting both delay and dispersion relative to its antecedent neural activity. These effects result in a loss of temporal information in the resulting measurements. However, because the general time of the delay is known, it is possible to use this information to capture the resulting BOLD signal excursion in a specific temporal interval.

Interleaved acquisition is a member of the larger class of evoked hemodynamic response techniques, also referred to as event-related fMRI techniques, that combine relatively brief stimulus presentation or task performance times with sufficient temporal sampling to capture the shape of the evoked hemodynamic response. In comparison with other event-related approaches, the interleaved techniques tend to employ longer repetition times (TR), allowing maximal longitudinal relaxation and therefore providing greater MRI signal contrast and potentially greater sensitivity to small task-related signal changes. While conventional methods of fMRI data acquisition data usually operate continuously, in the interleaved techniques the gradients are off during periods of task execution and subsequently activated to acquire data after the task interval has completed. These interleaved techniques utilize the same basic pulse sequence parameters as those used in conventional continuous fMRI acquisitions, differing only in the addition of a 9–12 s gap between the onsets of successive acquisitions. By interleaving task

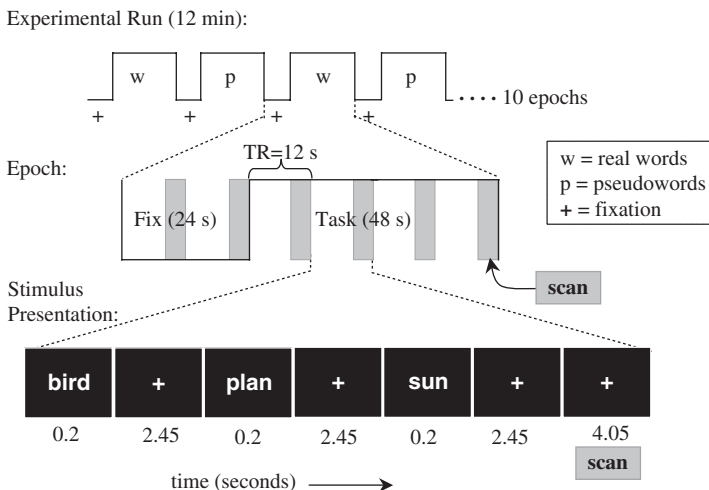


Figure 3. Timing diagram of an interleaved data acquisition experiment.

performance and data acquisition, the interleaved techniques allow the subject to perform the tasks under relatively quiet experimental conditions. The details of the event timing are shown in Figure 3.

### 1.5. Advantages and Disadvantages of Interleaved Acquisition Techniques

Acoustic gradient noise during task performance causes measurable BOLD-contrast signal modulations in both cortical and subcortical structures (Bandettini et al., 1998). In conventional continuous data acquisition procedures, the gradient noise occurs equally in the control and task conditions. Assuming that the signal changes due to gradient noise are linearly additive with other sources of signal change (e.g. task-related changes), in principle it should be possible to subtract the noise effects. In this way categorical data analysis approaches based on image subtraction, such as Student's *t*-test, can be used to identify regions that exhibit task-related activity. Although it is reasonable to assume that linear additivity for gradient noise might hold for many cortical regions, it is less likely that this assumption is reasonable for cortical areas known to be responsive to auditory stimuli (Talavage & Edmister, 2004). Non-linear effects and interference are particularly to be expected at frequencies near those generated by the gradients.

By employing a longer TR than customarily used in conventional continuous imaging, interleaved acquisition techniques have potential increased sensitivity to small signal changes, resulting from the improved contrast-to-noise that occurs with longer TR intervals as a result of more fully recovered longitudinal relaxation. However, the loss of signal due to incomplete recovery of the longitudinal magnetization with shorter TRs has been shown to be outweighed by the increased statistical power gained by the larger number of samples collected in that situation (Constable & Spencer, 2001).

When using interleaved approaches, the proportion of time spent in performance of the target task is reduced, resulting in a situation in which the subject makes relatively fewer responses during the imaging session than during a continuous acquisition experiment utilizing the same overall run length. Therefore, interleaved techniques may be less sensitive because of the smaller number of samples collected in the same acquisition period. However, the reduced power related to the smaller sample size may be balanced by the fact that, when using interleaved gradients, the subjects alternate more frequently between target and control tasks, with an attendant reduction in response habituation and therefore greater task-related signal modulation.

Also with interleaved acquisition techniques, susceptibility artifacts resulting from jaw and tongue motion can be greatly reduced. Experiments involving verbal responses are necessarily associated with orofacial movement that can induce significant susceptibility artifacts in medial temporal and orbitofrontal cortical regions. As these areas are involved in language production and understanding, it is obvious that tasks involving orofacial movement could be associated with signal drop-out or image distortion in the very regions that would be the principal objects of study. Speech is also likely to be associated

with periodic head motion, resulting from the mechanical coupling of the jaw and skull such that relatively small jaw movements can result in large rotational movements of the skull. Since even small amounts of uncorrected interscan motion can result in both false-positive and false-negative effects in statistical maps, even when using interleaved techniques it is advisable to take precautions to minimize head motion using comfortable restraints that allow unfettered jaw motion.

To illustrate the application of this method, Figure 4 shows an example of interleaved acquisition used to detect activity related to single word reading. An interleaved design was used to reduce the problem of susceptibility artifacts resulting from head and jaw movements. This is an important issue in understanding the brain mechanisms responsible for language processing, because evidence from PET studies, that are uncontaminated by motion related susceptibility artifacts, indicates that covert and overt naming responses engage different neural processing systems (Bookheimer et al., 1995). Results from a single subject are shown for the comparison of reading words silently versus fixation (bottom) and reading words aloud versus fixation (top).

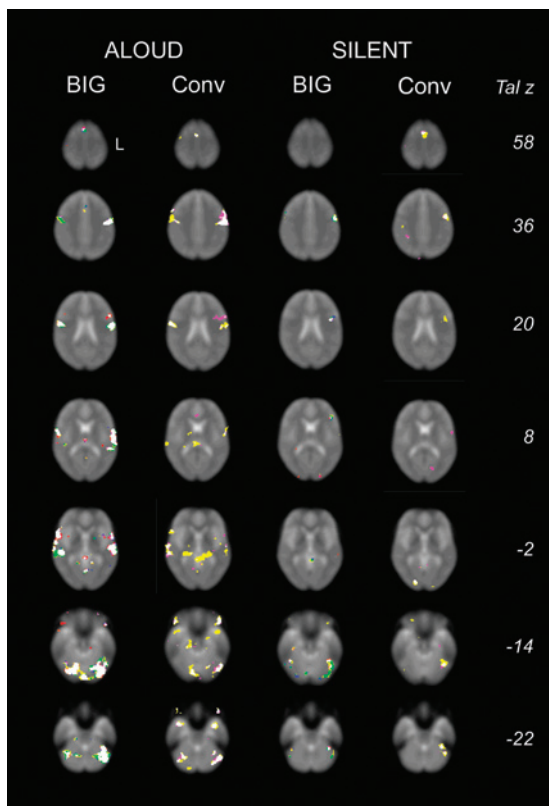


Figure 4. Interleaved and continuous acquisition compared.

Table 1  
fMRI studies employing speech responses.

Reference	Speech task	Block vs. event-related	Continuous vs. interleaved	Comments
Abrahams et al. (2003)	Verbal fluency, confrontation naming	Block	Interleaved	
Abrahams et al. (2004)	Verbal fluency, confrontation naming	Block	Interleaved	
Aldenkamp et al. (2003)	Naming task	Block	Interleaved	
Baciu, Rubin, Decorps, and Segebarth (1999)	Word fluency	Block	Continuous	Soft articulation requested; comparison of aloud and silent word fluency
Barch et al. (1999)	Stroop task, verb generation, noun reading	Block	Continuous	Avoid scanning at throat and mouth, discard data acquired during speech; comparison of aloud and silent Stroop task and noun reading
Barch Brover, Sabb, and Noll (2000)	Verb generation	Event-related	Continuous	
Barrett, Pike, and Paus (2004)	Reading before and after mood induction	Event-related	Interleaved	
Birn, Bandettini, Cox, and Shaker (1999)	Speech production	Block and event-related	Continuous	

Birn, Cox, and Bandettini (2004)	Word reading	Block and event-related	Continuous	Discard images acquired during speech
Burgund, Lugar, Miezin, and Peterson (2003)	Object naming	Event-related	Continuous	
Burton, Noll, and Small (2001)	Word repetition	Block	Continuous	
de Zubicaray, Wilson, McMahon, and Muthiah (2001)	Picture-word task	Event-related	Interleaved	
de Zubicaray, McMahon, Eastburn, and Wilson (2002)	Picture-word task	Event-related	Interleaved	
Dietz, Jones, Gareau, Zeffiro, and Eden (2005)	Word reading	Block	Interleaved	Comparison of aloud and silent word reading
Dietz, Jones, Twomey, Zeffiro, and Eden, (in press)	Word reading	Block	Interleaved and continuous	Comparison of aloud and silent word reading
Eden et al. (2004)	Word reading	Block	Interleaved	
Frenck-Mestre, Anton, Roth, Vaid, and Viallet (2005)	Word and sentence reading	Block	Continuous	
Grabowski et al. (2006)	Object naming	Event-related	Continuous	Adaptive pacing algorithm
Haller, Radue, Erb, Grodd, and Kircher (2005)	Sentence generation, word and sentence reading	Event-related	Continuous	
Hashimoto and Sakai (2003)	Sentence reading	Block	Interleaved	

*(Continued)*



Table 1  
(Continued)

Reference	Speech task	Block vs. event-related	Continuous vs. interleaved	Comments
He et al. (2003)	Chinese word and pinyin reading	Block	Continuous	Comparison of aloud and silent Chinese word and pinyin reading
Heim, Opitz, and Friederici (2002)	Picture naming, grammatical gender	Event-related	Continuous	
Huang, Carr, and Cao (2002)	Letter naming, animal name generation	Event-related	Continuous	Comparison of aloud and silent letter naming and animal name generation
Jung, Prasad, Qin and Anderson (2005)	Name repetition (in new order)	Event-related	Continuous	Active noise cancellation
Kan & Thompson-Schill (2004)	Picture naming	Block	Interleaved	Comparison of aloud and silent picture naming
Kemeny et al. (2006)	Sentence generation	Block	Continuous	
Kemeny, Ye, Birn, and Braun (2005)	Object and action naming	Event-related	Continuous	
Kircher, Brammer, Levelt, Bartels, and McGuire (2004)	Continuous speech production	Event-related	Continuous	
Leger et al. (2002)	Picture naming and picture/word rhyming	Block	Continuous	

Liegeois et al. (2003)	Verb generation, word repetition	Block	Interleaved	Comparison of aloud and silent verb generation
Martin et al. (2005)	Picture naming	Block	Continuous	
McCarthy, Blamire, Rothman, Gruetter, and Shulman (1993)	Word generation, word repetition	Block	Continuous	Comparison of aloud and silent word generation
Naeser et al. (2004)	Speech production	Block	Continuous	Dynamic susceptibility contrast fMRI technique
Nelles et al. (2003)	Read and generate words	Event-related	Continuous	
Neumann et al. (2003)	Sentence reading	Event-related	Continuous	
Neumann et al. (2005)	Sentence reading	Event-related	Continuous	
Owen, Borowsky, and Sarty (2004)	Word naming	Block	Continuous	
Palmer et al. (2001)	Word stem completion	Event-related	Continuous	Comparison of aloud and silent word stem completion
Peck et al. (2004)	Word generation	Event-related	Continuous	
Phelps, Hyder, Blamirc, and Shulman (1997)	Word repetition, antonym generation, word generation	Block	Continuous	

*(Continued)*

Table 1  
(Continued)

Reference	Speech task	Block vs. event-related	Continuous vs. interleaved	Comments
Preibisch et al. (2003a)	Sentence reading	Event-related	Continuous	
Preibisch et al. (2003b)	Word and sentence reading	Event-related	Continuous	
Riecker, Ackermann, Wildgruber, Dogil, and Grodg (2000)	Speech and melody production	Event-related	Continuous	Comparison of aloud and silent speech and melody production
Rosen, Ojemann, Oilinger, and Petersen (2000)	Word stem completion	Event-related	Continuous	Comparison of aloud and silent word stem completion
Shuster and Lemieux (2005)	Word reading	Event-related	Continuous	Comparison of aloud and silent word reading
Small et al. (1996)	Word reading	Block	Continuous	
Turkeltaub, Eden, Jones, and Zeffiro (2002)	Word reading	Block	Interleaved	
Viswanath, Karmonik, King, Rosenfield, and Mawad (2003)	Speech production	Event-related	Continuous	
Yetkin et al. (1995)	Word generation	Block	Continuous	Comparison of aloud and silent word generation

With the interleaved acquisition mechanism, jaw movement does not occur during data acquisition, resulting in reduced signal drop-out and geometric distortion effects in the derived statistical maps. The data show that reading aloud and reading silently make different demands on the inferior frontal gyrus (IFG).

Because of their lengthy TR intervals, interleaved techniques have limited temporal resolution and are optimal for experimental designs requiring imaging of the entire brain at the relatively low sampling rates of six times per minute or less. Two to four seconds are required to image the entire brain utilizing echo-planar imaging. As the hemodynamic response to a brief movement takes ten seconds to complete, it is possible to image the entire cerebrum during the peak of the hemodynamic modulation.

Because of their relative insensitivity to acoustic and motion artifacts, the interleaved techniques produce activity maps that are comparable to or better than those derived using continuous acquisition, possibly because the reduced inter-scan head motion would result in less temporal image misregistration and therefore higher resulting levels of statistical significance. In addition, interleaved techniques employ a longer TR that can result in increased sensitivity to small signal changes due to the improved contrast-to-noise. In Table 1 we present a guide to some of the recent fMRI work using speech responses to study the neural mechanisms of cognitive processing. Both continuous and interleaved techniques are employed using both block and event-related timing arrangements.

## 1.6. Summary of Recommendations

The proscription against the use of speech responses in functional MRI experiments is not warranted, even though jaw motion can be associated with dramatic artifacts in statistical maps. The use of event-related and interleaved data acquisition techniques can greatly reduce the effects of acoustic gradient noise on detection and estimation of spatial and temporal patterns of task-related activity. Very soon, improvements in MR imaging equipment will include both better acoustic isolation of the subject and active noise-cancellation capabilities, allowing ecologically valid communication studies in much quieter experimental environments.

## REFERENCES

Abrahams, S., Goldstein, L. H., Simmons, A., Brammer, M. J., Williams, S. C., Giampietro, V. P., Andrew, C. M., & Leigh, P. N. (2003). Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. *Human Brain Mapping*, 20, 29–40.

Abrahams, S., Goldstein, L. H., Simmons, A., Brammer, M., Williams, S. C., Giampietro, V., & Leigh, P. N. (2004). Word retrieval in amyotrophic lateral sclerosis: A functional magnetic resonance imaging study. *Brain*, 127, 1507–1517.

- Aldenkamp, A. P., Boon, P. A., Deblaere, K., Achten, E., Backes, W. H., Boon, P., Hofman, P., Troost, J., Vandemaele, P., Vermeulen, J., Vonck, K., & Wilmink, J. (2003). Usefulness of language and memory testing during intracarotid amobarbital testing: Observations from an fMRI study. *Acta Neurologica Scandinavica*, *108*, 147–152.
- Baciu, M. V., Rubin, C., Decorps, M. A., & Segebarth, C. M. (1999). fMRI assessment of hemispheric language dominance using a simple inner speech paradigm. *NMR in Biomedicine*, *12*, 293–298.
- Bandettini, P. A., Jesmanowicz, A., Van Kylen, J., Birn, R. M., & Hyde, J. S. (1998). Functional MRI of brain activation induced by scanner acoustic noise. *Magnetic Resonance in Medicine*, *39*, 410–416.
- Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior cingulate and the monitoring of response conflict: Evidence from an fMRI study of overt verb generation. *Journal of Cognitive Neuroscience*, *12*, 298–309.
- Barch, D. M., Sabb, F. W., Carter, C. S., Braver, T. S., Noll, D. C., & Cohen, J. D. (1999). Overt verbal responding during fMRI scanning: Empirical investigations of problems and potential solutions. *NeuroImage*, *10*, 642–657.
- Barrett, J., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience*, *19*, 458–464.
- Belliveau, J. W., Kennedy, D. N., Jr., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., & Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, *254*, 716–719.
- Birn, R. M., Bandettini, P. A., Cox, R. W., & Shaker, R. (1999). Event-related fMRI of tasks involving brief motion. *Human Brain Mapping*, *7*, 106–114.
- Birn, R. M., Cox, R. W., & Bandettini, P. A. (2004). Experimental designs and processing strategies for fMRI studies involving overt verbal responses. *NeuroImage*, *23*, 1046–1058.
- Biswal, B., DeYoe, A. E., & Hyde, J. S. (1996). Reduction of physiological fluctuations in fMRI using digital filters. *Magnetic Resonance in Medicine*, *35*, 107–113.
- Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, *3*, 93–106.
- Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T. A., Gaillard, P. W., & Theodore, W. H. (2000). Activation of language cortex with automatic speech tasks. *Neurology*, *55*, 1151–1157.
- Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., & Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences, USA*, *93*, 14878–14883.
- Burgund, E. D., Lugar, H. M., Miezin, F. M., & Peterson, S. E. (2003). Sustained and transient activity during an object-naming task: A mixed blocked and event-related fMRI study. *NeuroImage*, *19*, 29–41.

- Burton, M. W., Noll, D. C., & Small, S. L. (2001). The anatomy of auditory word processing: Individual variability. *Brain and Language*, *77*, 119–131.
- Cho, Z. H., Chung, S. C., Lim, D. W., & Wong, E. K. (1998). Effects of the acoustic noise of the gradient systems on fMRI: A study on auditory, motor, and visual cortices. *Magnetic Resonance in Medicine*, *39*, 331–335.
- Constable, R. T., & Spencer, D. D. (2001). Repetition time in echo planar functional MRI. *Magnetic Resonance in Medicine*, *46*, 748–755.
- de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture-word task: An event-related fMRI study using overt vocal responding. *NeuroImage*, *16*, 1084–1093.
- de Zubicaray, G. I., Wilson, S. J., McMahon, K. L., & Muthiah, S. (2001). The semantic interference effect in the picture-word paradigm: An event-related fMRI study employing overt responses. *Human Brain Mapping*, *14*, 218–227.
- Dietz, N. A., Jones, K. M., Gareau, L., Zeffiro, T. A., & Eden, G. F. (2005). Phonological decoding involves left posterior fusiform gyrus. *Human Brain Mapping*, *26*, 81–93.
- Dietz, N. A., Jones, K. M., Twomey, R., Zeffiro, T. A., & Eden, G. F. (in press). A comparison of continuous versus interleaved fMRI data acquisition for the study of reading aloud.
- Eden, G. F., Jones, K. M., Cappell, K., Gareau, L., Wood, F. B., Zeffiro, T. A., Dietz, N. A., Agnew, J. A., & Flowers, D. L. (2004). Neural changes following remediation in adult developmental dyslexia. *Neuron*, *44*, 411–422.
- Eden, G. F., Joseph, J. E., Brown, H. E., Brown, C. P., & Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. *Magnetic Resonance in Medicine*, *41*, 13–20.
- Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*, 89–97.
- Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience*, *7*, 357–375.
- Frenck-Mestre, C., Anton, J. L., Roth, M., Vaid, J., & Viallet, F. (2005). Articulation in early and late bilinguals' two languages: Evidence from functional magnetic resonance imaging. *NeuroReport*, *16*, 761–765.
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*, 346–355.
- Frith, C. D., Friston, K., Liddle, P. F., & Frackowiak, R. S. (1991). Willed action and the prefrontal cortex in man: A study with PET. *Proceedings of the Royal Society of London, Series B*, *244*, 241–246.

- Glover, P., Hykin, J., Gowland, P., Wright, J., Johnson, I., & Mansfield, P. (1995). An assessment of the intrauterine sound intensity level during obstetric echo-planar magnetic resonance imaging. *British Journal of Radiology*, *68*, 1090–1094.
- Grabowski, T. J., Bauer, M. D., Foreman, D., Mehta, S., Eaton, B. L., Graves, W. W., Defoe, D. L., & Bolinger, L. (2006). Adaptive pacing of visual stimulation for fMRI studies involving overt speech. *NeuroImage*, *29*, 1023–1030.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., & Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*, 213–223.
- Haller, S., Radue, E. W., Erb, M., Grodd, W., & Kircher, T. (2005). Overt sentence production in event-related fMRI. *Neuropsychologia*, *43*, 807–814.
- Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*, *20*, 22–28.
- Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *The Journal of Neuroscience*, *14*, 6336–6353.
- He, A. G., Tan, L. H., Tang, Y., James, G. A., Wright, P., Eckert, M. A., Fox, P. T., & Liu, Y. (2003). Modulation of neural connectivity during tongue movement and reading. *Human Brain Mapping*, *18*, 222–232.
- Heim, S., Opitz, B., & Friederici, A. D. (2002). Broca’s area in the human brain is involved in the selection of grammatical gender for language production: Evidence from event-related functional magnetic resonance imaging. *Neuroscience Letters*, *328*, 101–104.
- Herscovitch, P., Markham, J., & Raichle, M. E. (1983). Brain blood flow measured with intravenous H<sub>2</sub><sup>15</sup>O. I. Theory and error analysis. *Journal of Nuclear Medicine*, *24*, 782–789.
- Hu, X., Le, T. H., Parrish, T., & Erhard, P. (1995). Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magnetic Resonance in Medicine*, *34*, 201–212.
- Huang, J., Carr, T. H., & Cao, Y. (2002). Comparing cortical activations for silent and overt speech using event-related fMRI. *Human Brain Mapping*, *15*, 39–53.
- Jung, K. J., Prasad, P., Qin, Y., & Anderson, J. R. (2005). Extraction of overt verbal response from the acoustic noise in a functional magnetic resonance imaging scan by use of segmented active noise cancellation. *Magnetic Resonance in Medicine*, *53*, 739–744.
- Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affect, & Behavioral Neuroscience*, *4*, 43–57.
- Kemeny, S., Xu, J., Park, G. H., Hosey, L. A., Wettig, C. M., & Braun, A. R. (2006). Temporal dissociation of early lexical access and articulation using a delayed naming task - An fMRI study. *Cerebral Cortex*, *16*, 587–595.

- Kemeny, S., Ye, F. Q., Birn, R., & Braun, A. R. (2005). Comparison of continuous overt speech fMRI using bold and arterial spin labeling. *Human Brain Mapping, 24*, 173–183.
- Kircher, T. T., Brammer, M. J., Levelt, W., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortex during pauses in speech. *NeuroImage, 21*, 84–90.
- Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H. -M., Brady, T. J., & Rosen, B. R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences, USA, 89*, 5675–5679.
- Larsen, B., Skinhoj, E., & Lassen, N. A. (1978). Variations in regional cortical blood flow in the right and left hemispheres during automatic speech. *Brain, 101*, 193–209.
- Larsen, B., Skinhoj, E., Soh, K., Endo, H., & Lassen, N. A. (1977). The pattern of cortical activity provoked by listening and speech revealed by rCBF measurements. *Acta Neurologica Scandinavica Supplementum, 64*, 268–269, 280–281.
- Le, T. H., & Hu, X. (1996). Retrospective estimation and correction of physiological artifacts in fmri by direct extraction of physiological activity from MR data. *Magnetic Resonance in Medicine, 35*, 290–298.
- Leger, A., Demonet, J. F., Ruff, S., Aithamon, B., Touyeras, B., Puel, M., Boulanouar, K., & Cardebat, D. (2002). Neural substrates of spoken language rehabilitation in an aphasic patient: An fMRI study. *NeuroImage, 17*, 174–183.
- Liegeois, F., Baldeweg, T., Connelly, A., Gadian, D. G., Mishkin, M., & Vargha-Khadem, F. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nature Neuroscience, 6*, 1230–1237.
- Martin, P. I., Naeser, M. A., Doron, K. W., Bogdan, A., Baker, E. H., Kurland, J., Renshaw, P., & Yurgelun-Todd, D. (2005). Overt naming in aphasia studied with a functional MRI hemodynamic delay design. *NeuroImage, 28*, 194–204.
- McCarthy, G., Blamire, A. M., Rothman, D. L., Gruetter, R., & Shulman, R. G. (1993). Echo-planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans. *Proceedings of the National Academy of Sciences, USA, 90*, 4952–4956.
- Naeser, M. A., Martin, P. I., Baker, E. H., Hodge, S. M., Sczernie, S. E., Nicholas, M., Palumbo, C. L., Goodglass, H., Wingfield, A., Samaraweera, R., Harris, G., Baird, A., Renshaw, P., & Yurgelun-Todd, D. (2004). Overt propositional speech in chronic nonfluent aphasia studied with the dynamic susceptibility contrast fMRI method. *NeuroImage, 22*, 29–41.
- Nelles, J. L., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2003). Automated method for extracting response latencies of subject vocalizations in event-related fMRI experiments. *NeuroImage, 20*, 1865–1871.
- Neumann, K., Euler, H. A., von Gudenberg, A. W., Giraud, A. -L., Lanfermann, H., Gall, V., & Preibisch, C. (2003). The nature and treatment of stuttering as revealed by fMRI a within- and between-group comparison. *Journal of Fluency Disorders, 28*, 281–409.



- Neumann, K., Preibisch, C., Euler, H. A., von Gudenberg, A. W., Lanfermann, H., Gall, V., & Giraud, A. L. (2005). Cortical plasticity associated with stuttering therapy. *Journal of Fluency Disorders, 30*, 23–39.
- Novitski, N., Alho, K., Korzyukov, O., Carlson, S., Martinkauppi, S., Escera, C., Rinne, T., Aronen, H. J., & Naatanen, R. (2001). Effects of acoustic gradient noise from functional magnetic resonance imaging on auditory processing as reflected by event-related brain potentials. *NeuroImage, 14*, 244–251.
- Novitski, N., Anourova, I., Martinkauppi, S., Aronen, H. J., Naatanen, R., & Carlson, S. (2003). Effects of noise from functional magnetic resonance imaging on auditory event-related potentials in working memory task. *NeuroImage, 20*, 1320–1328.
- Obrist, W. D., Thompson, H. K., Jr., King, C. H., & Wang, H. S. (1967). Determination of regional cerebral blood flow by inhalation of 133-Xenon. *Circulation Research, 20*, 124–135.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences, USA, 89*, 5951–5955.
- Owen, W. J., Borowsky, R., & Sarty, G. E. (2004). fMRI of two measures of phonological processing in visual word recognition: Ecological validity matters. *Brain and Language, 90*, 40–46.
- Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *NeuroImage, 14*, 182–193.
- Peck, K. K., Moore, A. B., Crosson, B. A., Gaiefsky, M., Gopinath, K. S., White, K., & Briggs, R. W. (2004). Functional magnetic resonance imaging before and after aphasia therapy: Shifts in hemodynamic time to peak during an overt language task. *Stroke, 35*, 554–559.
- Phelps, E. A., Hyder, F., Blamire, A. M., & Shulman, R. G. (1997). fMRI of the prefrontal cortex during overt verbal fluency. *NeuroReport, 8*, 561–565.
- Poncellet, B. P., Wedeen, V. J., Weisskoff, R. M., & Cohen, M. S. (1992). Brain parenchyma motion: Measurement with cine echo-planar MR imaging. *Radiology, 185*, 645–651.
- Preibisch, C., Neumann, K., Raab, P., Euler, H. A., von Gudenberg, A. W., Lanfermann, H., & Giraud, A. L. (2003a). Evidence for compensation for stuttering by the right frontal operculum. *NeuroImage, 20*, 1356–1364.
- Preibisch, C., Raab, P., Neumann, K., Euler, H. A., von Gudenberg, A. W., Gall, V., Lanfermann, H., & Zanella, F. (2003b). Event-related fMRI for the suppression of speech-associated artifacts in stuttering. *NeuroImage, 19*, 1076–1084.
- Price, C. J., Wise, R. J., & Frackowiak, R. S. (1996a). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex, 6*, 62–70.

- Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S., & Friston, K. J. (1996b). Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain*, *119*, 919–931.
- Raichle, M. E., Martin, W. R., Herscovitch, P., Mintun, M. A., & Markham, J. (1983). Brain blood flow measured with intravenous  $H_2^{15}O$ . II. Implementation and validation. *Journal of Nuclear Medicine*, *24*, 790–798.
- Ravicz, M. E., & Melcher, J. R. (2001). Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: Examination of noise conduction through the ear canal, head, and body. *The Journal of the Acoustical Society of America*, *109*, 216–231.
- Ravicz, M. E., Melcher, J. R., & Kiang, N. Y. (2000). Acoustic noise during functional magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *108*, 1683–1696.
- Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport*, *11*, 1997–2000.
- Rosen, H. J., Ojemann, J. G., Ollinger, J. M., & Petersen, S. E. (2000). Comparison of brain activation during word retrieval done silently and aloud using fMRI. *Brain and Cognition*, *42*, 201–217.
- Roy, C. S. C. (1890). On the regulation of the blood supply of the brain. *The Journal of Physiology*, *11*, 85–108.
- Rumsey, J. M., Horwitz, B., Donohue, B. C., Nace, K., Maisog, J. M., & Andreason, P. (1997). Phonological and orthographic components of word recognition. A PET-rCBF study. *Brain*, *120*, 739–759.
- Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F. M., & Petersen, S. E. (1997). Top-down modulation of early sensory cortex. *Cerebral Cortex*, *7*, 193–206.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*, 20–31.
- Slifer, K., Cataldo, M. F., Cataldo, M. D., Llorente, A. M., & Gerson, A. C. (1993). Behavior analysis of motion control for pediatric neuroimaging. *Journal of Applied Behavioral Analysis*, *26*, 469–470.
- Small, S. L., Noll, D. C., Perfetti, C. A., Hlustik, P., Wellington, R., & Schneider, W. (1996). Localizing the lexicon for reading aloud: Replication of a PET study using fMRI. *NeuroReport*, *7*, 961–965.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain* (M. Rayport, Trans.). New York: Thieme Medical Publishers, Inc.
- Talavage, T. M., & Edmister, W. B. (2004). Nonlinearity of fMRI responses in human auditory cortex. *Human Brain Mapping*, *22*, 216–228.

- Talavage, T. M., Edmister, W. B., Ledden, P. J., & Weisskoff, R. M. (1999). Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Human Brain Mapping, 7*, 79–88.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage, 16*, 765–780.
- Viswanath, N. S., Karmonik, C., King, D., Rosenfield, D. B., & Mawad, M. (2003). Functional magnetic resonance imaging (fMRI) of a stutterer's brain during overt speech. *Journal of Neuroimaging, 13*, 280–281.
- Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. (1998a). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography, 22*, 139–152.
- Woods, R. P., Grafton, S. T., Watson, J. D., Sicotte, N. L., & Mazziotta, J. C. (1998b). Automated image registration: II. Intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography, 22*, 153–165.
- Yang, S., Ross, T. J., Zhang, Y., Stein, E. A., & Yang, Y. (2005). Head motion suppression using real-time feedback of motion information and its effects on task performance in fMRI. *NeuroImage, 27*, 153–162.
- Yetkin, F. Z., Hammeke, T. A., Swanson, S. J., Morris, G. L., Mueller, W. M., McAuliffe, T. L., & Haughton, V. M. (1995). A comparison of functional MR activation patterns during silent and audible language tasks. *American Journal of Neuroradiology, 16*, 1087–1092.

## **SECTION 2**

### **LANGUAGE COMPREHENSION**

This page intentionally left blank

## Chapter 6

# Speech Perception within a Biologically Realistic Information-Theoretic Framework

*Keith R. Kluender and Michael Kieft*

### 1. INTRODUCTION

During the second half of the 20th century, research concerning speech perception stood relatively distinct from the study of audition and other modalities of high-level perception such as vision. Contemporary research, however, is beginning to bridge this traditional divide. Fundamental principles that govern all perception, some known for more than a century, are shaping our understanding of perception of speech as well as other familiar sounds.

Investigators of speech perception traditionally attempted to explain how listeners perceive the spoken acoustic signal as a sequence of consonants and vowels, collectively referred to as *phonetic segments* or *units*. When one describes speech sounds in this way, brackets are used to surround phonetic symbols such as [j] (the ‘y’ sound in ‘yes’) and [o] (as in ‘oh’). By contrast, *phonemes* are more abstract linguistic units that roughly correspond to letters in written language, and are transcribed surrounded by slashes (/j/ and /o/.) Morphemes are the smallest meaningful units of language, roughly corresponding to words (e.g., ‘dog’, ‘taste’, as well as ‘dis’- and ‘-ful’) with phonemes being the smallest units that can change the meaning of a morpheme (e.g., ‘yo’ versus ‘go’) (Trubetskoy, 1969). Within this scheme, the experimental study of speech perception classically has corresponded more or less to the lowest division of labor generally agreed upon by linguists and psycholinguists.

To the extent that speech perception researchers’ task is to deliver minimal units to those who study language, an important caveat must be applied to this inherited division of labor. There is no clear experimental evidence demonstrating that either phonetic segments or phonemes are real outside of linguistic theory (e.g., Lotto, 2000), and the appeal of phonetic segments and phonemes may arise principally from experience with

alphabetic writing systems (e.g., Morais, Bertelson, Cary, & Alegria, 1986; Morais, Cary, Alegria, & Bertelson, 1979; Port, in press). One ought not be sanguine about whether speech perception really is about recognizing consonants and vowels *per se*. Listeners probably do not extract phonemes preliminary to recognizing words. There may be nowhere in the brain where phonemes reside independent of words that they comprise.

Nevertheless, conceptualizing speech perception as a process by which phonemes are retrieved from acoustic signals is tradition. Within this tradition, research in speech perception often has been focused on problems concerning *segmentation* and *lack of invariance*. The problem of segmentation refers to the fact that, if phonetic units exist, they are not like typed letters on a page. Instead, they overlap extensively in time, much like cursive handwriting. The problem of lack of invariance (or, problem of variability) is related to the segmentation problem. Because speech sounds are produced such that articulations for one consonant or vowel overlaps with production of preceding ones, and *vice versa*, every consonant and vowel produced in fluent connected speech is dramatically colored by its neighbors. Some of the most recalcitrant problems in the study of speech perception are the consequence of adopting discrete phonetic units as a level of analysis, a level that is not discrete and may not be real. In connected speech, acoustic realization of the beginning and end of one word also overlaps with sounds of preceding and following words, so the problems of invariance and segmentation are not restricted to phonetic units.

This being said, either morphemes or words are the first units of language that stand more or less on their own accord.<sup>1</sup> It is possible, even likely, that speech perception is a series of non-discrete processes along the way from waveforms to words. In this chapter, speech perception will be described as a continuum of processes operating on the acoustic signal with varying levels of sophistication. The consistent theme will be common principles that define how these processes work.

Following some preliminaries concerning broad principles that govern perception, a framework for conceptualizing perception of speech will be presented. In part, this approach is modest because many of the central premises are derivative of what is known about domain-general processes of perception and learning. In addition, this approach is conservative by virtue of avoiding *ad hoc* claims concerning processing of speech in any unique way, while also avoiding reliance upon higher-level cognitive processes. The central claim is that perception of speech works the same way perception works for other modalities and for other environmental sources. Speech perception follows a handful of general principles that are implemented in both sophisticated and not-so-sophisticated ways through the chain of processing from periphery through central nervous system.

---

<sup>1</sup> Very similar skepticism concerning the status of morphemes arises. For example, Seidenberg and Gonnerman (2000) describe morphemes as graded units arising from correlations among sound, meaning, and spelling.

## 2. SOME FUNDAMENTALS OF PERCEPTION

### 2.1. The Inverse Problem

For many years much of the study of speech perception was conducted in isolation from the study of perception more generally to mostly ill effect. In part, this state of affairs was encouraged by the focus of language researchers (linguists and psycholinguists) seeking to know more about elemental aspects of language use. Consistent with appreciation for apparently unique characteristics of human language, early speech researchers were encouraged to believe that perception of speech may be as unique as language itself. For this and other historical reasons, research in speech perception was often naïve to developments in related areas of perception.

An enduring distraction for investigators studying speech perception has concerned the extent to which articulatory gestures (e.g., Fowler, 1986; Liberman & Mattingly, 1985), acoustic patterns, patterns of sensory stimulation (e.g., Diehl & Kluender, 1989), or some combination (e.g., Nearey, 1997; Stevens & Blumstein, 1981) serve as proper objects of speech perception. Controversies concerning appropriate objects of perception generated a fair bit more heat than light. However, debates concerning objects of perception cannot be resolved because the question itself is ill-posed, if not outright misleading. There are no objects of perception, either for speech or for perception in general. There is an *objective* for perception, which is to maintain adequate agreement between an organism and its world in order to facilitate adaptive behavior. Success with this objective does not require objects of perception.

Within this functional framework, perceptual success does not require recovery or representations of the world *per se*. Perceivers' subjective impressions may be of objects and events in the world, and the study of perceptual processes may lead to inspection of real-world objects and events, patterns of light or sound pressure waves, transduction properties, or neural responses. By and large, however, viewing perception with a focus toward either distal or proximal properties falls short of capturing the essential functional characteristic of perception – the relationship between an organism's environment and its actions.

This depiction of success in perception as essentially functional, discarding any sense of perceiving true reality, might seem novel to some readers. However, these ideas are classic, having become so broadly accepted that most mention seems to have lapsed in instruction to modern students of perception. Beginning at least with Helmholtz (e.g., 1866/1969), it has been understood that perceiving the true state of the world is impossible. Helmholtz himself was led to this understanding by British Empiricist philosophers (e.g., Hume, 1748/1963; Berkeley, 1837/1910). Nevertheless, contemporary discourse in the field of perception often betrays this fact.<sup>2</sup>

---

<sup>2</sup> One may question whether denying the possibility of veridical recovery is arguing against a straw man. In some ways, this is trivially correct because there are limitations upon biological sensors in both range and precision of transduction (e.g., hearing across only 20–20,000 Hz with limitations of precision in both frequency and amplitude.) Consequently, humans cannot hear environmental sounds that are heard by elephants (lower frequencies) and bats (higher frequencies). A more interesting criticism is that it is always impossible to know



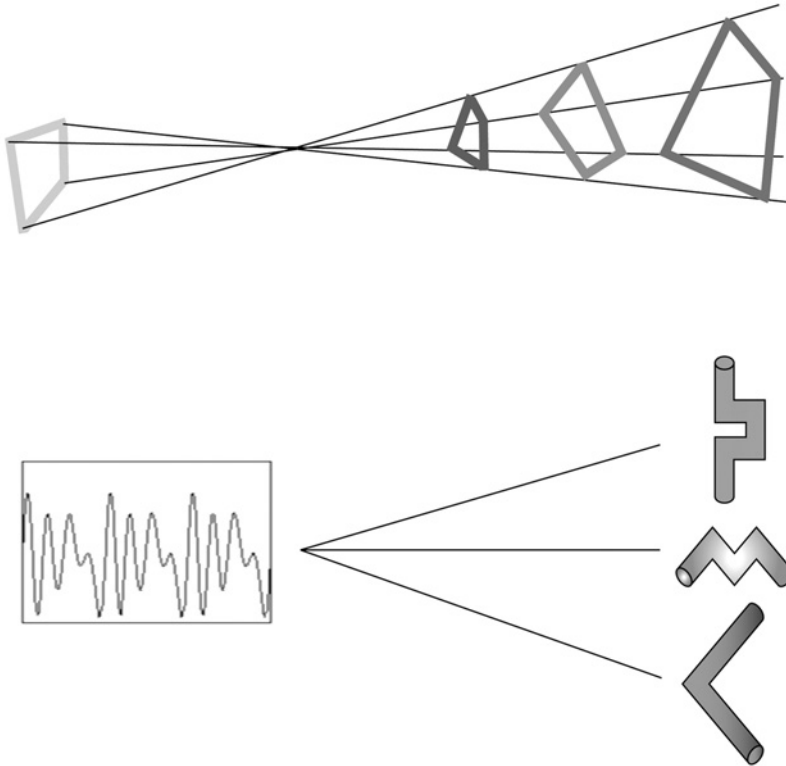


Figure 1. An infinite number of external 3-dimensional objects give rise to the same 2-dimensional retinal image (top, left). An infinite number of sound producing sources (characterized on right as resonator shapes) give rise to the same waveform available to the ear (bottom, left).

Much contemporary work in perception is concerned, in one way or another, with addressing the *inverse problem* (Figure 1). The inverse problem emerges from the simple fact that information available to sensory transducers (eyes, ears, etc.) is inadequate to authentically reconstruct a unique distal state of affairs. In vision, for any 2-dimensional projection, there are an infinite number of possible 3-dimensional objects that could give rise to exactly the same 2-D image. In audition, for any sound-pressure wave,

---

(continued)

Truth (with a capital 'T'). Instead, all one can hope is to evaluate function, whether or not something works. Within this tradition of *pragmatism* (e.g., Peirce, 1878; James, 1897), one can assign 'truth' (with a lowercase 't') on functional grounds. The parallel here is that perception cannot provide Veridical (with a capital 'V') recovery of the environment, but it can supply veridical recovery as measured by whether perception gets the job done for the organism. The present approach is intended to be consistent with the pragmatic rendition of 'truth' and of 'veridical' in as much as the only evaluative measure is whether perception is successful for the organism. Here, these terms are avoided in the interest of being faithful to the vernacular within which 'truth' and 'veridical' are taken to imply some real portrayal of the world.

there are an infinite number of sound producing events that could give rise to that waveform. These are facts of physical optics and acoustics, not theory. Information available to sensory transducers is inadequate to reconstruct an authentic optical or acoustic distal environment.

For speech perception, the inverse problem presents one of the two major reasons why appeals to articulatory gestures cannot in principle or in practice make one's theory of speech perception simpler or more successful. There is a lawful mapping from characteristics of physical sound sources to the waveforms they produce. The inverse mapping, from waveforms to sound sources, is indeterminate. There are very limited cases for which it is theoretically possible to solve the inverse problem in acoustics. For example, Jenison (1997) has demonstrated that characteristics of movement of a sound source could be derived from conjoint detection of interaural-time-delay, Doppler shifts, and sound intensity. However, it is unlikely that this theoretical possibility has biological plausibility because biological transducers lack the precision required for the three variables, and because extreme environmental conditions required approach biological limits of detection (e.g., extremely fast moving objects to yield sufficient Doppler shifts) fall outside the domain of normal perceptual experience. More typical is the case of attempting to solve the inverse from waveform to simpler 2-D surfaces (e.g., the shape of a drum.) Mathematicians have formally proved that even this relatively simple translation from waveform to plane geometry is impossible (Gordon, Webb, & Wolpert, 1992).

Because multiple sound sources yield the same waveform, waveforms can never be more complex than characteristics of physical sources. Researchers within the field of speech perception have long been familiar with appeals to perception via articulatory gestures as a simplifying construct, and there have been a series of efforts to extract gestures in order to facilitate machine speech recognition, albeit with very limited success. What physics demands, however, is that depiction of speech in terms of articulatory gestures can give only the illusion of simplicity. Because scientists are much better at measuring details of sounds than they are at measuring details of articulator activity, articulatory gestures appear simpler only because they are defined more abstractly and are measured with less precision. Because multiple resonator configurations can give rise to the same waveform, the acoustic waveform available to listeners always underestimates variability in articulation.

For all of the discussion that follows regarding specific issues concerning speech perception, speech typically will be described as sounds. This is not because sounds are legitimate objects of perception. This is because, along the chain of events from creating patterns of sound-pressure to encoding these patterns in some collection of neural firings to eliciting behavior, waveforms are public, easily measurable, and simpler than alternatives.

## 2.2. Why Perception Seems Veridical

If perceiving the true state of the world is impossible, one might ask why phenomenal experience is not fuzzy and uncertain. To effectively guide behavior, and not leave the

organism pondering multiple possibilities, all that is required is that the perceptual system come to the same adaptive output every time it receives functionally the same input. It is this deterministic nature of perception that prevents being paralyzed among myriad alternatives. Phenomenal experience of certain reality does not depend on authentic rendering of the world. Instead, phenomenal experience of a clear and certain world is the consequence of perceptual systems reliably arriving at deterministically unique outputs. It is this reliability that encourages certainty (Hume, 1748/1963), but reliability is not validity.

On rare occasions, perceptual systems do not converge on a unique output and are left oscillating between equally fitting outputs when sensory inputs are not singly determinate (usually in response to impoverished stimuli.) Many readers are familiar with bistability when viewing Necker cubes. One such auditory experience is encountered when listening to a repeating synthesized syllable intermediate between [da] and [ta] or any other pair of similar speech sounds. When two perceptual outputs fit the input equally well, phenomenal experience oscillates between two percepts (Tuller, Case, Ding, & Kelso, 1994).

### 2.3. Information for Perception

If there are no objects *of* perception, how should one think about information *for* perception? Information for perception does not exist in the objects and events in the world, nor does it exist in the head of the perceiver. Instead, information exists in the relationship between an organism and its world. It may be useful to consider the contrast between information *about* and information *for*. When one discusses objects of perception, it is information *about* that is typically inferred. Implicit in such efforts is the notion that one needs to solve the inverse problem. By contrast, if the objective of a successful perceptual system is to maintain adequate agreement between an organism and its world in order to facilitate adaptive behavior, then information *for* successful perception is nothing more or less than information that resides in this relationship (or agreement).

This way of viewing information as a relationship is consistent with one of the fundamental characteristics of Shannon information theory (Shannon, 1948; Weiner, 1948). Some readers may be familiar with Fletcher's pioneering applications of information theory to speech (Fletcher, 1953/1995). However, the application here will be more akin to the approach of Attneave (1954, 1959) and Barlow (1961), an approach that remains highly productive (e.g., Barlow, 1997, 2001; Simoncelli & Olshausen, 2001; Schwartz & Simoncelli, 2001). One important point of Shannon's information theory is that information exists only in the relationship between transmitters and receivers; information does not exist in either *per se*, and it does not convey any essential characteristics about either transmitters or receivers. Within this information-theoretic sense, perceptual information exists in the agreement between organisms and their environments. This agreement is the objective of perception (Figure 2).

Within a sea of alternative perceptual endpoints, agreement between the organism and environment is arriving at the alternative that gives rise to adaptive behavior. Information is transmitted when uncertainty is reduced and agreement is achieved

between organism and environment. The greater the number of alternatives (uncertainty, unpredictability, variability, or entropy) there are, the greater the amount of information that potentially can be transmitted (Figure 2a). There is no information when there is no variability. When there is no variability, there is total predictability and hence, no

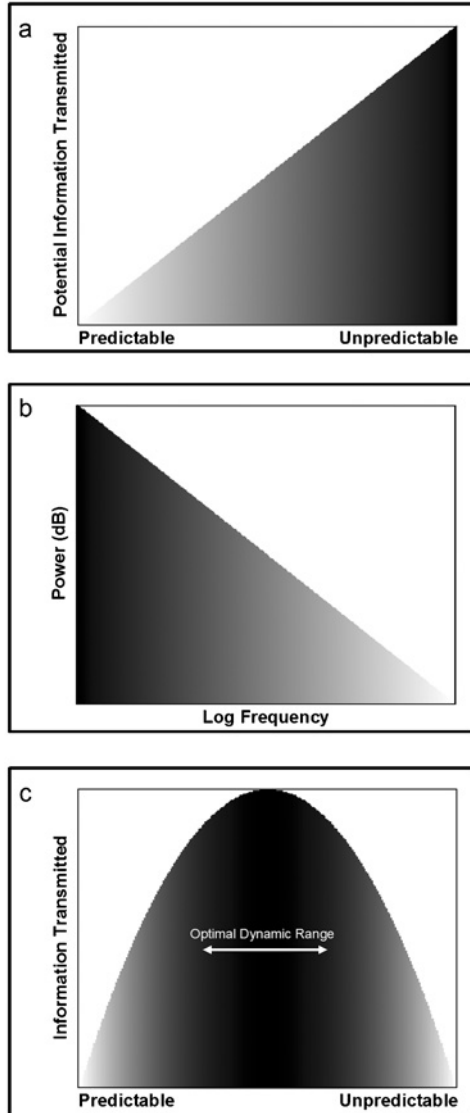


Figure 2. (a) The greater the number of alternatives (uncertainty, unpredictability, variability, or entropy) there are, the greater the amount of information that potentially can be transmitted. There is no new information in what stays the same or is predictable. (b) Relative power of energy flux in natural environments approximates  $1/f$ . (c) Information transmission optimized relative to energy flux in the environment. A sensorineural system should optimize dynamic range about this maximum.

information transmitted. There is much that stays the same in the world from time to time and place to place, but there is no information in stasis. Uncertainty is reduced consequent to the perceiver's current experiences (context) as well as past experiences with the environment (learning).

Shannon and his Bell Telephone Laboratory engineer colleagues were concerned with evaluating equipment, not listeners. Answers to questions about what equipment can do are different from answers to questions about what biological perceivers naturally do (Licklider & Miller, 1951). Although the amount of theoretical potential information transmitted is maximized at maximum entropy (total unpredictability or randomness), it is not advantageous for biological systems to shift dynamic range as far as possible toward this maximum. In natural environments, this would result in diminishing returns if the system adjusts to register the last bits of near-random energy flux. Instead, biological systems should optimize the efficiency with which they capture information relative to the distribution of energy flux in real environments. The best estimate of statistics of natural environments is  $1/f$  (pink) noise (Figure 2b). This simple power law with a negative exponent ( $f^{-1}$ ) is scale-invariant, and it is a ubiquitous characteristic across many systems from radioactive decay to fluid dynamics, biological systems, and astronomy. As one would expect, spectral density of fluctuations in acoustic power of music and speech vary as  $1/f$  (Voss & Clarke, 1975, 1978). Efficient information transmission for sensorineural systems with limited dynamic range may be depicted best as the product of the positive exponential growth in information and the negative exponential of  $1/f$ . This yields the quadratic function shown in Figure 2c describing optimal transmission of information relative to energy flux in the environment.

#### 2.4. Sensory Systems Respond to Change (and little else)

Given these facts about information, it is true and fortunate that sensorineural systems operate as they do. Sensorineural systems respond only to change relative to what is predictable or does not change. Perceptual systems do not record absolute levels whether loudness, pitch, brightness, or color. Relative change is the coin of the realm for perception, a fact known at least since Ernst Weber in the mid-18th century, and has been demonstrated perceptually in every sensory domain. Humans have a remarkable ability to make fine discriminations, or relative judgments, about frequency and intensity. The number of discriminations than can be made numbers in the hundreds or thousands before full dynamic range is exhausted. Yet, most humans are capable of reliably categorizing, or making absolute judgments about only a relatively small number of stimuli regardless of physical dimension (e.g., Miller, 1956; Gardner & Hake 1951). This sensory encoding of change, and not absolute characteristics, is another major reason why veridical recovery is biologically impossible.

Sacrifice of absolute encoding has enormous benefits along the way to maximizing information transmission. Although biological sensors have impressive dynamic range given their evolution via borrowed parts (e.g., gill arches to middle ear bones), this dynamic range is always a fraction of the physical range of absolute levels available from

the environment and essential to organisms' survival. This is true whether one is considering optical luminance or acoustic pressure. The beauty of sensory systems is that, by responding to relative change, a limited dynamic range shifts upward and downward to optimize the amount of change that can be detected in the environment at a given moment.

The simplest way that sensory systems adjust dynamic range to optimize sensitivity is via processes of adaptation. Following nothing, even a subtle sensory stimulus can trigger a strong sensation. However, when a level of sensory input is sustained over time, constant stimulation loses impact. This sort of sensory attenuation due to adaptation is ubiquitous, and has been documented in vision (Riggs, Ratliff, Cornsweet, & Cornsweet, 1953), audition (Hood, 1950), taste (Urbantschitsch, 1876, cf. Abrahams, Krakauer & Dallenbach, 1937), touch (Hoagland, 1933), and smell (Zwaardemaker, 1895, cf. Engen, 1982). There are increasingly sophisticated mechanisms supporting sensitivity to change with ascending levels of processing, and several will be discussed in this chapter. Most important for now is the fundamental principle that perception of any object or event is always relative – critically dependent on its context.

### 3. CONTRAST AND LOW-LEVEL SPEECH PERCEPTION

#### 3.1. Contrast in General

Because it is only change that is perceived, perception at any particular time or place always depends on temporally or spatially adjacent information. Many instances of sensitivity to change are revealed through demonstration of contrast. For example, a gray region appears darker against a white background and lighter next to a black background (see e.g., Anderson & Winawer, 2005). While examples abound for all modalities, sensory contrast has been most amply demonstrated in studies of visual perception, with contrast being a fundamental process that reveals edges of objects and surfaces (Figure 3).

Contrast effects are ubiquitous, and of course, they exist for audition (Cathcart & Dawson, 1928/1929; Christman, 1954). Forms of auditory contrast are important for several aspects of speech perception. Over the past few years, multiple studies have provided evidence that simple processes of spectral contrast contribute to solving one of the most, if not the most, difficult questions concerning speech perception, coarticulated speech. Coarticulation is the spatial and temporal overlap of adjacent articulatory activities, and it is reflected in the acoustic signal by severe context-dependence. Acoustic information specifying one speech sound varies substantially depending on surrounding sounds.

#### 3.2. Contrast and Perception of Coarticulated Speech

The problem for speech perception is how listeners hear a speech sound such as [d] when acoustic characteristics change dramatically depending upon sounds that precede and follow (e.g., vowels [e] versus [o]) (Figure 4). Coarticulation presents a major challenge to automatic speech recognition (ASR) systems, which largely identify speech sounds on

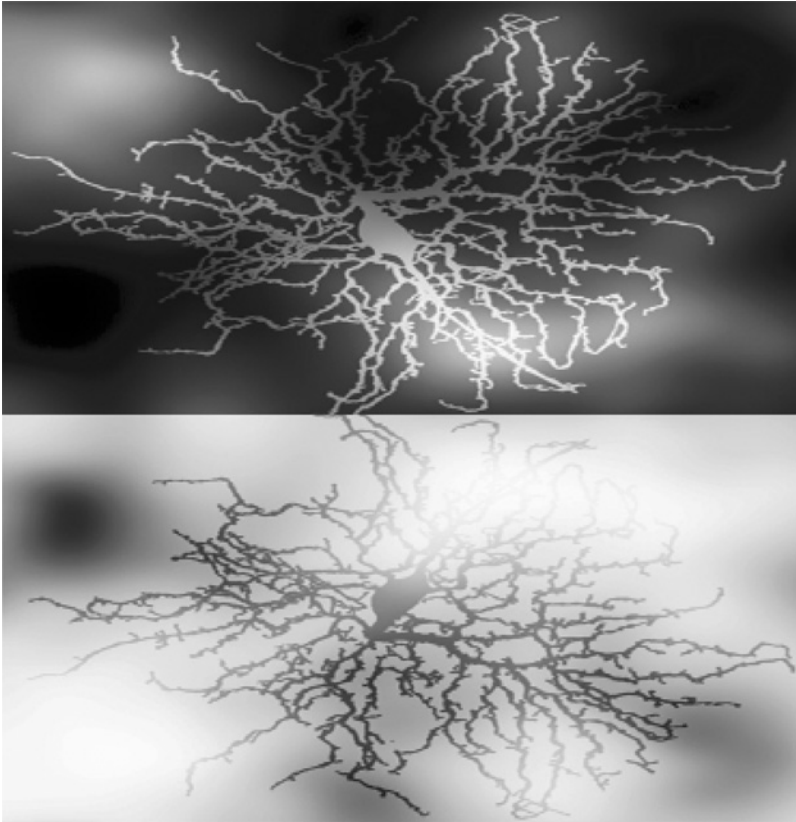


Figure 3. Contrast effects are ubiquitous in perception. For this example of lightness contrast, absolute lightness/darkness of these mirror-image neurons is identical.

the basis of template matching. Instead of storing a single template for [d], multiple templates must be stored for [d] following all other possible speech sounds, and each of those templates must be stored multiply for every instance of [d] preceding all other possible speech sounds. For ASR, this strategy using a geometrically expanding set of templates can be made to work so long as one has sufficient memory and sufficient processing speed to sort through templates. Not surprisingly, progress in ASR over decades is closely correlated with speed of microprocessors and price of memory (Lippman, 1996).

There is a consistent pattern to coarticulation that suggests a simpler solution. Adjacent sounds always assimilate toward the spectral characteristics of one another. Owing to mass and inertia of articulators (as well as planning), articulatory movements are compromises between where articulators have been and where they are headed. Because the acoustic signal directly reflects these articulatory facts, the frequency spectrum assimilates in the same fashion that speech articulation assimilates.

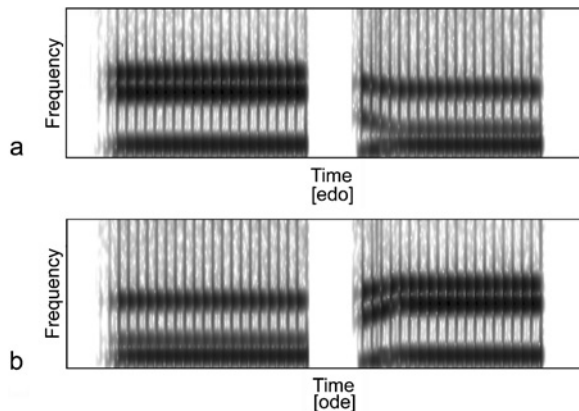


Figure 4. Schematic spectrograms of [edo] (top) and [ode] (bottom.) Note that acoustic properties of [d] depend upon characteristics of preceding and following vowel sounds.

Lindblom (1963) provided some of the best early evidence concerning how context systematically influences speech production. He reported that the frequency of the second formant ( $F_2$ ) was higher in the productions of [did] (‘did’) and [dud] (‘dud’) than for the vowels [i] and [u] in isolation, and that  $F_2$  was lower for vowels in [bib] and [bub]. In both contexts,  $F_2$  frequency approached that of flanking consonants, which are higher for [d] than for [b]. In a subsequent study, Lindblom and Studdert-Kennedy (1967) demonstrated that perception of coarticulated vowels is complementary to these facts of articulation. Listeners reported hearing /i/ (higher  $F_2$ ) more often in [wVw] context, and /u/ more often in [jVj] context. Consonant context affected vowel perception in a manner complementary to the assimilative effects of coarticulation. Lindblom and Studdert-Kennedy (1967) wrote: “It is worth reiterating... that mechanisms of perceptual analysis whose operations contribute to *enhancing contrast* in the above-mentioned sense are precisely the type of mechanisms that seem well suited to their purpose given the fact that the slurred and sluggish manner in which human speech sound stimuli are often generated tends to reduce rather than sharpen contrast” (p. 842, italics added).

One of the most thoroughly investigated cases for perceptual context dependence concerns the realization of [d] and [g] as a function of preceding liquid (Mann, 1980) or fricative (Mann & Repp, 1981). Perception of /d/ as contrasted with perception of /g/, can be largely signaled by the onset frequency and trajectory of the third formant ( $F_3$ ). In the context of a following [a], a higher  $F_3$  onset encourages perception of /da/ while a lower  $F_3$  onset results in perception of /ga/. Onset frequency of the  $F_3$  transition varies as a function of the preceding consonant in connected speech. For example,  $F_3$ -onset frequency for [da] is higher following [al] in [alda] than when following [ar] in [arda]. The offset frequency of  $F_3$  is higher for [al] owing to a more forward place of articulation, and is lower for [ar].



Perception of /da/ and /ga/ has been shown to be complementary to the facts of production much as it is for CVCs. Listeners are more likely to report hearing /da/ (high  $F_3$ ) when preceded by the syllable [ar] (low  $F_3$ ), and hearing /ga/ (low  $F_3$ ) when preceded by [al] (high  $F_3$ ) (Mann, 1980; Lotto & Kluender, 1998). In subsequent studies, the effect has been found for speakers of Japanese who cannot distinguish [l] and [r] (Mann, 1986), for prelinguistic infants (Fowler, Best, & McRoberts, 1990), and for avian subjects (Lotto, Kluender, & Holt, 1997.) The same pattern of findings has been replicated for perception of /d/ and /g/ following fricatives [s] and [ʃ] such that listeners are more likely to report hearing /d/ (high  $F_3$ ) following [ʃ] (lower frequency noise) and hearing /g/ (low  $F_3$ ) following [s] (higher frequency noise) (Mann & Repp, 1981) (Figure 5).

Coarticulation *per se* can be dissociated from its acoustic consequences by combining synthetic speech targets with nonspeech flanking energy that captures minimal essential spectral aspects of speech. Lotto and Kluender (1998) replaced [al] and [ar] precursors with nothing more than constant-frequency sinusoids set to the offset frequencies of  $F_3$  for [al] and [ar] syllables. Perception of following [da-ga] shifted just as it did following full-spectrum [al] and [ar].

Holt, Lotto, and Kluender (2000) replicated the Lindblom and Studdert-Kennedy findings with CVCs using the vowels [ɛ] and [ʌ] flanked by stop consonants [b] and [d]. They replaced flanking [b] and [d] with FM glides that tracked the center frequency of only  $F_2$  for [b] or [g]. Again, the pattern of results for flanking nonspeech FM glides mimicked that for full-spectrum [b] and [d] syllable-initial and syllable-final transitions. On the basis of the results for VCCVs (Lotto & Kluender, 1998) and these results for CVCs, one can conclude that much of perceptual accommodation for coarticulation is not restricted to speech-like signals. All of the findings are consistent with spectral contrast, whereby the spectral composition of context serves to diminish or enhance the perceptual efficacy of spectral components for adjacent sounds.

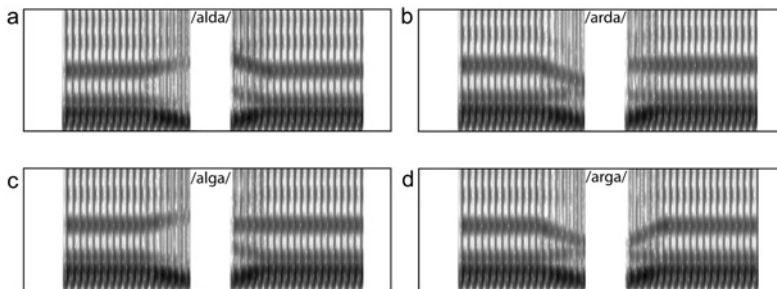


Figure 5. Owing to coarticulation, acoustic properties of any speech sound become more similar to the properties of sounds preceding and following. This assimilation is a property of all fluent connected speech. Here, acoustic characteristics of [d] (e.g.,  $F_3$ ) following [r] (upper right) are very similar to those of [g] following [l] (lower left). Listeners hear the same consonant vowel (CV) as /d/ following [ar] and as /g/ following [al].

In keeping with typical usage, the term *contrast* has been used in a largely descriptive way thus far. There are a large number of experimental precedents for spectral contrast – often called *auditory enhancement*, and these precedents provide more specific hypotheses. Summerfield, Haggard, Foster, and Gay (1984) established the existence of an *aftereffect* in vowel perception. When a uniform harmonic spectrum was preceded by a spectrum that was complementary to a particular vowel with troughs replacing peaks and *vice versa*, listeners reported hearing a vowel during presentation of the uniform spectrum. The vowel percept (for the uniform spectrum) was appropriate for a spectrum with peaks at frequencies where there were troughs in the preceding spectrum.

Summerfield et al. (1984) noted that perceiving vowel sounds in uniform spectra (following appropriate complementary spectral patterns) has a well-known precedent in psychoacoustics. The oft-reported finding is that, if just one member of a set of harmonics of equal amplitude is omitted from a harmonic series and is reintroduced, then it stands out perceptually against the background of the pre-existing harmonics (Schouten, 1940; Green, McKey, & Licklider, 1959; Cardozo, 1967; Viemeister, 1980; Houtgast, 1972). Viemeister (1980), for example, demonstrated that the threshold for detecting a tone in a harmonic complex is 10–12 dB lower when the incomplete harmonic complex (missing the target tone) precedes the tone as compared to when the onset of the inharmonic complex is coincident with that for the target tone. This was referred to as an *enhancement effect*. Viemeister (1980) then examined a number of properties of this effect, finding that the complex need not be harmonic and that noise maskers or bandpass noise signal also served to enhance the detection of the tone. He also found the effect over a wide range of intensities for maskers and targets.

Summerfield and colleagues (1984, 1987) suggested that their demonstration of vowel aftereffects may be rooted in peripheral sensory adaptation. One could suggest that neurons adapt, and the prominence of the added harmonic is due to the fact that neurons tuned to its frequency were not adapted prior to its onset. Alternatively, some researchers (e.g., Houtgast, 1974; Moore & Glasberg, 1983) have suggested that rapid adaptation serves mostly to enhance onsets selectively, with *suppression* being a process through which differences in level of adjacent spectral regions in complex spectra (e.g., formants in speech signals) are preserved and/or enhanced.

Viemeister and Bacon (1982) showed that, not only was an enhanced target tone more detectable, the tone also served as a more effective masker of a following tone. They suggested that suppression must be included in an adaptation scenario to place it in closer accord to this finding. Different frequency components of a signal serve to suppress one another, and two-tone suppression has been cast as an instance of lateral inhibition in hearing (Houtgast, 1972). Investigators have argued that suppression helps to provide sharp tuning (e.g., Wightman, McGee, & Kramer, 1977; Festen & Plomp, 1981), and with respect to speech perception, Houtgast (1974) has argued that this process serves to sharpen the neural projection of a vowel spectrum in a fashion that enhances spectral peaks.

Many neurophysiological observations bear on enhancement effects. In particular, a number of neurophysiological studies of auditory nerve (AN) recordings (e.g., Smith & Zwislocki, 1971; Smith, 1979; Smith, Brachman, & Frisina, 1985) strongly imply a role for peripheral adaptation. Delgutte and colleagues (Delgutte, 1980, 1986, 1996; Delgutte, Hammond, Kalluri, Litvak, & Cariani, 1996; Delgutte & Kiang, 1984) have established the case for a much broader role of peripheral adaptation in perception of speech. He notes that peaks in AN discharge rate correspond to spectro-temporal regions that are rich in phonetic information, and that adaptation increases the resolution with which onsets are represented. Delgutte notes neurophysiological evidence that “adaptation enhances *spectral contrast* between successive speech segments” (p. 3, italics added). This enhancement arises because a fiber adapted by stimulus components close to its CF is relatively less responsive to subsequent energy at that frequency, while stimulus components not present immediately prior are encoded by fibers that are unadapted — essentially the same process offered by psychoacousticians but now grounded in physiology. Delgutte also notes that adaptation takes place on many timescales, and is sustained longer with increasing level in the auditory system.

Inspired by the vowel aftereffect studies by Summerfield and his colleagues (1984; 1987), Coady, Kluender, and Rhode (2003) sought to make clearer the connections between experiments using very simple nonspeech flanking stimuli (e.g., FM glides) and Summerfield’s studies using rich spectra that were complementary to those for vowel sounds. Although sine waves and FM glides have often been used as nonspeech proxies for formants, such sounds have limited resemblance to speech formants. While it is true that spectrograms illustrate formants as bands of energy and formant transitions as bands of energy traversing frequency, such descriptions can be misleading. For example, if fundamental frequency ( $f_0$ ) is constant, individual harmonics of the fundamental do not change frequency at all, and all that changes are relative amplitudes of harmonics. Individual frequency components of the speech spectrum change frequency no more than  $f_0$  changes.

Coady and colleagues (2003) used VCV sequences for which the initial vowel ([e] or [o]) affects perception of the following consonant (/ba/ or /da/). In addition to creating synthetic vowels [e] and [o], they created spectral complements of these vowels [~e] and [~o] by creating troughs where formants occurred for [e] and [o]. These precursor vowel-complements altered perception in a fashion opposite that of normal (noncomplement) vowels because troughs in energy increased excitability within a frequency range for which excitability was attenuated when a frequency prominence (formant) was present in a normal vowel. In addition, they demonstrated that these perceptual effects relied substantially upon spectral characteristics of onsets of sounds.

It appears that the same underlying processes account for effects of both very simple nonspeech precursors and spectrally rich vowel-like complements. Although more complex or domain-limited theories of speech perception have been proposed to explain perception of coarticulated speech, the above patterns of perception with both simple and complex stimuli suggest that spectral contrast is an important part of the explanation for perceptual accommodation of coarticulated speech.

### 3.3. Broader Spectral and Temporal Effects

Contributions of spectral contrast to perception of coarticulated speech are narrowly focused in both time and frequency. Processes through which the auditory system maximizes detection of spectral change operate over durations on the order of less than 1/2 second, and spectral components of interest are relatively local (e.g., formants.) In keeping with the fundamental principle that, in the interest of maximizing transmission of new information, perceptual systems respond primarily to change, long-term signal characteristics that do not change should also alter perception in similar ways.

For vision, perceivers maintain color constancy in the face of changes in level or spectral composition of illumination, respectively. The visual system adjusts for the spectral composition of illumination (e.g., sunlight versus tungsten or fluorescent lighting), while maintaining relatively consistent perception of color under widely varying viewing conditions. Analogous challenges arise for hearing. When auditory experiments are conducted in the laboratory, experimenters typically endeavor to maintain consistent response across all frequencies through the use of high-quality audio equipment and headphones. However, in real-world listening environments, the spectrum is virtually always colored by characteristics of the listening environment. Energy at some frequencies is reinforced by reflective properties of surfaces, while energy at other frequencies is dampened by absorption properties of materials and shapes of objects in the environment. For hearing to be most effective, listeners must adapt to reliable spectral characteristics in order to be maximally sensitive to the most informative characteristics of sounds.

Kiefte and Kluender (2005b) used vowel sounds to examine how auditory systems may adapt to predictable (redundant) spectral characteristics of the acoustic context in order to be more sensitive to information-bearing characteristics of sounds. Simple vowel sounds are useful in this application because it is known that listeners use both spectrally narrow (formant peaks) and broad properties (gross spectral tilt) to perceive vowel sounds. For example, low frequency  $F_1$  is heard as /u/ (as in ‘boot’) when accompanied by a low frequency  $F_2$ , and as /i/ (as in ‘beet’) when accompanied by higher frequency  $F_2$ . In addition, gross spectral tilt, the relative balance between low- and high-frequency energy, is quite different for these vowels. The vowel [u] has more low- than high-frequency energy, resulting in a gross spectral tilt of rapidly declining energy as a function of increasing frequency. In contrast, [i] has relatively more high-frequency energy, and energy decreases much more gradually with increasing frequency. When listening to isolated vowel sounds, listeners use a combination of both formant frequencies and gross spectral tilt to identify vowels (Kiefte & Kluender, 2005a) (Figure 6).

Kiefte and Kluender created a matrix of vowel stimuli that varied perceptually from /u/ to /i/ in two dimensions: center frequency of  $F_2$  and gross spectral tilt. Along one dimension, center frequency of  $F_2$  varied from low ([u]) to high ([i]). Along the second

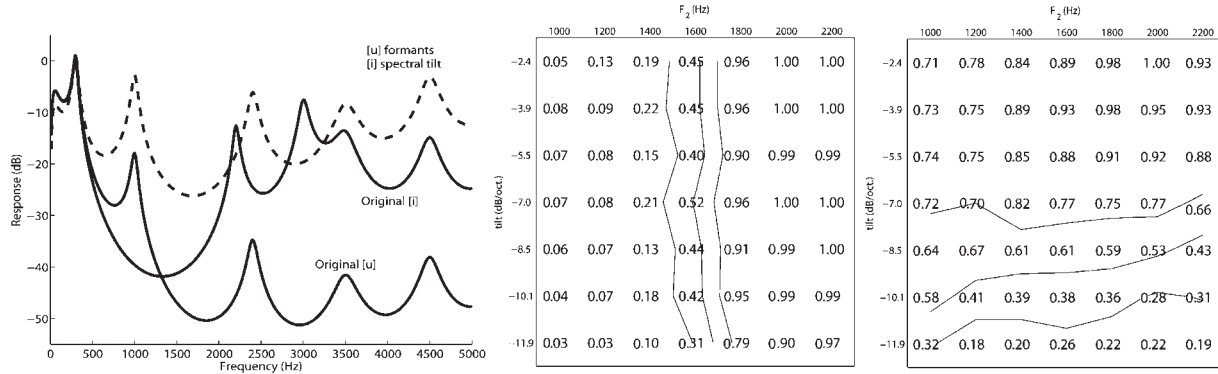


Figure 6. (a) Spectra of vowel sounds [u] and [i] vary both in center frequency of  $F_2$  and overall spectral tilt, and listeners use both acoustic properties to identify these vowels. Kiefte and Kluender (2005) created sounds that varied in each dimension independently. (b) When listeners heard these vowel sounds following a precursor that was filtered to have the same spectral tilt as a target vowel, listeners used  $F_2$  frequency exclusively when identifying the vowels. (c) The complementary pattern of results, exclusive use of tilt, obtained when precursors include a constant-frequency spectral peak matching  $F_2$  of the target sound.

dimension, gross spectral tilt varied in the same stepwise fashion, from a spectral tilt characteristic of [u] to one characteristic of [i]. Listeners identified this matrix of vowel sounds preceded by a synthesized rendition of the sentence “You will now hear the vowel...” When long-term spectral tilt of context sentences was altered to match that of the following vowel, listeners relied virtually exclusively on the frequency of  $F_2$  when identifying /i/ and /u/. This pattern of performance indicates that tilt, as a predictable spectral property of acoustic context, was effectively canceled out of perception.

Encoding of predictable characteristics such as spectral tilt requires only the most abstract characterization of the spectrum, with little spectral detail. Changes in spectral tilt can be due to changes in acoustic properties of the physical environment, or particular to speech, changes in emotional state or speaker identity (Klatt, 1982). Neither of these properties change rapidly and both are relatively stable across time relative to the rapid spectral changes found in speech. Kiefte and Kluender then tested whether perception also compensates for local spectral properties when these properties are stable across time.

Using the same sentence context, sentences were processed with a single-pole filter, which corresponded exactly to the frequency and bandwidth of  $F_2$  of the target, yielding intact sentences with an additional constant-frequency spectral peak added throughout. When presented with a stimulus in which  $F_2$  was an unchanging acoustic property of the context, listeners relied largely on global spectral characteristics (tilt) for identification of the target vowel. Effects of preceding context are not restricted to gross spectral properties. Perceptual cancellation of predictable spectral characteristics also occurs for local, relatively narrowband spectral characteristics of the acoustic context.

In order to rule out the possibility that these effects relied on listeners’ linguistic knowledge or on familiarity with the same context sentence over many trials, Kiefte and Kluender replicated these findings using passages of backward speech of varying lengths. For both conditions, tilt-matched and  $F_2$ -matched between context and vowel sound, effects of preceding context were closely replicated. Perceptual cancellation of predictable acoustic context does not depend on preceding context being speech (Holt, 2005), nor does it depend on the context being identical trial to trial. Listener performance provides evidence that the auditory system is quite adept at factoring out predictable characteristics of a listening context, and is consequently more sensitive to informative changes in spectral composition across time.

Underlying mechanisms by which the auditory system calibrates for characteristics of acoustic context have not yet been extensively investigated, and are not yet understood. Recent findings provide evidence that primary auditory cortex (AI) neurons encode spectral shape with respect to both broad and narrow complex spectral shapes (Barbour & Wang, 2003), and neurons in AI are sensitive to the relative probabilities of pure tones of different frequencies in an extended sequence of tones (Ulanovsky, Las, & Nelken, 2003).

## 4. MAXIMIZING TRANSMISSION OF SPEECH INFORMATION WITH MULTIPLE DIMENSIONS

### 4.1. Speech Perception Uses Multiple Sources of Information

A signature property of speech perception is its extreme resilience in the face of dramatic signal degradation. For example, listeners understand speech at signal-to-noise ratios less than 0 dB, and they understand speech either when all energy is removed above 1500 Hz or when all energy is removed below 1500 Hz. Listeners can understand speech when the only information available is fluctuations in amplitude of eight or so bands of noise across frequency (Shannon, Zeng, Wygonski, Kamath, & Ekelid, 1995), and some listeners can understand speech consisting of little more than sine waves that track the center frequencies of formants (Remez, Rubin, Pisoni, & Carrell, 1981). In large part, these as well as other demonstrations of perceptual resilience can be explained by the fact that listeners can rely on experience with speech that far exceeds experience with any other type of sounds. This power of experience exploits the high degree of redundancy within the speech signal. Redundancy does not distinguish speech from other objects and events in the world. For example, Attneave (1954) notes that information received by the visual system is redundant with sensory events that are highly interdependent in both space and time, and this is simply because “the world as we know it is lawful (p. 183)”.

Every consonant and vowel is defined by multiple acoustic properties. For example, the distinction between [b] and [p] as in [aba] and [apa] includes at least 16 different acoustic differences (Lisker, 1978). Further, no single acoustic attribute from these 16 is necessary to signal the distinction. Much of this redundancy in speech sounds results from the complexity of speech signal. This acoustic complexity is due to the fact that the structures that produce speech, the larynx and vocal tract, also are complex. Human vocal tracts are highly complex structures, having evolved as compromises upon structures from primate ancestry, as they still must accomplish other duties such as drinking, chewing, and swallowing. Talkers morph their vocal tracts into widely varying shapes with different lengths. The surfaces inside vocal tracts vary considerably from tooth enamel to fleshy lips and soft palate, resulting in wildly varying absorption and reflection properties throughout the length of the vocal tract. Talkers also intentionally produce multiple acoustic cues in order to make differences between speech sounds more perceptible (Kingston & Diehl, 1994; Kluender, 1994) resulting in yet greater redundancy in the acoustic signal.

Because multiple attributes are used perceiving speech, the presence of one attribute can compensate for the absence of another, as increasing the magnitude of one source of information serves to compensate for decrease of another. There have been many demonstrations of these “trading relations” (e.g., Repp, 1982) (Figure 7).

Using multiple stimulus attributes is common to perception across modalities. For example, multiple monocular and binocular cues contribute to visual perception of distance. While individual neurons rarely provide high fidelity, populations of neurons acting in concert robustly encode information even when signals are substantially degraded. Implicit to

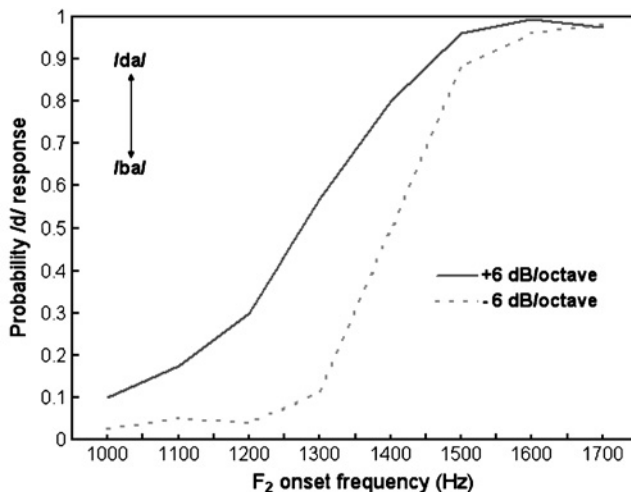


Figure 7. Listeners’ use of multiple acoustic attributes is commonly demonstrated by experiments in which two attributes are independently manipulated. In this example from Alexander and Kluender (2005), when relative spectral tilt is positive (solid), listeners hear sounds with lower  $F_2$  onset frequency as /d/. One attributes “trades” against another, as more of one compensates for less of another.

population encoding is the fact that relationships between activities of multiple neurons conspire for effective perception. It is the correlation of activity across neurons (*i.e.*, redundancy) that makes sensorineural systems robust. Exploiting correlations among multiple attributes in speech perception provides another example of maximizing the performance of perceptual systems by extracting predictability in the service of emphasizing change. The quintessential example of combining multiple acoustic attributes is *categorical perception* (Liberman, Harris, Hoffman, & Griffith, 1957). Preliminary to discussion of categorical perception, some consideration of perceptual constancy and categorization is in order.

#### 4.2. Perceptual Constancy and Categorization

Although rarely recognized as being so (Kluender, 1988), perceptual constancy and categorization share a great deal in common. A classic definition of categorization is that it permits treating discriminably different examples as functionally equivalent. A virtue of categorization typically is presented as efficiently directing responses to functionally equivalent objects or events. Similarly, perceptual constancy maintains when discriminably different exposures (varying with size, orientation, etc.) are treated as equivalent. For example, the apparent size of an object remains the same even when brought nearer or farther from the perceiver. And, perceived shape stays the same across rotations. The simple observation that nonhuman animals manage to navigate their worlds is ample testimony to their ability to maintain perceptual constancy. Nonhuman animals also have been shown to “categorize” both visual images and acoustic stimuli such as speech (Herrnstein, 1984; Kluender, Diehl, & Killeen, 1987).



Both perceptual constancy and categorization can be treated as effectively the same process, differing only in level of abstraction. There is one obvious objection to equating perceptual constancy and categorization, this being that a literal identity relation is held between two presentations in perceptual constancy, but not in categorization. Kluender (1988, 1994) has argued that this apparent distinction is weak and of little functional importance. In addition, to the extent that veridical recovery of objects and events is impossible, perception of true identity is fantasy. Confusion between perceptual constancy and perceptual categorization is common in descriptions of speech perception as somehow arriving at appropriate phonetic categories. However, there have been a fair number of instances for which researchers adopted perceptual constancy as the preferred description (see e.g., Kuhl, 1978, 1979, 1980, 1987).

Categorization may be intuitively separate from constancy because members of categories (e.g., penguin, starling, emu) are more different from one another than for perceptual constancy (e.g., penguin at multiple angles or distances). For some curious reason, speech perception researchers often get this backward when they label as categorization the identification of stimuli that differ in small ways. Instead, here it is proposed that categorical perception be thought of as perceptual constancy. To the extent that categorization is only a more abstract manifestation of constancy, choosing constancy may not be a particularly provocative choice. If one considers perception of speech to involve perceptual constancy, commonalities with similar perceptual achievements are revealed and surplus cognitive content typically ascribed to categorization is avoided.

### 4.3. Categorical Perception

Categorical perception is the most well-known pattern of perceptual performance with speech sounds. Three features define categorical perception: a sharp labeling (identification) function, discontinuous discrimination performance (near-perfect across identification boundary and near-chance to either side), and the ability to predict discrimination performance purely on the basis of labeling data (Wood, 1976). All three defining features of categorical perception arise naturally from the principle of discovering (and perceptually absorbing) predictability in the interest of maximizing sensitivity to change.

Returning to the fact that speech sounds are comprised of multiple acoustic attributes, many of which are redundant, one acoustic attribute serves to predict the occurrence of another. Through experience, perceptual processes come to absorb these correlations in a way that increases efficiency. When perceptual systems encode correlations among attributes, there are two consequences. First, there is a decrease in sensitivity to differences between two sounds that share the same pattern of correlation among the same set of attributes. Second, two sounds with different patterns of correlation become easier to distinguish. For speech, detection of differences between functionally different speech sounds is optimized to the extent that perceptual processes

absorb redundancies across acoustic attributes that covary as properties of the same consonant or vowel (Figure 8).

#### 4.3.1. *Principal components analysis: An analogy*

This perceptual processing can be compared with the statistical technique Principal Component Analysis (PCA; see e.g., Dillon & Goldstein, 1984). For PCA, one begins with a correlation matrix of multiple variables, created to assess the degree to which each variable is correlated with every other variable across many observations. From this correlation matrix, it is possible to determine weighted combinations of variables, vectors, that account for as much shared variance as possible. To the extent that multiple observations reveal covariance among variables, a limited number of vectors (few relative to the number of variables) can account for a high percentage of the total variance across observations. PCA is being used here only as analogy because it is unlikely that real neurons adhere to formal restrictions on how vectors are chosen, and the ways PCA fails as analogy are themselves illuminating.

First, PCA is a linear analysis, and it is well-known that sensory processes are nonlinear. Second, PCA assumes normally distributed values, and the real world complies

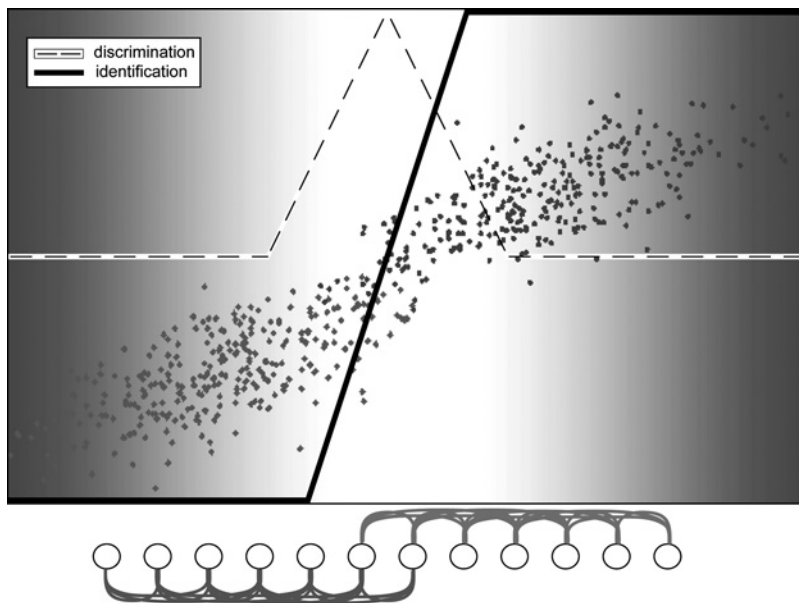


Figure 8. Categorical perception can be explained by auditory systems treating sounds with attributes that share the same correlation structure as equivalent. By this explanation, detection of differences between two complex sounds is greatest when each sound corresponds to a different pattern of experienced correlation.

with this assumption only to varying extents. A related analysis, independent component analysis (ICA; see e.g., Hyvärinen & Oja, 2000) does permit violations of assumption of normality, and may come a bit closer to modeling neural processing. Third, PCA, but not networks of neurons, requires that vectors be ordered from most to least amount of variance accounted for, and these vectors must be orthogonal (Eigenvectors.)

The issue concerning orthogonality is interesting in two ways. First, while maximal efficiency is achieved if every vector shares no variance with any other vector, achieving this goal is unlikely in a neural system. A second point is more informative. Here, perception is being construed as a cascade of processes, each working to extract redundancy from the outputs of earlier processes. To the extent that outputs of prior processes even approach orthogonality, this would seem to imply that seizing upon correlation again would become increasingly implausible. The solution to this seeming dead end is that, with every successive reduction of redundancy, information over which processing operates expands in space, frequency, time, and any other dimension of interest. Thus, statistical relationships that hold relatively locally do not constrain correlations at the next coarser grain of processing.

Finally, there is a parallel between practical use of PCA and the prior argument that perceptual systems should efficiently optimize, not maximize, information transmission. Figure 2c depicts a sensorineural system tuned to focus dynamic range in a way that optimizes information transmission relative to the distribution ( $1/f$ ) of energy flux in the environment while neglecting the last bits of potential information. At an extreme, PCA also permits characterization of all of the variance across observations when the number of Eigenvectors is equal to the number of observations. However, the virtue of PCA is that the majority of, but not all, variance can be accounted for by relatively few vectors. Efficiencies are gained by capturing correlations within a few vectors. More Eigenvectors are discarded than are saved, despite the fact that some snippets of information are lost.

From the analogy to PCA, it is easy to envision efficient sensorineural encoding that extracts reliable correlations between multiple attributes across observations. This process can be instantiated in perceptrons (Rosenblat, 1958), the simplest connectionist networks. Most important to the analogy is that experience is encoded efficiently in a way that decreases sensitivity to differences between stimulus inputs that share the same correlation vector (or attractor) and increases sensitivity to differences between inputs that correspond to different vectors (see e.g., Olshausen & Field, 1997; Simoncelli & Olshausen, 2001).

#### 4.3.2. *“Phonemes” as correlations?*

Through experience with redundant attributes, simplified encoding of inputs as patterns of correlation serves as grist for consequent processing. It could be argued that vectors in a correlation matrix correspond to the putative linguistic units called phonemes

(Kluender & Lotto, 1999); however, the same basic principles apply continuously along the chain of processing, and such a demarcation would be artificial. The grain of synthesis steadily increases, from attributes that are spectrally and temporally local to those that extend across multiple attributes derived from preceding processing through to lexical organization. For example, Kingston and Diehl's (1994) hypothesized intermediate perceptual properties (IPPs) may correspond to earlier analysis, and it is likely that statistical properties of acoustically simpler vowel and consonant sounds are extracted prior to those for more complex speech sounds. Further, as described below, additional reification of phoneme-like dimensions may await lexical organization.

Perhaps, no real harm may be done if one suggests that some correlations are phonemes *per se*, if only as a placeholder. However, it is important to distinguish two ways of thinking about this. The first way is common or even typical. One could suggest that the task for a listener is to identify consonants and vowels as individual psychological entities, and those entities are represented by correlations among attributes.

The second way, suggested here, is that consonants and vowels are revealed much more by what they are *not* than by what they are. Through experience, processes of speech perception become especially sensitive to acoustic differences that distinguish different consonants and vowels as a consequence ignoring differences among multiple acoustic properties that share a history of co-occurrence. What matters are distinctions between speech sounds, which are enhanced perceptually, not consonants and vowels themselves. We hear the sounds of a language by virtue of learning how they are distinguished from all other consonants and vowels. This idea was most explicitly expressed by linguists Roman Jakobson and Morris Halle (1971) in their classic book *Fundamentals of Language* "All phonemes denote nothing but mere otherness" (p. 22). Later in this chapter, this perspective becomes useful for understanding how infants learn to talk.

#### 4.4. Categorical Perception

Returning to categorical perception, one can understand how these patterns of performance emerge naturally from perceptual systems exploiting redundancies among attributes in the service of maximizing sensitivity to change. Following experience with correlations among acoustic attributes, listeners are relatively unlikely to detect differences among complex sounds that share the same correlation structure. The fact that modest changes are perceptually neglected is consistent with the fact that listeners can understand speech when some acoustic properties (e.g., energy above 1500 Hz) are absent. When the correlation structure is not violated too severely, perception overcomes perturbations, or even absence, of some attributes that normally contribute to the correlation. All that is required are sufficient attributes to get to the right vector.

This lack of sensitivity to perturbations among inputs accounts for the finding that discrimination performance is near chance for different instances of the same consonant or vowel. If the same correlation structure best fits two different stimuli, the same perceptual consequences obtain. Complementary to this lack of discrimination for sounds

that share the same correlation structure, discrimination performance is exquisite when the speech sounds to be discriminated are associated with two competing correlation structures. For these cases, discrimination is especially good because, by virtue of perceptual processes extracting redundancies within separate correlation structures, detection of change (information transmission) is optimized.

If categorical perception is only another example of perceptual constancy operating within general principles of perceptual organization, why did so many researchers (e.g., Studdert-Kennedy, Liberman, Harris, & Cooper, 1970) believe categorical perception to be unusual? The reason is because categorical perception was routinely contrasted with psychophysical data from experiments employing very simple stimuli (typically unidimensional) of limited ecological significance. Equally important, comparisons were made to stimuli with which subjects have little or no experience before coming to the experimental session. Classic psychoacoustic experiments using pure tones, noise bursts, and inharmonic complexes have great utility for interrogating operating characteristics of sensory transduction absent content or experience. Thresholds for energy detection or sensory change are valuable things to know, but these are not informative with respect to perception as it guides real activities in a real world.

When investigators use stimuli that are complex and familiar, signature response patterns of categorical perception are found. Categorical perception has been reported for musical intervals (Burns & Ward, 1974, 1978; Smith, Kemler Nelson, Grohskopf, & Appleton, 1994) and tempered triads (Locke & Kellar, 1973). Visually, humans categorically perceive human faces (Beale & Keil, 1995) and facial expressions (Etcoff & Magee, 1992; Calder, Young, Perrett, Etcoff, & Rowland, 1996; de Gelder, Teunisse, & Benson, 1997; Young, Rowland, Calder, Etcoff, Seth, Perrett, 1997), as well as faces of different species (Campbell, Pascalis, Coleman, Wallace, & Benson, 1997). When human observers are trained with artificial categories, they demonstrate increased perceptual sensitivity for items that are categorized differently (Goldstone, 1994). When monkeys are trained to respond differentially to clear examples of cats versus dogs (initially novel categories for monkeys), behavioral responses to stimuli along a morphed cat/dog series exhibit sharp crossovers at the series midpoint (Freedman, Riesenhuber, Poggio & Miller, 2001). Rather than being specific to speech, categorical perception is a general property of any perceptual system consequent to experience with rich regularities of natural objects and events.

Categorical perception appears to be an emergent property of any perceptual system that is shaped by experience. Damper and Harnad (2000) reviewed evidence from human and animal listeners as well as from neural network models. They concluded that any number of generalized learning mechanisms can account for categorical perception. Models ranging from simple associative networks (e.g., Anderson, Silverstein, Ritz, & Jones, 1977) and back-propagation networks with no hidden units (e.g., Damper, Gunn, & Gore, 2000) to multilayer networks (e.g., Harm & Seidenberg, 1999) all exhibit categorical perception. Because categorical performance arises from a variety of simple learning algorithms that seize upon reliable statistics in their inputs, Damper and Harnad

(2000) conclude that specialized processing is not necessary, and that “any general learning system operating on broadly neural principles ought to exhibit the essentials of [categorical perception] (p. 862)”.

#### 4.4.1. *Categorical perception by people with language impairments*

Based upon the fundamental and simple nature of categorical perception, as well as its ubiquity, one would expect individuals with language learning impairments also to exhibit this pattern of performance. Following several reports to the contrary (e.g., J. Sussman, 1993; Thibodeau & Sussman, 1979), Coady, Kluender, and Evans (2005) provided evidence that, in fact, children with specific language impairment (SLI) exhibit all three hallmarks of categorical perception. Children with SLI perform best, however, when speech sounds are most natural and meaningful (words) (Coady, Evans, Mainela-Arnold, & Kluender, in press).

#### 4.4.2. *Multi-modal interactions are expected*

Thus far, and for the remainder of this contribution, discussion typically will be restricted to auditory perception of speech. This should not be taken to imply that other modalities do not contribute to understanding speech. The approach outlined here is explicitly associationist and driven by experience. Whenever non-auditory information is redundant with auditory information, those correlations should contribute to efficient encoding of speech. For example, listeners have a wealth of experience simultaneously hearing speech and viewing talkers’ faces, and the McGurk effect (McGurk & MacDonald, 1976) is evidence of the profound effects visual information can have on the way speech sounds are perceived. Also, whenever people are talking, they both hear the sounds they are producing and they experience the movements in their own vocal tracts. While ideally, this occurs less than half the time people hear speech, simultaneous activities of both hearing and talking provide exquisite conditions for extraction of correlations.

## 5. EXPERIENCE AND SOUND CONTRASTS IN THE NATIVE LANGUAGE

### 5.1. Vowels

Experience is essential for the development of every sensorineural system. The profound role of experience is especially clear for speech perception. There are literally thousands of languages in use around the world, most without writing systems. Across a survey of only 317 representative languages, Maddieson (1984) describes 558 different consonants, 260 different vowels, and 51 diphthongs used by talkers around the world. Different languages use different subsets from this broad assortment of sounds, and acoustic distinctions that are communicatively necessary in one language should be ignored by speakers of another. Experience plays a critical role in tuning speech perception to the distributions of sounds within one’s language environment. Much, if not most,

of this development as a native listener takes place during the first year of life. Given the fact that multiple stimulus attributes come to be used collectively in speech perception, this leaves one with the question of how infants come to hear speech contrasts in a way that is appropriate to their native language environment.

One of the most challenging obstacles for the infant is to use acoustic information that distinguishes one speech sound from another in the face of sometimes widely varying acoustic properties, many of which do not distinguish speech sounds in their language. Acoustic differences that convey a contrast in one language may be of little or no relevance to another language. Some of these differences simply may be unrelated (orthogonal) to distinctions used in a particular language and would not occur in a language context. In addition, clearly audible differences such as gender of talker, speaking rate, emotional state, and other factors have profound effects on the acoustic signal, yet the language learner must learn to understand speech across these variations. Careful study of the ways in which infants overcome such challenges help to inform models of speech perception.

At least by the age of 6 months, infants have the ability to distinguish stimuli by vowel type even when different instances of the vowel differ considerably between presentations (Kuhl, 1983). In a reinforced head turn paradigm, Kuhl trained infants to turn their heads only when the vowel of the background stimulus changed during presentation of the closely related vowels [a] (as in 'tot') and [ɔ] (as in 'taught') spoken by a male talker. When tested on novel vowels produced by women and children (adding random variation in pitch contour in addition to shifting absolute frequencies of formants), infants provided the correct response on the first trial demonstrating that they recognized the novel instances as consistent with training vowels despite talker changes. Note that, by the shared covariance account offered above, the capacity to distinguish vowels across variation in irrelevant acoustic characteristics is a natural consequence of encoding stimuli on the basis of attributes that tend to co-occur. Attributes such as those accompanying changes in talker are irrelevant to particular consonants and vowels, so they do not play much role in phonetic distinctions.

While these studies attest to the ability of infants to respond to distinctions between vowels in the face of irrelevant variation, later studies have investigated how perception may be structured along acoustic/auditory dimensions that are directly relevant to distinguishing vowel sounds. What has become most apparent is that the degree to which infants treat as equivalent acoustically different instances of the same vowel is critically dependent on their experience with a particular language. For example, 6-month-old infants detect differences between vowel sounds differently depending on whether they lived in an English-speaking (Seattle) or Swedish-speaking (Stockholm) home (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

Further evidence for the role of experience can be found in experiments in which performance by European starlings (*Sturnus vulgaris*), having learned statistically-controlled distributions of renditions of Swedish and English vowels, was highly correlated with performance of adult human listeners (Kluender, Lotto, Holt, & Bloedel,

1998). A simple linear association network model, exposed to the same vowels heard by the birds, accounted for 95% of the variance in avian responses. Consistent with the principle that consonants and vowels are defined mostly by what sounds they are not, both human goodness judgments (Lively, 1993) and starling response rates illustrate an anisotropy such that peak responses are skewed away from competing vowel sounds more than they are defined by centroids of vowel distributions.

## 5.2. Consonants

Perception of differences between consonants is similarly tuned by experience. Werker and her colleagues (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Logan, 1985; Werker & Lalonde, 1988; Werker & Tees, 1983; Werker & Tees, 1984a, 1984b) have demonstrated that, as a function of experience with consonants in their native language, infants' tendency to respond to differences between some consonants that are not in their language begins to attenuate. The series of studies by Werker and Lalonde (1988) permits a relatively complete description of the phenomenon. They exploited the fact that speakers of English and Hindi use place of articulation somewhat differently for stop consonants. While for English, three places of articulation are used for voiced stop consonants: labial, alveolar, and velar (e.g. /b/, /d/, and /g/, respectively), in Hindi four places are used: labial, dental, retroflex, and velar (e.g. /b/, /ḍ/, /ḍ/, and /g/, respectively.) They created a synthetic series that varied perceptually from /b/ to /d/ (for native-English speaking adults) and from /b/ to /ḍ/ to /ḍ/ (for native-Hindi speaking adults).

Using the same reinforced head turn procedure used by Kuhl, they found that 6- to 8-month-old infants from English-speaking families responded to changes in stimulus tokens that crossed perceptually from the English /b/ to /d/ and also responded to changes between Hindi stops [ḍ] and [ḍ]. A different group of infants from English speaking families aged 11- to 13-months of age responded reliably only to the English [b]-[d] contrast, and not to the Hindi [ḍ]-[ḍ] contrast. Essentially, 6- to 8-month-old infants responded in a manner typical of by native-Hindi adults, while 11- to 13-month-olds responded like native-English adults treating both dental and retroflex stops as being the same. Werker and her colleagues have found analogous results in studies using different consonant contrasts from different languages (Figure 9).

For vowels and consonants, perception of speech is shaped during the first year of life in ways that respect the statistics of the linguistic environment. The challenge for the infant learning the speech sound distinctions in his or her native language is precisely this. Infants must learn how acoustic/auditory attributes tend to co-occur, and those correlations among attributes define perceptual organization that optimizes sensitivity to change.

### 5.2.1. Learning to talk

Here, it is worthwhile to revisit the claim that infants are learning distinctions between sounds, not consonants and vowels as entities *per se*. Information transmission is optimized by maximizing sensitivity to differences; this is the benefit of consolidating



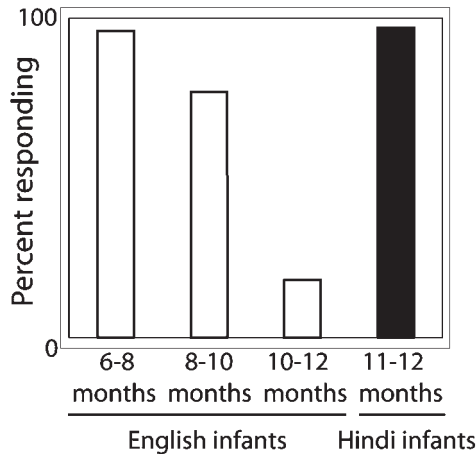


Figure 9. Perception of speech sounds is tuned by experience. As a consequence of experience with consonants in their native language, infants' tendency to respond to differences between some consonants that are not in their language begins to attenuate. Here, 6- to 8-month-old infants from English-speaking homes respond in a manner typical of native-Hindi adults and 11- to 12-month-old Hindi infants when hearing dental and retroflex Hindi stops. Before they are a year old, infants from English environments respond like native-English adults, treating both consonants the same. (Adapted from Werker and Lalonde, 1988.)

redundant attributes. Emphasizing the ways that sounds are different, versus how they are the same, helps illuminate issues concerning infants and young children learning how to produce speech sounds.

Owing to the developmental course of supralaryngeal anatomy and control, it is impossible for small developing vocal tracts to produce adult-like sounds of a language (e.g., Kent & Miolo, 1995; Kent & Vorperian, 2005; Vorperian, Kent, Lindstrom, Kalina, Gentry, & Yandell, 2005). The infant vocal tract begins more like the single tube not unlike that of a chimpanzee. While this configuration facilitates simultaneous drinking and breathing, it impedes production of many speech sounds. The larynx begins too high with a vocal tract too short, and will undergo drastic restructuring across the first 6 years (Vorperian, Kent, Gentry, & Yandell, 1999; Vorperian et al, 2005). What is a neotalker to do?

Mimicking speech sounds of the adult is not an option. However, it is possible for the developing vocal tract to produce sounds that are different in ways similar to how adult speech sounds differ. For example, shorter vocal tracts have resonances at higher frequencies than do longer vocal tracts, so center frequencies of formants are much higher for children than for full-size adults. This makes it impossible for children to produce vowel sounds with formants approximating the same frequencies as those for adults. Different vocal tract architectures make it fruitless for young children to try to match articulatory or auditory targets. However, the child is able to preserve acoustic contrasts in speech proportional to those heard from adult talkers. For example, the same relative

change in formant frequency can be produced irrespective of vocal tract size (Hillenbrand & Nearey, 1999). By definition, recasting the problem as change renders a solution that is inherently relative. One ancillary advantage is that the solution is unitless. In perceptual systems that have little or no access to absolute measures of anything, this quality is both attractive and essential.

### 5.3. Second-Language Perception

The same principles that explain categorical perception and development of perceptual organization during the first year of life extend to predicting how difficult, or easy, it is to learn new speech contrasts in a second language. For the case of the sounds of a single language, correlated attributes distinguish each consonant or vowel from others in a fashion that maximizes sensitivity to differences. The same construct, habitual co-occurrence of acoustic attributes, constrains and predicts how listeners perceive familiar and unfamiliar sounds from a second language. There are three basic patterns of interaction between perceptual organization for the native language and the mapping of sounds from a second language.

First, acoustic attributes of two contrasting non-native speech sounds can be equally well-correlated with attributes corresponding to only a single native consonant or vowel. Consider the case for formant patterns contributing to categorization of stop consonants with varying place of articulation. For example, both dental and retroflex stops such as those found in Hindi are acoustically realized in a manner quite similar to that for English alveolar stops. Given the range of ways English [d] is produced in the contexts of other speech sounds, there is ample overlap with acoustic attributes associated with both dental and retroflex stops. When [d] is produced in the environment of the retroflex continuant [ɻ] as in “drew”, English [d] shares multiple acoustic commonalities with the Hindi retroflex stop [ɖ]. Similarly, when English [d] is produced in the environment of a dental fricative such as [θ] in words like “width”, it is acoustically quite similar to the Hindi dental stop [ɗ] (Polka, 1991). Given the facts about the distributions of acoustic attributes for alveolar stops in fluent English, attributes consistent with dental or retroflex stops are well correlated with attributes that co-occur in alveolar stops. Dental-like or retroflex-like acoustic attributes are accommodated within correlation structures for English alveolar stops via an increase in overall variance reflective of the observed variability in English alveolar stop production. Werker and Lalonde’s (1988) adult identification data are entirely consistent with this scenario. Stimuli that are identified by native-Hindi listeners as dental or retroflex are all assimilated into the set of stimuli identified as alveolar by native-English listeners. Best, McRoberts, and Sithole, (1988) referred to a similar process as *single-category* assimilation in her taxonomy of contrasts within an articulatory framework.

An analogous example of difficulty perceiving a distinction between two sounds is the well-known inability of native-Japanese listeners to detect the distinction between English [r] and [l] (e.g., Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura 1975). Japanese sounds include a consonant called a *rhotic flap*, and acoustic

characteristics of these flaps overlap extensively with those of both English [r] and [l] (Sato, Lotto, & Diehl, 2003). Consequently, acoustic attributes of [r] and [l] are equally well-correlated with attributes corresponding to a single Japanese sound, and native-Japanese listeners are unable to hear, and consequently produce, the English distinction.

A related way non-native contrasts can be assimilated with native contrasts involves cases for which attributes of one sound from of a non-native contrast are very well-correlated with attributes of a single native consonant or vowel, being effectively the same. Attributes of a second non-native sound fit less well with the correlation structure of the same native sound, but they do correspond better with that sound than with any other native sound. One example of this is the Farsi distinction between velar and uvular stops. Native-English listeners do not lose the ability to discriminate Farsi velars from uvulars. Instead, they perceive the Farsi voiced velar and uvular stops as being good and poor instances, respectively, of English /g/ (Polka, 1992). In this case, Farsi velar stops are perceived as good English velar stops because they share most or all of the acoustic/auditory attributes that comprise the correlated structure of English /g/. Farsi uvular stops share fewer attributes with those that co-occur for English [g]. Farsi uvulars somewhat, but not completely, fit the correlation structure of English [g]. A related process has been referred to as *category-goodness* assimilation by Best et al. (1988).

The third way native and non-native contrasts can interact can be found in cases where the native language does not exactly share a contrast with a non-native language, but the native language does have a similar contrast that facilitates perception of the non-native contrast. For example, French does not include a voicing distinction for dental fricatives such as /ð/-/θ/ (as in “than” and “thank”), yet native-French listeners can discriminate voiced from voiceless English fricatives by perceiving them as versions of French dental stops /d̥/ and /t̥/, respectively (Jamieson & Morosan, 1986). Best et al. (1988) label this type of assimilation *two-category* because each sound of the non-native contrast maps more or less on to a different sound in the native language. Within the framework of correlated attributes, one would explain the fact that French listeners perceive the English fricatives as versions of French stops is because attributes of the dental fricatives are reasonably well-correlated with attributes of the French dental stops as produced with typical allophonic variation in fluent speech.

This scenario leaves one only those non-native distinctions that are roughly orthogonal to contrasts within the native language. Across the broad domain of possible mouth sounds, there always will remain some attributes that are not well-correlated with any of attributes for any speech sounds within a given language. For example, attributes of the click sounds of Zulu are correlated with no sounds of English, and their perception should be, and is relatively unaltered by the process of learning English phonemes (Best et al., 1988). It should be noted that all of the patterns of performance above are consistent with Flege’s (1995) speech learning model (SLM) and with patterns of experience-dependent learning of speech contrasts (e.g., Flege, Bohn, & Jang, 1997; Imai, Flege, & Wayland, 2002). However, Flege’s explanations of underlying processes that give rise to these patterns of performance are distinct at least in level of analysis, and he may or may not agree with specific aspects of the present authors’ explanations.

## 6. TO THE LEXICON AND BEYOND

### 6.1. Lexical Development and the Emergence of Phonemes (or something like them)

In the introduction to this chapter, discussion of phonetic segments and phonemes as independent entities was decidedly circumspect. Throughout the foregoing, consonants and vowels have been described either as sounds or as correlations among acoustic attributes in the service of maximizing information transmission. They have not been described as inherently linguistic or as a discrete stage in processing. To borrow Angell's (1907) dichotomy between functionalism and structuralism, discussion has been more about *how* and *why*, and less about the structuralist *what* of linguistic theory. Particular emphasis has been about *how*, and the focus now turns to *why*. The why of speech perception is to recognize words, and the end goal must be getting from the acoustic signal to words that have meaning. Within the information-theoretic perspective adopted here, one can construe the process of speech perception as one of successively reducing uncertainty sufficiently to arrive at words.

Over the years, some researchers have made a case that speech perception really is word perception without intermediate levels of analysis. There have been simulations of lexicons constructed directly from acoustic/auditory input *sans* phonemes (e.g., Klatt, 1980; Johnson, 2000), and a number of investigators have argued for the primacy of “holistic” (word-size) organization in lexical development (e.g., Charles-Luce & Luce, 1990, 1995; Jusczyk, 1986; Walley, 1993; Walley, Metsala, & Garlock, 2003). For example, Charles-Luce and Luce (1990) argued that emergence of acoustic/auditory detail, such as consonants and vowels of words, within the lexicon is a consequence of – not antecedent to – learning more words. By such an account, as the number of words in the lexicon grows, increasing degrees of detail are required to sort each word from all the others.

Kluender and Lotto (1999), by contrast, suggested that neither words nor phonetic units may serve exclusively to structure the developing lexicon. Werker and Curtin (2005), within their developmental framework for Processing Rich Information from Multidimensional Interactive Representation (PRIMIR), argue that infants have access to information for both phonetic units and the ways these units are grouped together into words. While PRIMIR is more of a framework than a detailed model, Werker and Curtin suggest that internal representations of phonemes become more firmly established and resistant to change as the lexicon grows.

The present emphasis of maximizing information transmission is consistent with concurrent development of statistical structures corresponding to consonants and vowels both preliminary to the lexicon and as an emergent property of lexical organization. In this chapter, speech perception has been described as a succession of processes operating on the acoustic signal with varying levels of complexity. Common to all these processes is maximizing efficiency of information transmission by absorbing redundancies across inputs. For example, spectral contrast operates early, with trading relations and

categorical perception operating later. Following these preliminary operations, one easily can imagine development of a nascent, but undetailed, lexicon of word forms. As a lexical space becomes increasingly populated, the covariance space becomes more complex. Predictable relationships among attributes, now with a phonetic segment-like grain, can be revealed resulting in a reduction in dimensionality of lexical items. The classic definition of phonemes is that they provide the minimal distinction between meaningful elements (morphemes) of language (Trubetskoy, 1969). Consequently, phonemes provide one of the most efficient ways to describe differences within the lexicon. Because phonemes provide efficient descriptors of lexical space, they emerge as dimensions of the developing lexical space by quite the same process that explains categorical perception, now operating over a larger time window.

One might suggest that positing phoneme-like dimensions as emergent properties of a lexical space obviates the need for anything resembling consonants and vowels preliminary to the lexicon. However, doing so violates principles of sensorineural organization, these being that redundancies are extracted continuously with ascending levels of processing. Werker and Curtin's (2005) proposal that phonemes become more established, and presumably increasingly tuned to phonotactic regularities in a language, is more consistent with persistent successive absorption of redundancy in the service of maximizing sensitivity to change.

### 6.1.1. *Note about dyslexia*

People with average and above average intelligence sometimes have a great deal of difficulty learning to read. The label, *dyslexia*, covers diverse disorders, many of which involve difficulties mapping from sounds to letters (e.g., Harm & Seidenberg, 1999; Joannisse, 2000). Speech sounds and putative phonemes each map inconsistently with orthography, but phonemes (as a linguistic abstraction) correspond sufficiently well for alphabetic writing systems to encourage belief in their existence (Port, in press). To the extent that phoneme-like dimensions are emergent properties of a lexical space, statistical approaches to extraction of redundant attributes may illuminate how one may think about dyslexia.

Return to the analogy of PCA with the lexical space conceptualized as a very large correlation matrix (with dimensionality reduced from that of raw auditory inputs following extraction of multiple redundancies through antecedent processing). Formal PCA yields a set of vectors through this matrix that optimally absorb correlations via the constraint of ordered Eigenvectors. In fact, for most every correlation matrix with much complexity, there are many near-optimal solutions. While the very best solution may capture 90% of the total variance using some fixed number of vectors, there would be many alternative solutions that capture 80–89% of the variance with the same number of vectors.

Now, imagine that people who learn to read normally simply are fortunate enough to have dimensions of their nascent lexical organization map relatively sensibly to orthography. People with dyslexia and equally strong intellect have great difficulty learning to

read because, during the years of developing their lexical space using auditory inputs alone, they developed perfectly functional lexical organization in a statistical sense, but their initial mapping of shared correlations was unfortunately skewed away from those that would map reasonably to letters. The present authors are not experts concerning dyslexia; however, it bears note that dyslexia-like etiology is consistent with, and even predicted by, the view of lexical development offered here.

## 6.2. Finding Word Boundaries

One final example of auditory perception using predictability to enhance sensitivity to change is found in studies demonstrating how infants find boundaries between words. In connected speech, acoustic realization of the beginning and end of one word also overlaps with sounds of preceding and following words. Unlike white spaces between words on a page, there are no silent intervals that mark beginnings and ends of words. Interestingly, perception is at odds with this acoustic reality. When listening to someone talk, most individual words stand out quite clearly as discrete entities. But listening to someone speak in a different language is often a very different experience. Every phrase or sentence may sound like a single very long word. This is the situation faced by infants (Figure 10).

Saffran and colleagues (Saffran, Aslin, & Newport, 1996) demonstrated that infants can use transitional probabilities between successive sounds within a speech stream as evidence for breaks between words. In their studies, they used streams of connected pseudowords, for which the probability of some sequences of consonant-vowels (CVs) was very high (1.0) while probability of other sequences was relatively low (0.33). Infants were sensitive to whether two sounds share a history of co-occurrence. When they heard successive CVs that rarely co-occurred with one another in their experience, they recognized this as a sign that there is a break between words. This discontinuity corresponds to a spike in information because one CV did not predict the occurrence of the next.

<p>tokibugopilagikobatipolutokibu  gopilatipolutokibugikobagopila  gikobatokibugopilatipolugikoba  tipolugikobatipolugopilatipolu  tokibugopilatipolutokibugopila  tipolutokibugopilagikobatipolu  tokibugopilagikobatipolugikoba  tipolugikobatipolutokibugikoba  gopilatipolugikobatokibugopila</p> <p>a</p>	<p><i>tokibugopilagikoba</i>TIPOLU<i>tokibu</i>  gopilaTIPOLU<i>tokibugikoba</i>gopila  <b>gikoba</b><i>tokibugopila</i>TIPOLU<b>gikoba</b>  TIPOLU<b>gikoba</b>tipolugopilaTIPOLU  <i>tokibugopila</i>TIPOLU<i>tokibugopila</i>  TIPOLU<i>tokibugopila</i><b>gikoba</b>TIPOLU  <i>tokibugopila</i><b>gikoba</b>TIPOLU<b>gikoba</b>  TIPOLU<b>gikoba</b>TIPOLU<i>tokibugikoba</i>  gopilaTIPOLU<b>gikoba</b><i>tokibugopila</i></p> <p>b</p>
--	--

Figure 10. There are no acoustic markers between most words in a stream of fluent speech, analogous to the strings of letters on the left (a). Following minutes of experience with streams of connected speech in which probability of some sequences of consonant–vowels (CVs) is very high, while probability of other sequences was relatively low, infants are sensitive to whether two sounds share a history of co-occurrence (Saffran, Aslin, & Newport, 1996).

Statistics of English support this emphasis on word boundaries, as the ends of most words cannot be identified prior to the onset of the next (Luce, 1986). Infant sensitivity to boundaries is yet another example of using predictability to enhance sensitivity to change, and hence enhance transmission of information.

Because this is a principle of perceptual systems most broadly, one expects this use of predictability to apply most generally. Indeed, these patterns of performance extend to infants experiencing tonal sequences (Saffran, Johnson, Aslin, & Newport, 1999), visual shapes (Kirkham, Slemmer, & Johnson, 2002), and visual feature combinations (Fiser & Aslin, 2002). In fact, even nonhuman primates (Hauser, Newport, & Aslin, 2001) exhibit this sensitivity to transitional probabilities.

## 7. SPEECH IN THE BRAIN

Given the plethora of relatively recent studies concerning speech processing using electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI), extensive depiction of speech processing in cerebral cortex alone requires one or more chapters. Here, a very brief review of current understanding will be placed within the framework of information processing provided above.

Hearing sounds of any kind activates primary auditory cortex (AI). Processing of complex sounds relies on additional areas of cortex adjacent to AI, which is functionally divided into ventral, anterior, and posterior sections. Neurons from AI project to the surrounding *belt* of cortex, and neurons from this belt synapse with neurons in the adjacent *parabelt* area. Just about any sound will cause activation in some part of AI. However, in the belt and parabelt areas, referred to as “secondary” or “associational” auditory areas, simple sounds such as sine waves and white noise elicit less activity, particularly if they have limited temporal structure. Thus, as in the visual system, processing proceeds from simpler to more complex stimuli farther along the auditory pathway, and there is also greater evidence of cross-modal processing (e.g., combining acoustic and optic information), particularly in parabelt areas. Of course, this general property of hierarchical organization is consistent with continuous and successive extraction of redundancy across increasing spans of space and time. As one might expect, areas beyond AI are activated when listeners hear speech and music. Further, at these early levels of cortical processing, activity in response to music, speech, and other complex sounds is relatively balanced across the two hemispheres.

When listening to speech, additional areas of both left and right superior temporal lobes along the superior temporal sulcus (STS) activate more strongly in response to speech than to non-speech sounds such as tones and noise (e.g., Binder, 2000). While language processing is typically lateralized to one hemisphere, this activity in response to speech signals is relatively balanced across both sides of the brain when researchers have been very careful to avoid higher-level effects of language (Zatorre & Binder, 2000).

At some point, however, processing of speech should become more lateralized as perceiving speech becomes part of understanding language. However, one challenge for researchers has been to create control stimuli that have all the complex properties of speech without being heard as speech. Because listeners are very good at understanding even severely distorted speech, it is very difficult to construct stimuli that are complex like speech without being heard as speech.

Liebenthal and colleagues (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005) adopted a creative way to control for acoustic complexity while varying whether sounds would be heard as speech. They synthesized speech syllables varying incrementally from [ba] to [da]. Nonspeech control stimuli were the same series of syllables, except characteristics of  $F_1$  transitions were flipped upside down, decreasing in center frequency following syllable onset. It is impossible for a human vocal tract to create such sounds, and listeners could not identify them as consonants.

During scanning, listeners participated in a categorical perception task, discriminating pairs of stimuli from one or the other series of stimuli. For the [ba-da] speech series, performance was typical for categorical perception experiments, with stimuli that would be labeled differently (e.g., /ba/ versus /da/) being almost perfect, while discrimination of other stimulus pairs was rather poor. For the nonspeech sounds, discrimination was above chance and fairly equivalent for all pairs of stimuli. Listening to both series of stimuli resulted in increased activation in STS in both temporal lobes. When sounds were speech [ba-da] however, there was increased activation in STS superior temporal cortex, a bit anterior from activation for control stimuli and mostly in the left hemisphere. Liebenthal et al. (2005) may have revealed a place in the left temporal lobe where experience with English [b] and [d] shapes neural activity (Figure 11).

Another study demonstrates how these results might suggest the next step of processing of speech into language. Scott and colleagues (Scott, Blank, Rosen, & Wise, 2000) controlled for acoustic complexity while changing whether sentences were intelligible. They created *spectrally rotated* sentences that could not be understood by listeners. Rotated speech signals were spectrographically upside down (*i.e.*, the frequency scale was inverted). They could not be understood by listeners, but they were as complex acoustically as right-side-up sentences. Rotated stimuli activated STS comparably to intact non-rotated sentences, suggesting that auditory processing in these areas is related more to complexity than to being speech *per se*. The essential difference between cortical responses to these two types of stimuli was that, on the left temporal lobe, activation in response to intact sentences continued further anterior and ventral (including superior temporal gyrus, STG) to the region activated by rotated sentences. Activation was somewhat anterior and ventral to the area in which Liebenthal and colleagues (2005) found activation for [b] and [d]. Relative left lateralization consequent to recognizability parallels findings for other environmental sounds (e.g., Lewis, Wightman, Brefczynski, Phinney, Binder, & DeYoe, 2004), albeit in different brain regions.



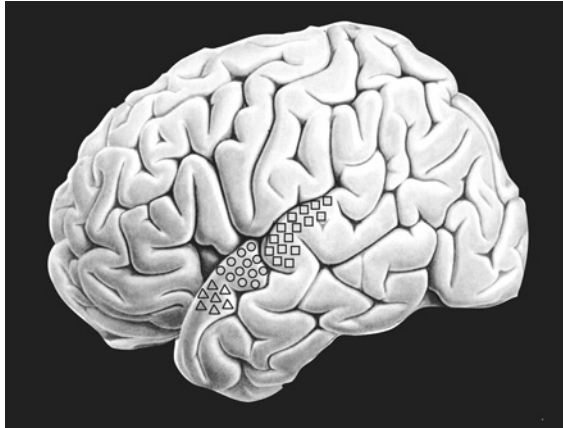


Figure 11. When listening to speech, as well as other complex sounds, cortical activity along the superior temporal sulcus (STS, squares) is relatively balanced across the hemispheres. When acoustic complexity is carefully controlled, there is evidence of increased activation in STS (circles), more anterior and mostly in the left hemisphere (Liebenthal, Binder, Spitzer, & Possing, 2005). Cortical activation in response to recognizable sentences appears further anterior and ventral (triangles), including superior temporal gyrus (STG) (Scott, Blank, Rosen, & Wise, 2000).

The extension of lateralized activation, increasingly anterior and ventral, discovered by Scott and her colleagues (2000) inspires a tantalizing hypothesis concerning the next level of processing beyond speech perception to word recognition. Based upon neuropsychological data from patients, cortical areas essential to semantic memory may reside within far anterior temporal lobe (see e.g., Rogers et al., 2004). Consider a scenario through which increasingly sophisticated redundancies are wrenched out of the speech signal, through differentiation of sounds such as [b] and [d], through detection of word boundaries, to regularities within word boundaries (*i.e.*, words) and concomitant associations with semantic properties of words. Such a notion is clearly speculative given the present state of knowledge; however, this would be an elegant view of successive processing along increasingly anterior and ventral areas of temporal lobe.

## 8. CONCLUSION

Speech perception is grounded in general principles that apply to other acoustic events and to other modalities. Classic principles that guide perception, none of which is wholly original to the authors, explain processes underlying multiple phenomena of speech perception. This information-theoretic model, operating from sensory transduction to word learning, is biologically realistic. It is intended that this framework will serve, not only to reveal processes underlying normative aspects of speech perception, but also to extend understanding of clinical conditions of speech and language processing. Finally,

beyond being amenable to study like any other form of perception, speech perception holds promise as a fertile domain for research that can reveal and extend fundamental understanding of perception most generally.

## ACKNOWLEDGMENTS

Comments provided by Joshua Alexander, Randy Diehl, Ray Kent, Timothy Rogers, Mark Seidenberg, and Christian Stilp on part of all of this chapter are greatly appreciated, but the authors have not accepted all their advice and are alone responsible for any shortcomings of editing or content. This Chapter was written in December 2005. The Work was supported by NIDCD DC04072 and SSHRC-410-2003-1819.

## REFERENCES

- Alexander, J. M., & Kluender, K. R. (2005). Contributions of gross spectral properties to perception of stop consonants. *Journal of the Acoustical Society of America*, *118* (3), (Part 2) 1993.
- Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.
- Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature*, *434*, 79–83.
- Angell, J. R. (1907). The province of functional psychology. *Psychological Review*, *14*, 61–91.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.
- Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results*. New York: Henry Holt and Company, Inc.
- Barbour, D. L., & Wang, X. (2003). Contrast tuning in auditory cortex. *Science*, *299*, 1073–1075.
- Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In: W. A. Rosenblith (Ed.), *Sensory communication* (pp. 53–85). Cambridge, MA: MIT Press.
- Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Philosophical Transactions of the Royal Society of London B*, *352*, 1141–1147.
- Barlow, H. B. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, *24*, 602–607.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

- Berkeley, G. (1837). *The works of George Berkeley*. London: Tegg (original works published in 1709, 1710, 1713).
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology, 14*, 345–360.
- Binder, J. R. (2000). The neuroanatomy of speech perception (Editorial). *Brain, 123*, 2371–2372.
- Burns, E. M., & Ward, J. D. (1974). Categorical perception of musical intervals. *Journal of the Acoustical Society of America, 55*, 456.
- Burns, E. M., & Ward, W. D. (1978). Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*, 456–468.
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*, 81–117.
- Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., & Benson, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proceedings of the Royal Society of London, Series B–Biological Sciences, 264*, 1429–1434.
- Cardozo, B. L. (1967). Ohm's law and masking. *IPO Annual Progress Report, Institute for Perception Research, 2*, 59–64.
- Cathcart, E. P., & Dawson, S. (1928–1929). Persistence 2. *British Journal of Psychology, 19*, 343–356.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17*, 205–215.
- Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language, 22*, 727–735.
- Christman, R. J. (1954). Shifts in pitch as a function of prolonged stimulation with pure tones. *American Journal of Psychology, 67*, 484–491.
- Coady, J. A., Kluender, K. R., & Rhode, W. S. (2003). Effects of contrast between onsets of speech and other complex spectra. *Journal of the Acoustical Society of America, 114*(4), 2225–2235.
- Coady, J. A., Kluender, K. R., Evans, J. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research, 48*, 944–959.
- Coady, J. A., Evans, J. L., Mainela-Arnold, E., & Kluender, K. R. (in press). Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. *Journal of Speech, Language, and Hearing Research*.
- Damper, R. I., Gunn, S. R., & Gore, M. O. (2000). Extracting phonetic knowledge from learning systems: Perceptrons, support vector machines and linear discriminants. *Applied Intelligence, 12*, 43–62.

- Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics*, *62*, 843–867.
- De Gelder, B. Teunisse, J. P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion*, *11*, 1–23.
- Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory nerve fibers. *Journal of the Acoustical Society of America*, *68*, 843–857.
- Delgutte, B. (1986). Analysis of French stop consonants with a model of the peripheral auditory system. In: J. S. Perkell, & D. H. Klatt (Eds), *Invariance and variability of speech processes* (pp. 131–177). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Delgutte, B. (1996). Auditory neural processing of speech. In: W. J. Hardcastle, & J. Laver (Eds) *The handbook of phonetic sciences*. (pp. 507–538). Oxford: Blackwell.
- Delgutte, B., Hammond, B.M., Kalluri, S., Litvak, L.M., & Cariani, P.A. (1996). Neural encoding of temporal envelope and temporal interactions in speech. In: W. Ainsworth, & S. Greenberg (Eds), *Auditory basis of speech perception* (pp. 1–9). European Speech Communication Association Keele, UK.
- Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve IV: Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America*, *75*, 897–907.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.
- Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis, methods and applications*. New York: Wiley.
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227–240.
- Festen, J. M., & Plomp, R. (1981). Relations between auditory functions in normal hearing. *Journal of the Acoustical Society of America*, *70*, 356–369.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA* *99*, 15822–15826.
- Flege, J. E. (1995). Second language speech learning: Theory, findings and Problems. In: W. Strange York, & M. D. Timonium (Eds), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–272).
- Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470.
- Fletcher, H. (1953/1995). *Speech and hearing in communication*. New York: Krieger.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics*, *48*, 559–570.
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. (2001). Categorical perception of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.
- Gardner, W. R., & Hake, H. W. (1951). The amount of information in absolute judgments. *Psychological Review*, *58*, 446–459.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178–200.
- Gordon, C., Webb, D. L., & Wolpert, S. (1992). One cannot hear the shape of a drum. *Bulletin of the American Mathematical Society*, *27*, 134–138.
- Green, D. M., McKey, M. J., & Licklider, J. C. R. (1959). Detection of a pulsed sinusoid in noise as a function of frequency. *Journal of the Acoustical Society of America*, *31*, 1146–1152.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528.
- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–B64.
- Helmholtz, H. v. (1866/1969). The recent progress of the theory of vision. In: R. M. Warren & R. P. Warren (Eds) *Helmholtz on perception: Its physiology and development* (pp. 61–138). New York: Wiley.
- Herrnstein, R. J. (1984). Objects, categories and discriminative stimuli. In: H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds), *Animal cognition* (pp. 233–261). Hillsdale, NJ: Erlbaum.
- Hillenbrand, J., & Nearey, T. M. (1999). Identification of resynthesised /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America*, *105*, 3509–3523.
- Hoagland, H. (1933). Quantitative aspects of cutaneous sensory adaptation I. *Journal of General Physiology*, *16*, 911–923.
- Holt, L. L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science*, *16*, 305–312.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, *108*(2), 710–722.
- Hood, J. D. (1950). Studies in auditory fatigue and adaptation. *Acta Oto-Laryngology Supplement*, *92*, 1–57.
- Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America*, *51*, 1885–1894.

- Houtgast, T. (1974). Auditory analysis of vowel-like sounds. *Acustica*, 31, 320–324.
- Hume, D. (1748/1963). An enquiry concerning human understanding. In: C. W. Eliot (Ed.), *The Harvard classics*. New York: P. F. Collier & Son Corporation (original published in 1748 entitled *My own life*).
- Hyvarinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13, 411–430.
- Imai, S., Flege, J., & Wayland, R. (2002). Perception of cross -language vowel differences: A longitudinal study of native Spanish learners of English. *Acoustical Society of America Journal*, 111(5), 2364–2364.
- Jakobson, R., & Halle, M. (1971). *The fundamentals of language*. Mouton: The Hague.
- James, W. (1897/1956). *The will to believe and other essays in popular philosophy*. Mineola, NY: Dover Publications, Inc.
- Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: acquisition of English /q/-/d/ contrast by francophones. *Perception and Psychophysics*, 40, 205–215.
- Jenison, R. L. (1997). On acoustic information for auditory motion. *Ecological Psychology*, 9(2), 131–151.
- Joanisse, M. F. (2000). *Connectionist phonology*. Unpublished Ph.D. dissertation, University of Southern California.
- Jusczyk, P. W. (1986). Towards a model for the development of speech perception. In: J. Perkell, & D. H. Klatt (Eds), *Invariance and variability in speech processes* (pp. 1–19). Hillsdale, NJ: Erlbaum.
- Kent, R. D., & Miolo, G. (1995). Phonetic abilities in the first year of life. In: P. Fletcher, & B. MacWhinney (Eds), *The handbook of child language* (pp. 303–333). New York: Blackwell.
- Kent, R. D., & Vorperian, H. K. (2005). In the mouths of babes: anatomic, motor, and sensory foundations of speech development in children. In: R. Paul (Ed.), *Language disorders from a developmental perspective*. Mahwah, NJ: Lawrence Erlbaum.
- Kiefte, M., & Kluender, K. R. (2005a). The relative importance of spectral tilt in monophthongs and diphthongs. *Journal of the Acoustical Society of America*, 117(3), 1395–1404.
- Kiefte, M., & Kluender, K. R. (2005b). Cancellation of reliable spectral characteristics in auditory perception. *Journal of the Acoustical Society of America*.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70, 419–454.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42.

- Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67(3), 971–995.
- Klatt, D. H. (1982). Acoustic, Speech, and Signal Processing, *IEEE International Conference on ICASSP '82*, 7, 1278–1281.
- Kluender, K. R. (1988). *Auditory constraints on phonetic categorization: Trading relations in humans and nonhumans*. Unpublished Ph.D. dissertation, University of Texas at Austin.
- Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In: M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173–217). San Diego, CA: Academic.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195–1197.
- Kluender, K. R., & Lotto, A. J. (1999). Virtues and perils of empiricist approaches to speech perception. *Journal of the Acoustical Society of America*, 105, 503–511 (invited contribution, peer reviewed).
- Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America*, 104, 3568–3582.
- Kuhl, P. K. (1978). Perceptual constancy for speech-sound categories. *N.I.C.H.D. Conference on Child phonology: perception, production, and deviation*, Bethesda, MD.
- Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66, 1668–1679.
- Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In: G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds) *Child phonology Vol. 2: Perception*. New York: Academic Press.
- Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6, 263–285.
- Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In: P. Salapatek, & L. Cohen (Eds), *Handbook of infant perception* (Vol. 2, pp. 257–381). New York: Academic Press.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants six-months of age. *Science*, 255, 606–608.
- Lewis, J. W., Wightman, F., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, 14, 1008–21.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, 21, 1–36.

- Licklider, J. C. R., & Miller, G. A. (1951). The perception of speech. In: S. S. Stevens (Ed.) *Handbook of experimental psychology* (pp. 1040–1074). New York: Wiley.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*(10), 162–163.
- Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, *35*, 1773–1781.
- Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, *42*, 830–843.
- Lippman, R. P. (1996). Speech perception by humans and machines. In: W. Ainsworth, & S. Greenberg (Eds), *Workshop on the auditory basis of speech perception* (pp. 309–316). Keele, England: Keele University.
- Lisker, L. (1978). Rapid versus rabid: A catalogue of acoustical features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research*, *SR-54*, 127–132.
- Lively, S. E. (1993). An examination of the perceptual magnet effect. *Journal of the Acoustical Society of America*, *93*, 2423.
- Locke, S., & Kellar, L. (1973). Categorical perception in a non-linguistic mode, *Cortex*, *9*, 355–369.
- Lotto, A. J. (2000). Reply to “An analytical error invalidates the ‘depolarization’ of the perceptual magnet effect.” *Journal of the Acoustical Society of America*, *107*, 3578–3580.
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*, 602–619.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, *102*, 1134–1140.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- Maddieson, I. (1984). *Patterns of sound*. Cambridge: Cambridge University Press.
- Mann, V. A. (1980). Influence of preceding liquid in stop-consonant perception. *Perception & Psychophysics*, *28*, 407–412.
- Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners’ perception of English “l” and “r”. *Cognition*, *24*, 169–196.
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, *69*, 548–558.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.



- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, *63*, 81–97.
- Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of (r) and (l) by native speakers of Japanese and English. *Perception and Psychophysics*, *18*, 331–340.
- Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulas for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, *74*, 750–753.
- Morais, J., Bertelson, P., Cary, L., Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*, 45–64.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323–331.
- Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*, 3241–3254.
- Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, *37*, 3311–3325.
- Peirce, C. S. (1878). How to make ideas clear. *Popular Science Monthly*, *12*, 286–302.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, *89*(6), 2961–2977.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, *52*(1), 37–52.
- Port, R. F. (in press). The graphical basis of phones and phonemes. In: M. Munro, & O. -S. Bohn (Eds), *Second-language speech learning: The role of language experience in speech perception and production*.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*, 947–950.
- Repp, B. H. (1982). Phonetic trading relations and context effects: New evidence for a speech mode of perception. *Psychological Bulletin*, *92*, 81–110.
- Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America*, *43*(6), 495–501.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*, 205–235.
- Rosenblat, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*, 386–408.

- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Sato, M., Lotto, A. J., & Diehl, R. L. (2003). Patterns of acoustic variance in native and non-native phonemes. *The Journal of the Acoustical Society of America*, *114*(4), 2392.
- Schouten, J. F. (1940). The residue, a new component in subjective analysis. *Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen*, *43*, 356–365.
- Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature: Neuroscience*, *4*, 819–825.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.
- Seidenberg, M. S., & Gonnerman, L. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*, 353–361.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.
- Shannon, R. V., Zeng, F. G., Wygonski, J., Kamath, V., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.
- Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1215.
- Smith, R. L. (1979). Adaptation, saturations, and physiological masking in single auditory-nerve fibers. *Journal of the Acoustical Society of America*, *65*, 166–178.
- Smith, R. L., Brachman, M. L., & Frisina, R. D. (1985). Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments. *Journal of the Acoustical Society of America*, *78*, 1310–1316.
- Smith, J. D., Kemler Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, *52*, 23–54.
- Smith, R. L., & Zwislocki, J. J. (1971). Responses of some neurons of the cochlear nucleus to tone-intensity increments. *Journal of the Acoustical Society of America*, *50*, 1520–1525.
- Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In: P. D. Eimas, & J. L. Miller (Eds), *Perspectives in the study of speech*. Hillsdale, NJ: Erlbaum.
- Summerfield, Q., Haggard, M. P., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception and Psychophysics*, *35*, 203–213.

- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 77, 234–249.
- Summerfield, Q., Sidwell, A., & Nelson, T. (1987). Auditory enhancement of changes in spectral amplitude. *Journal of the Acoustical Society of America*, 81, 700–707.
- Sussman, J. E. (1993). Perception of formant transition cues to place of articulation in children with language impairments. *Journal of Speech and Hearing Research*, 36, 1286–1299.
- Thibodeau, L. M. Sussman, H. M. (1979). Performance on a test of categorical perception of speech in normal and communication disordered children. *Journal of Phonetics*, 7, 375–391.
- Trubetzkoy, N. S. (1969). *Principles of phonology* (C. Baltaxe, Trans.) Berkeley: University of California Press (original work published in 1939).
- Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology Hum Percept Perform*, 20(1), 3–16.
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6, 391–398.
- Urbantschitsch, (1876), cf. Abrahams, Krakauer & Dallenbach (1937). Gustatory adaptation to salt. *American Journal of Psychology*, 49, 462–469.
- Viemeister, N. F. (1980). Adaptation of masking. In: G. van den Brink, & F. A. Bilsen (Eds), *Psychophysical, physiological, and behavioral studies in hearing* (pp. 190–197). Delft: University Press.
- Viemeister, N. F., & Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *Journal of the Acoustical Society of America*, 71, 1502–1507.
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117(1), 338–350.
- Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results. *International Journal of Pediatric Otorhinolaryngology*, 49, 197–206.
- Voss, R. F., & Clarke, J. (1975). '1/f noise' in music and speech. *Nature*, 258, 317–318.
- Voss, R. F., & Clarke, J. (1978). "1/f noise" in music: Music from 1/f noise. *Journal of the Acoustical Society of America*, 63, 258–263.
- Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286–350.

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: It's role in the development of phoneme awareness and early reading ability. *Reading and Writing, 16*, 5–20.

Weiner, N. (1948). *Cybernetics*. New York: Wiley.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*, 349–355.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24*, 672–683.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37*, 35–44.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37*, 278–286.

Werker J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*, 1866–1878.

Wightman, F. L., McGee, T., & Kramer, M. (1977). Factors influencing frequency selectivity in normal and hearing-impaired listeners. In: E. F. Evans, & J. P. Wilson (Eds), *Psychophysics and physiology of hearing* (pp. 295–310). New York: Academic Press.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60*, 1381–1389.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A. & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotional recognition. *Cognition, 63*, 271–313.

Zatorre, R. J., & Binder, J. (2000). Functional and structural imaging of the human auditory system. In: A. Toga & J. Mazziotta (Eds), *Brain mapping the systems* (pp. 365–402). San Diego, CA: Academic Press.

Zwaardemaker, H. (1895). *Die physiologie des geruchs*. Leipzig: Engelmann.

This page intentionally left blank

## Chapter 7

### The Perception of Speech

*Jennifer S. Pardo and Robert E. Remez*

A talker can expect a listener to grasp the rough dimension of any sincere and appropriate message, though only by saying it. For talker and listener, speech is a medium, a link in a commonplace causal chain by which pleasantries or philosophies are exchanged, cooperation is negotiated and compliance is compelled. But, does an essay about speech belong in a book about language? To a newcomer, it is self-evident that conversational partners know what each other says simply by hearing the sounds of spoken words. From this perspective, the fundamentals of speech perception surely lie in psychoacoustics, an essential reduction of speech perception to sensory resolution and auditory categorization. Even so, the newcomer might already notice the difference in auditory quality in the speech of children and adults, or in face to face and in telephone speech, and suspect that the perception of spoken messages entails more than acute hearing. To the old hand familiar with cognitive psychology and the historic place of speech within it, the motivation to study speech perception might seem well and truly relieved now that affordable devices transcribe words from sound. On the contrary, this essay like its companions in this volume was produced by a typing hand and not by a dictating voice, despite the mathematical ingenuity of the engineers – far exceeding that of cognitive psychologists – who create speech-to-text devices. For the reader of any degree of experience, this part of the *Handbook* explains why the descriptive and theoretical puzzles provoked by speech perception have proven to be so enduring, psychologically and linguistically, and in doing so claims a role for speech in language.

Our characterization of the perception of speech ranges across three of its facets. First, we discuss the historic aim of research on speech, which has been to understand how acoustic properties evoke an impression of linguistic form. This line of research is mature, and a sizeable literature beginning with classical sources presents a consistent expression of competing views and evidence. Ideological commitments aside, it is a singular merit of this research tradition that it introduced a generous assortment of theoretical conceptualizations to perceptual psychology. Even when innovation happened to spring from other sources, the well established techniques and research paradigms within the study of speech perception provided a ready means to calibrate the explanatory

adequacy of a principle. This portion of the essay exposes contemporary viewpoints about perceptual organization and analysis of speech and notes the questions that lead the research forward.

Second, the ordinary perceptual resolution of the linguistic properties of speech is accompanied by an irreducible impression of the talker as well as the message. Research about the recognition of individuals from their speech takes its origin in forensic projects – studies to determine whether a known talker and an unidentified talker are the same – and in artifactual methods to create a vocal identification technology. In contrast to these humble roots, more recent cognitive studies emphasize the perceptual effects of variation in phonetic form across individuals and instances. The evident perceptual interchange of linguistic, individual (or, indexical) and situated properties promised to overturn the classic conceptualization of the acoustic-to-phonetic projection, and this portion of our essay describes the partial success of this project and the questions that remain for a complete causal account.

The third section of our essay characterizes self-regulatory speech perception in which an individual talker's self-perception modulates the production of speech. This theme is contrary to Lashley's founding arguments in psycholinguistics. He held that the rate of production of vocal actions was too rapid to permit monitoring by proprioception, and many studies since have recounted adequate unmonitored articulation, for instance, concurrent to mandibular somatic sensory blockade. This literature about the control of coordination in vocal movement is supplemented and elaborated by more recent studies that identified effects of self-monitoring in other sensory modalities. These findings show that talkers adjust subtle – and, less subtle – properties of articulatory expression as a consequence of phonetic perception, albeit at a slower pace than Lashley stipulated, and in varied social conditions.

Throughout, our essay is organized by psycholinguistic questions, rather than by concerns with specific research methods. Although the investigations that we describe are largely the yield of functional studies of normal adults, we have referred to research about special populations or using special methods when we aimed to secure premises in our argument. We also direct the reader to other discussions when technical matters or special perceivers hold intrinsic interest or importance.

## **1. PERCEPTUAL ORGANIZATION AND ANALYSIS OF SPEECH**

A listener intent on grasping a talker's message must sample physical effects of speech that vary regularly if unpredictably, a consequence of a talker's vocal acts. The regularity as well as the unpredictability derive from a common cause; the linguistic governance of speech deploys formal attributes designated in the talker's language, and these drive the regularities. At the same time, no expression is an exact repetition of a prior one, and whether the departures from stereotypy are attributed to chance or to a specific cause – to a talker's enthusiasm, or haste, or influenza – exact patterns never recur. The central

problem in research on speech has been to understand how perception of regular linguistic attributes is evoked by such unpredictably varying acoustic causes.

None of the acoustic constituents of speech is unique to speech, although some features of speech are characteristic: a cyclical rise and fall of energy associated with a train of syllables, amplitude peaks and valleys in the short-term spectrum, and variation over time in the frequency at which the peaks and valleys occur (Stevens & Blumstein, 1981). In addition to noting these attributes, it is fair to say that natural speech is an acoustic composite of whistles, clicks, hisses, buzzes and hums, a discontinuous and often aperiodic result of the continuous movement of articulators. In following a speech signal, a listener tracks an intermittent pattern of heterogeneous acoustic constituents; there is no single element nor set of them that defines speech, therefore, no simple way for a perceiver to distinguish speech piecemeal from the acoustic effects of other sources of sound. Despite all, a perceiver often tracks the speech of a specific talker sampling by ear and eye, two kinds of *perceptual organization* that also combine multimodally, and resolves the linguistic properties in the sensory effects – that is to say, *perceptual analysis* of the symbolic properties of speech succeeds. We discuss these in turn.

### 1.1. Perceptual Organization

The ability to track an individual's speech amid other sounds retains the characterization applied long ago by Cherry (1953), the *cocktail party problem*. Such get-togethers can pose many challenges for participants; this specific cocktail party problem is solved by perceivers who understand spoken messages despite the concurrent intrusions of acoustic elements very much like those composing the target speech stream. The sources of unrelated sounds surely include the clinking of glasses and popping of corks, although other extraneous acoustic moments are similar to an attended speech stream because they come from the speech of other talkers. Indoors, the direct sound mixes with late arriving reflections from the ceiling, floor and walls of the attended speech signal itself.

To gauge the means of resolving the sound produced by a single individual, the contrast between visual and auditory attention is instructive. In attending to a visible object or event, a perceiver typically turns to face it, bringing the light reflected by the object of interest to the fovea of the retina. In this retinal region, receptors are densest and pattern acuity is best, for which reasons visual attention will often coincide with a foveated object. A listener's attention to the audible world achieves spatial and spectral focus psychologically, without the selective benefit of a heading at which auditory pattern acuity peaks. In addition, the visible world contains opaque, translucent and transparent objects; the audible world is largely transparent. A listener cannot presume that a sound arriving from a certain direction stems from the visible object at the same heading, for sounds produced by other sources at the same direction are likely to propagate around intermediate objects to impinge on the listener. Despite all, perception often reciprocates the patterned variation of a speech stream with its discontinuities, dissimilarities among components and similarities between its components and those of unattended utterances and other



events. This perceptual function is fast, unlearned, keyed to complex patterns of sensory variation, tolerant of anomalous sensory quality, nonsymbolic and dependent on attention whether elicited or exerted (Remez, Rubin, Berns, Pardo, & Lang, 1994). The evidence to characterize the function and the limits of its effectiveness stems from several lines of research.

#### *1.1.1. Fast*

Whether speech occurs in the clear or in noise, it is quickly resolved perceptually if it is resolved at all. Classic studies of the persistence of the auditory trace of speech indicate such fast resolution, for they show that discrimination based on an auditory form of speech becomes poor very rapidly. Before the sensory trace fades, the auditory effects of speech are resolved into a coherent perceptual stream. The estimates of the rate of decay vary, though we can be certain that little of the raw auditory impression of speech is available after 100 ms (Elliott, 1962); and, none after 400 ms (Howell & Darwin, 1977; Pisoni & Tash, 1974). For the perceiver, the perishable auditory form creates an urgent limit on integration of the diverse constituents of speech; auditory properties available to perception are simply lost if integration is delayed. For a theorist, the evident long-term adaptive flexibility exhibited in natural perception cannot be attributed to unelaborated representations of the auditory features of speech without denying this basic psychoacoustic limit (see Grossberg, 2003). In contrast to the natural perceiver, urgency does not constrain artefactual recognizers. The schemes that they employ inherently surpass the physiological characteristics of an auditory system. They can sample and hold acoustic representations of speech analogous to the initial auditory sensory forms; indeed, they can hold them as long as electricity powers the memory (Klatt, 1989). Such superhuman systems have had wide theoretical influence despite indifference to the critical first step of urgent perceptual organization (Picheny, 2003).

#### *1.1.2. Unlearned*

Evidence that perceptual organization of speech is unlearned derives from studies of 14 week old infants, who integrated acoustic elements of speech composed through synthesis to be both spectrally and spatially disparate (Eimas & Miller, 1992; cf. Hollich, Newman, & Jusczyk, 2005). Listeners at this young age are hardly aware of linguistic properties in the speech they apprehend, and the perceptual coherence of the diverse constituents can be attributed to precocious sensitivity to vocalization independent of phonetic impressions, and well in advance of linguistic sensitivity. If experience plays a bootstrapping role in perceptual organization during the first three months of life, this is unlikely to entail arduous tutelage, nor sleep learning via exposure to adults whispering in the nursery.

#### *1.1.3. Keyed to complex patterns of sensory variation*

The amplitude peaks and valleys in the spectrum of speech are natural resonances of the column of air enclosed within the anatomy of the upper airway. These resonances, or

formants, are set ringing by the regular pulsing of the larynx, which produces harmonic excitation; or, by the production and release of air pressure behind an approximation or occlusion, as in the case of stop consonants; or, by sustained turbulence, as in the case of frication and aspiration. Acoustic changes in the spectrum are nonuniform across the formants. Specifically, the independent control of the articulators that produces formant frequency variation causes uncorrelated differences across the formants in the extent, rise and fall and temporal relation of frequency and amplitude change. Equal change in the first, second, third, nasal and fricative formants is uncharacteristic of vocal sound production, and aggregation of the sensory correlates of speech in perceptual organization occurs without evident reliance on similarity of change across the resonances. In some acoustic transforms of speech spectra, the frequency variation of the resonances is obscured without loss of perceptual coherence. In one version aiming to model the diminished frequency resolution imposed by an electrocochlear prosthesis (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) the coarse shape of the short-term spectrum envelope was represented in the power of 3 or 4 noise bands, each spanning a large portion of the frequencies of speech. Over time, asynchronous amplitude variation across the noise bands creates a derivative of speech without harmonic excitation and broadband formants, yet the frequency contours of individual resonances are absent. The effectiveness of such variants of speech spectra exposes the basis for the perceptual organization of speech, which lies in detecting a sensory pattern that coincides with phonologically governed articulation. Although these remain to be characterized formally, to a first approximation it is clear that the patterns of sensory variation are complex.

#### *1.1.4. Tolerance of anomalous auditory quality*

Perceptual organization of speech is tolerant of anomalous auditory quality, as research with sinewave replicas of utterances has shown (Remez, Rubin, Pisoni, & Carrell, 1981). In this synthetic acoustic signal, the natural products of vocalization are eliminated by imposing the pattern of a speech spectrum on elements that are not vocal in origin. Precisely, three or four pure tones are set to vary in frequency and amplitude in the pattern of the estimated formant peaks of a speech sample. A fourth tone is intermittently used to replicate fricative formants, brief bursts or nasal murmurs. The integration of the tones to compose an intelligible utterance occurs despite the persistence of the weird quality of a sinewave voice, evidence that neither natural acoustic correlates of speech nor auditory impressions of a legitimate voice are required for perceptual organization to occur. Studies with chimerical signals provide independent corroboration that the perceptual organization of speech is indifferent to the specific acoustic constituents of a signal and to the nonvocal auditory quality that can result (Smith, Delgutte, & Oxenham, 2002). To create an acoustic chimera, a coarse grain representation of the spectrum envelope of speech is excited with an arbitrarily chosen source. The result is a composite exhibiting the influence of each aspect, the spoken utterance and the arbitrary source. Like tone analogs of speech, a chimera is intelligible linguistically, evidence that its constituents are grouped to compose a signal fit to analyze as speech. Phenomenally, it retains the quality of the excitation, whether noisy, or harmonic, or, indeed, multiple, as in the instance shown in Figure 1, for which the excitation was provided from an acoustic

sample of a musical ensemble. In each of these critical cases, the perceptual coherence survived the inventory of arbitrary and nonvocal short-term properties by tracking the time-varying properties, which derived from speech even when the acoustic elements did not. This series of findings eliminates as implausible any characterization of perception warranting meticulous attention to elemental speech sounds or their correlated qualitative effects. Instead, they make evident the causal properties of a spectrotemporal pattern superordinate to the momentary constituents.

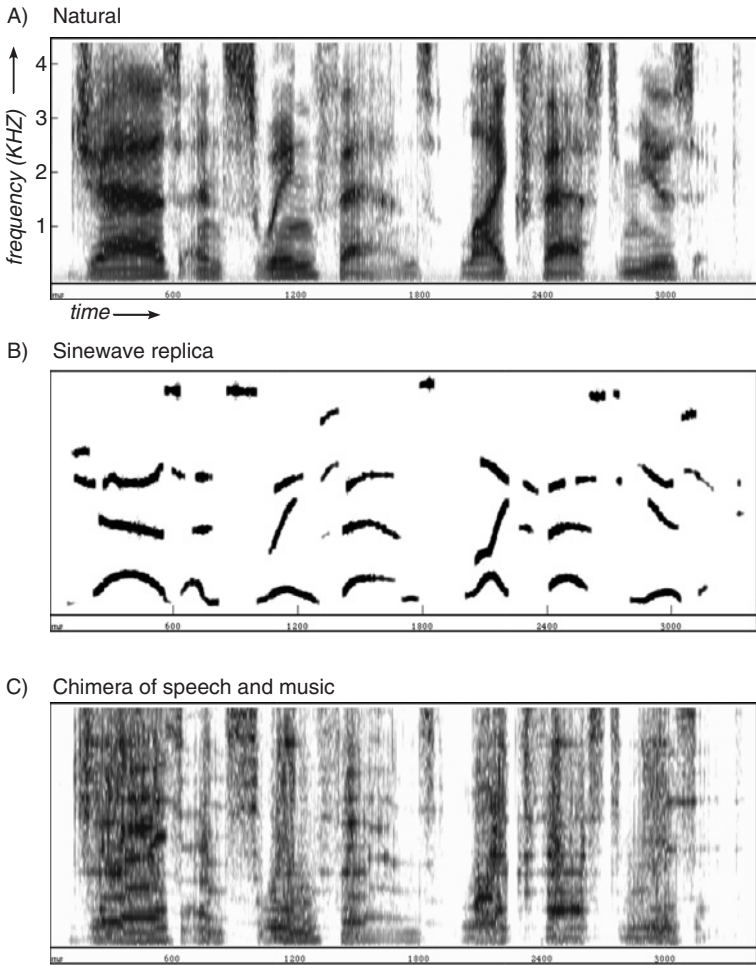


Figure 1. Inventory of acoustic constituents of three kinds of intelligible sentence: (A) natural speech; (B) synthetic replica composed of time-varying sinusoids; and (C) acoustic chimera made by exciting the changing spectrum envelope of speech with the fine structure of a nonspeech sample recorded by Count Basie in 1939.

### 1.1.5. *Nonsymbolic*

A study using patterned sinewave lures revealed that perceptual organization is nonsymbolic in its effects. That is to say, perceptual organization occurs by sensitivity to speechlike variation, and is distinct from the analytical finesse that creates an impression of linguistic form (Remez, 2001). In this test, the components of a sinewave sentence were arrayed dichotically, separating the tone analog of the first, third and fourth formants in one ear from the analog of the second formant in the other. The challenge to organization was to resolve the coherent variation among the tones composing the sentence despite the spatial dislocation of its constituents. After establishing that listeners tolerate spatial dissimilarity in amalgamating the tones perceptually, a variety of lures was introduced in the frequency band of the second formant in the ear opposite the true tone analog of the sentence, which also contained the tone analogs of the first, third and fourth formants. In this kind of presentation, perceptual organization is challenged to resist the lure presented in the same ear as the first, third and fourth tone, and to appropriate the second tone that completes the sentence. Some lures were easy to resist; those that were constant in frequency or those that alternated brief constant frequency tone segments did not harm the organization of dichotically arrayed components. Other lures were far more difficult to withstand: a gradient of speechlike variation was created by straining (or, more commonly, squashing) the frequency variation of the lure to vary from speechlike to constant frequency at the average frequency of the second formant; the speechlike variant was actually a temporally reversed second formant analog. The lure could not complete the sentence because its variation was never coherent with the first, third and fourth tone. Nor did it evoke impressions of linguistic form; the lure in the clear sounded like a warbling pitch pattern. Nonetheless, it interfered with organization in proportion to its frequency variation, the most when it varied with the pattern of a second formant; less when it varied in frequency at the pace and pattern of a second formant but over a reduced frequency range; and not at all when it was squashed to a constant frequency. In other words, the propensity of the lure to interfere with perceptual organization depended on the speechlike properties of its frequency variation alone, because this single time-varying tone did not evoke phonetic impressions that competed with those of the sentence.

### 1.1.6. *Requires attention*

A subjective impression of inevitability often accompanies the perception of utterances, but this belies the actuality; the perceptual organization of speech requires attention, whether elicited or exerted. This is seen, again, in studies of sinewave replicas in which the auditory quality is so little like speech listeners are unlikely to organize the tone complexes spontaneously (Remez et al., 1981; Remez, Pardo, Piorkowski, & Rubin, 2001). When a sinewave replica of an utterance is heard simply as a set of contrapuntal tones, none of the auditory qualities is vocal, hence nothing about the experience compels the perception of linguistic properties. Physiological measures were consistent with the hypothesis that no covert aggregation of the tones occurs if the perception of auditory form takes place without apprehension of phonetic attributes (Liebenthal, Binder, Piorkowski, & Remez, 2003). However, a listener who is informed that the tones

compose a kind of synthetic speech is readily capable of transcribing a sinewave sentence on this instruction alone; no training, extensive exposure or special hints are needed. In this condition, the aim of hearing the tones phonetically permitted attention to the coherence, albeit abstract, of the speechlike variation in the tones. Moreover, only the sinewave patterns derived from natural utterances are amenable to organization by virtue of the exercise of attention. An arbitrary or incomplete physical spectrum does not evoke an impression of phonetic attributes simply because a perceiver intends to resolve a speech stream. Because natural and much high-quality synthetic speech elicits phonetic attention by virtue of intrinsically vocal auditory qualities, the role of attention in speech is easily overlooked. More generally, this finding of a contingency of organization on attention in the case of speech anticipated the claim that auditory perceptual organization in the main depends on attention (Carlyon, Cusack, Foxton, & Robertson, 2001). It also falsifies the claim that speech perception is accomplished by a modular faculty, because the contingency of perceptual organization on attention contradicts the premise of autonomous and mandatory action immune to influence by belief.

## 1.2. Audiovisual Perceptual Organization

The classic formulation of perceptual organization took the cocktail party as the critical setting, and much of the ensuing research examined the proficiency with which a perceiver pulls the speech of a talker of interest from a lively acoustic background. Although this conceptualization has been productive, a different slant is needed to describe a listener who can also see the talker. It has been well established that in this situation a perceiver treats speech as a multimodal event, sampling visually and auditorily (Sumbly & Pollack, 1954). In multimodal speech perception, the formal characterization of finding and following a speech stream remains much the same as the auditory instance, with a twist. Rival conceptualizations have characterized multimodal perceptual organization in parallel streams converging at the end or, as a single multimodal stream in which visible and audible features interact continually. To caricature the two perspectives, the first pictures the perceiver as both a blind listener and a deaf viewer huddling within a single skin, resigned to negotiate any discrepancy between utterances that each determines after perception concludes. The second perspective conceives of the perceiver as an auditory-visual synesthete in whom visible and audible ingredients blend so thoroughly from the start that no residue remains to distinguish the sensory core of the phonetic forms. Whichever conceptualization comes closer to the truth, the visual resolution of speech poses the familiar challenge to organization: the physical effects are regular albeit unpredictable, and none of the effects in detail is unique to speech. This is so whether the level of description is a stream of light reflected from the surfaces of the face, a 2 1/2-d sketch of an as yet unresolved face in a visual scene, or a description of the face as a familiar kind of object in motion.

### 1.2.1. Intersensory combination

In considering the problem of multimodal perception, it is also natural to speculate about the grain of analysis at which intersensory combination occurs (Rosenblum,

2005; Lachs & Pisoni, 2004). Principally, this is a puzzle to solve only if the second alternative conceptualization, of the amalgamated multimodal stream, proves to be true. If speech perception occurs separately in visual and auditory modalities, then perceptual organization proceeds in parallel for visible and audible samples, and any interaction between modalities occurs after perceptual analysis has resolved the phonemic form. Because the phonemic properties are set by their contrastive linguistic function, and not by the specific transient sensory or motor forms of their expression, the dimensions of visually perceived speech intrinsically match those of auditorily perceived speech. Under this condition, the scientific puzzle of understanding the alignment of visible and audible streams is obviated.

The opposite conceptualization, in which intersensory blending occurs in organization, poses a puzzle, for there is no obvious dimension common to vision and hearing. The sensory qualities of these two modalities are largely incommensurate – hue, brightness and saturation do not form tight analogies to pitch, loudness and timbre – and it should be evident that the acoustic transparency permitting a listener to hear the changes originating in the action of the glottis and tongue body have no counterpart in a visible face, in which the larynx and all but the lips and the tip of the tongue are out of sight. Some research proposes the existence of a common intersensory metric, an intermediary permitting the visual and auditory streams to blend in a form exclusive to neither sense (Massaro & Stork, 1998). Variants of this proposal cast auditory sensation – pitch, loudness and timbre – as the common metric into which visual form is also cast for a spoken event (Kuhl, 1991), and a kind of shallow representation of visual and auditory primitives in articulatory parameters (Rosenblum & Gordon, 2001), about which there is much more to say when we turn to phonetic analysis.

Evidence favoring each conceptualization of intersensory relation exists in the technical literature, chiefly in studies of audiovisual merger. That is, synthesis and digital editing of video and audio components have been used to create phonemically discrepant visible and audible presentations, with which to determine the nature of multisensory combination. In the original use of this method (McGurk & McDonald, 1976), an audio [ba] and a video [ga] were resolved as a fusion, [da]. In tests of this kind, it is possible to fix the identifiability of auditory and visual components independent of tests of their combined effect, and such findings are subsumed well within a parallel model of perceptual organization in which concurrent perceptual states are reconciled if the perceiver organizes them as bound to the same talker. Alternatively, some audiovisual phenomena are simply not well described by parallel and segregated organization in each modality.

In one of these intriguing cases, a video of a face was presented with an amplified electroglottograph signal correlated with the pulsing of the larynx of the depicted talker (Rosen, Fourcin, & Moore, 1981). The appearance of the face was unexceptional; the electroglottograph sounded like an intermittent buzz changing in pitch in the range of the voice. Some of the syllables and words could be resolved phonetically by watching the face, and the audible buzz evoked no impressions of words at all. Overall, the

conditions for unimodal visual speech perception were barely met, and were not met at all for unimodal auditory speech perception. In this circumstance, multimodal perception should be poor in as much as the cumulative effect of poor visual and no auditory perception remains poor. Instead, the combination was fine, arguably reflecting the effectiveness of auditory and visual streams in combination when separately neither stream was adequate. This finding among many others offers evidence of a common dimensionality for viewed and heard speech preliminary to analysis.

### 1.2.2. *Mismatch tolerance*

Among the best clues to the nature of multimodal perceptual organization are the results of studies of the tolerance of spatial and temporal discrepancy across the modalities. In one notion, vision and hearing supply discrepant but complementary samples of speech (Bernstein, Auer, & Moore, 2004), and this simplification describes both audiovisual presentations in which visible and audible patterns coincide and cases of intersensory competition. The organization of fine grain discrepancy between viewed and heard speech scales up to coarse grain discrepancy, and this functional similarity over scale variation is surprising. At the finest grain, auditory and visual streams are mismatched simply because of the disparity in the aspects of the physical acts of articulation that each provides, and not only because the primitives of auditory and visual sensation differ. Indeed, the enterprise of multimodal research rests on findings that fine grain discrepancies introduced by a scientist's method are resolved in perception, often without eliciting an impression of disparity in the seen and the heard speech. And, at the coarser grain as well as the finer, perceptual integration of discrepant sensory samples is robust.

In several studies, perceptual integration survived spatial displacement of audio and video sufficient to notice (Bertelson, Vroomen, & de Gelder, 1997); and, temporal misalignment of audio and video sufficient to notice (Bertelson et al., 1997; Munhall, Gribble, Sacco, & Ward, 1996). To be more precise, the merging of phonetic features used to index the perceptual integration of vision and hearing persisted under conditions of sizeable spatial and temporal divergence. Such findings can leave the researcher without a convenient explanation because the theory of first resort fails to apply. Specifically, the very tolerance of mismatch blocks the psychologist's automatic and tiresome appeal to similarity as the engine of integration; the integrated streams are dissimilar, displaced and lagged. And, the conditions created within the audiovisual display introduce discrepancies at a scale that surpasses ordinary experience by an order of magnitude. Appeals to likelihood can seem clichéd in psychological explanation, but this procrustean tactic must fail in these instances. The relative divergence of the integrated streams is just unfamiliar.

### 1.2.3. *A unimodal and multimodal contour*

How, then, does a perceiver apprehend the disparate sensory samples of speech as a coherent progressive event? When organization is veridical, the auditory or visual effects are grouped despite dissimilarity and discontinuity of the sensory constituents of a

perceptual stream. The familiar principles of perceptual organization deriving from Gestalt laws of figural organization (Bregman, 1990) cannot be responsible for unimodal organization, for they invoke one or another form of similarity among the sensory constituents. If a role for this conventional account seems unlikely to suffice in unimodal organization, it is utterly implausible for explaining the cases of multimodal organization in which some form of binding appears to occur intermodally in advance of analysis. Although the discussions in the technical literature generally portray binding as a process of sorting analyzed features into bundles coextensive with objects, it appears as though the urgency of auditory perceptual organization compelled by the fast-fading sensory trace imposes a different order. Instead, binding of the sensory constituents of the spoken source must occur before analysis, and perhaps this is the cause of the condition that the perceptual organization of speech requires attention. It is tempting to speculate that there is a single set of organizational functions that applies regardless of the assortment of samples arrayed across the modalities, and some studies of neural metabolism correlated with perceptual organization (Liebenthal et al., 2003) are consistent with this view – but evidence of its existence holds far less value than evidence of its characteristics would.

### 1.3. Perceptual Analysis

A perceiver who resolves a stream of speech in a raucous or tranquil scene might also be able to resolve its linguistic form. These facets of perception are contingent. Certainly, the circumstance in which a listener knows that someone is speaking but cannot make out the words is familiar to us, although the inverse – linguistic impressions of a spoken event in the absence of an impression of someone speaking – might merit a thorough reappraisal of mental status. The perceptual resolution of linguistic form has been a topic within the technical study of speech for more than seventy years, and the longevity of this concern is due to the intriguing complexity of this type of sensitivity. Although it has taken a variety of guises, in each the central challenge has been to understand the perceptual ability to apprehend the expression of a small number of linguistic forms under conditions that vary without end.

Long ago, research on the perceptual analysis of speech adopted a focus on the ultimate constituents of language. That is, the linguistic properties that speech expresses are componential, and the components are hierarchically nested. Utterances in the form of sentences are composed of clauses, within which phrases are nested; phrases comprise words, words are composed of syllables and each syllable can be a series of phoneme segments. Phonemes are grouped by distinctive features, that is, by virtue of the coincidence of disjunctive attributes that, together, constitute a system of contrasts across the segmental inventory of a language.

To researchers of the first generation of psycholinguists, the componential nature of linguistic structure was theoretically significant, though the focus on ultimate constituents in speech perception research was also practical (Miller, 1965). It is not sensible to focus on sentences as irreducible objects of perception— there is an infinite



number of them, and of phrases, too. Although languages differ in the number of words that they sustain at any moment in history, the set of these is also large. To consider a specific instance, in English, a language often studied in psycholinguistics, words derive from Germanic and from Romance heritages, and for this reason English is said to incorporate more words than is typical of other languages with calmer history. A focus on individuals of a specific group can restrict the study to vocabulary in regular spoken use – in contrast to the far larger recognizable vocabulary – which imposes an inventory of roughly 15,000 items (Miller, 1951). If this is a saving from the infinity of sentences and a large lexical stock, even greater economy is achieved by considering that whatever the word, in English it is composed from a supply of three dozen or so phonemes expressing perhaps a dozen and a half contrast features. Taxonomies of phonemes and the features on which the classes are sorted can become controversial from time to time, depending on the rise or fall in value of one or another kind of evidence. Even with such disputes, there has been good agreement that the perception of speech entails the perceptual resolution of elementary linguistic attributes available in a brief spoken sample; larger structures of linguistic form are produced cognitively by aggregating the elementary constituents provided by speech perception. We defer a discussion right now of the consequences of the phonetic expression of phonemic contrasts, but not for long.

Setting a perceptual focus that is linguistic, segmental and contrastive defines the products of perception, although consensus about the effects has not tempered the disagreements about the causes of perception. This dispute among perspectives concerns the kind of perceptual analysis yielding the linguistic objects. Proponents have divided on its essential nature. Either the perception of speech depends on auditory sensitivity and categorization, or on articulation, or on linguistic function. Each of the proposals is old, and the stalemate is apparently perpetual. We will offer a recommendation, but first we expose some of the technical details.

### 1.3.1. *A general auditory account*

The roots of the auditory approach run deep. Among the earliest reports in experimental psychology are studies of the likeness of simple whistles and buzzes to speech sounds (Kohler, 1910; Modell & Rich, 1915). Although the correlations were only rough, they licensed the claim that *vocality* is a primitive auditory sensory quality. The argument held that because vocal impressions are elicited by simple acoustic attributes they are fundamental in human sensory experience; therefore, a talker's ability to evoke a listener's phonemic states depends on producing sounds that hit auditory targets given subjectively and intrinsically. There are more technically sophisticated versions of this antique claim at large today (Kuhl, 1991), but the germ of the idea is similar. Indeed, such findings are perennially welcome in psychology due to a resilient eagerness for sensory reduction of perceptual impressions of the structures and motions of the world. The draw of this explanation is that it permits a description of perception to attribute an incidental role to the objects and events that ultimately cause sensory states: All of the explanatory action pertains to the sensory pathway and neural centers of associative learning. Indeed, it has become commonplace recently for the justification to invoke a perceiver's ability

to learn the statistical characteristics of the distribution of sensory states with which phoneme contrasts allegedly coincide. This premise invokes a hypothetical norm in its attempt to accommodate the variability in the acoustic form of each phoneme due to the variety of talkers, rates of speech, and attitudes expressed concurrent to language production, each of which precludes an acoustically uniform expression of a phoneme across different occasions. In one expression of this idea (Diehl, Kluender, Walsh, & Parker, 1991), the auditory system is viewed as a nonlinear conduit of the acoustic effects of speech in which contrast is created by means of enhancement of some auditory elements relative to others. Admittedly, adherence to a general auditory perspective is only weakly justified by psychoacoustics or auditory physiology (Diehl, Lotto, & Holt, 2004).

The perspective on speech perception offered in a general auditory approach has a goal, to pursue a model of the phonemically interested listener as a trainable ear and little else. In a recent review, Diehl et al. (2004) argued that the explanatory detail presently accrued under this rubric is too thin to permit a falsifying test, but this reservation seems unduly gloomy. Even if precise predictions of experimental findings are not readily produced from the principles underlying the approach, it is sensible to ask if the premises of the model attach importance to false assertions. Specifically, if the ambition of the model is not mistaken, its allure is surely diminished by two well established properties of speech perception: (1) the fleeting nature of auditory forms; and, (2) the irrelevance of auditory norms. First, in this class of accounts, perception is based on the varying sensory correlates of speech sounds, and a listener's personal history of experience with /d/, for instance, is encoded to generate a long-term probability distribution in which more and less typical auditory manifestations of /d/ are calibrated. The success of a listener in recognizing instances of this phoneme would necessarily depend on the likelihood that an as yet unidentified sensory form can be assimilated to a longstanding likely auditory representation of /d/ among other segments in the language. But, classic psychoacoustic research revealed that the auditory properties of speech are exceedingly fragile, and are difficult to protect for even a quarter of a second (for instance, Howell & Darwin, 1977). This limit must be a mild embarrassment, at least, to a conceptualization relying on the durability of raw auditory impressions of speech. Although such representations are reasonably chosen for instrumental applications such as speech-to-text systems, these are constrained by circuit design and not by physiology (Klatt, 1989). To survive in a listener's memory, short-lived auditory properties acquire a different form, possibly in the dynamic dimensions of the sources that produced them (Hirsh, 1988), and when a listener remembers a sound, it is more likely that the recalled quality is generated rather than replayed from a faithful inscription in memory of the original auditory form.

A second problem for a general auditory account of speech perception is its reliance on auditory manifestations of the phoneme contrasts graded by likelihood. Even to entertain this premise, we must be credulous momentarily about the prior claim that unelaborated auditory forms of speech are retained well in memory; this suspension of criticism permits us to review the assertion that a spoken phoneme is identified by a normative assessment of its sensory form. In short, the robustness of intelligibility over widely varying natural conditions of acoustic masking and distortion show clearly that

neither goodness nor typicality in auditory quality is requisite for speech perception. Indeed, intelligible sentences are perceived from patterns dissimilar to speech in acoustic detail and in auditory effect (Remez et al., 1981; Shannon et al., 1995; Smith et al., 2002). But, what is the shape of a distribution of the auditory attributes of a phoneme?

To be truthful, no one knows. There is a single study of actual incidence, of the exposure of a single infant to speech produced by one adult (van de Weijer, 1997). This means that claims about sensitivity reciprocating the distributions of the acoustic or auditory forms of speech are hopeful, and without empirical foundation (Saffran, Aslin, & Newport, 1998); at least, no claim is grounded empirically yet. But, it is not difficult to recognize the implausibility of the claim that auditory typicality determines the perception of phonemes. The typical auditory forms of speech must be those sensory states evoked by exposure to the acoustic products of vocalization. After all, an overwhelming majority of instances must be those in which a listener perceives speech because a talker spoke. These days, the pervasiveness of the experience of speech over the telephone also contributes to normative distributions, and so does speech produced by talking toys and gadgets. Overall, the probability distribution must represent this kind of typical experience composed chiefly of acoustic vocal products with minimally distorted variants at the improbable ends of the distribution.

In fact, listeners are evidently not fussy about the acoustic constituents or the auditory qualities of intelligible signals. Neither natural broadband resonances nor harmonic excitation nor aperiodic bursts and frictions nor any specific set of acoustic correlates of a phoneme is required for perception (see Figure 1). Instead, a listener perceives speech as if the commitment to the particular sensory realization of the linguistic contrasts is flexible. This readiness to find functional contrasts in the least expected acoustic or auditory form opposes the fixity of an auditory norming rationale. Indeed, such acoustic norms – some without auditory warping – form the basis of speech-to-text devices often aimed at the typical expressions of just a single individual (Picheny, 2003); even so, we are still typing.

Before turning to consider an account of perception grounded in articulation, it is useful to note that there are important questions about auditory function in speech that do not depend on the claim that phoneme categories coalesce out of auditory form. At the most elementary, the acoustic correlates of each linguistic contrast are multiple: the speech stream itself is a composite of dissimilar acoustic elements. Attention to the auditory quality of constituents of a speech stream – an aperiodic burst, a second formant frequency transition, a noisy hiss – can occur concurrently with attention to the linguistic properties – an unvoiced fricative of coronal articulatory place. This kind of *bistable* perception in which attention can hold auditory form or its superordinate, or both, is not well understood outside of musical contexts. Moreover, if qualitative attributes of speech are retained in a durable form, the dimensionality of such knowledge is not well explored (Hirsh, 1988). At the largest grain, the flexibility of the standards for perceiving the linguistic elements of speech is well evident, yet the function by which a perceiver resolves linguistic properties in specific instances, especially those evoking novel auditory form, remains a tough puzzle.

### 1.3.2. *An articulation-based account*

Modern linguistic description took shape with phoneme contrasts already described in the dimensions of articulation. The technology required to portray acoustic properties did not exist, and in the resort to articulatory dimensions to describe the sounds of speech, Joos (1948) says linguists made a virtue of necessity. However, this practice was unsatisfactory even as articulatory description, largely because the method presumed anatomical and functional states of articulators without direct evidence. For instance, the classical notions of articulatory contrasts in vowel *height* and *advancement* were designated by intuition, not by observation, and ultimately proved to be inaccurate portraits of the tongue shape and motion discovered in x-ray fluoroscopy, electromyography and magnetic resonance imaging (Honda, 1996).

When methods for direct measurement of sound became available to supplement impressionistic descriptions, it had a paradoxical effect on the restlessness with old fashioned articulatory description. As the basic properties of speech acoustics were described technically, a problem emerged for proponents of acoustic description; indeed, the conceptualization of articulation was challenged as well. Research on production and perception alike failed to find counterparts to the theoretical description of phonemes in articulatory, acoustic or auditory components. Each perspective in its own way had presupposed that speech was a semaphore, with every phoneme a kind of vocal act or pose, or every segment a kind of acoustic display. Instead, whether construed as acts of articulation, their physical acoustic products or their psychoacoustic effects, an apparent lack of invariance was evident in the correspondence of the linguistically contrastive phoneme segments and their expressive manifestations. In each domain, the relation of a phoneme to its articulatory and acoustic correlates proved to be one-to-many.

Of course, the mere existence of variety among the physical or physiological correlates of a linguistic component is not troublesome to perceptual explanation. If the articulatory, acoustic or auditory tokens of different phonemes correspond uniquely to types, the lack of invariant form is insignificant because the correlates of one type are not shared with any other. The critical finding about the relation of phoneme to correlate was the nonexclusive relation between type and token. One of the clearest instances is the /pi-/ka-/pu/ phenomenon (Lieberman, Delattre, & Cooper, 1952) in which a single acoustic element evokes an impression of a *labial* consonant and a *palatal* consonant depending solely on the vowel with which it is presented.

A key explanatory innovation occurred in response to such findings. A new sense of the idea of *coarticulation* was created to describe the relation of production and perception; a history of coarticulation in phonetic linguistics is offered by Kühnert and Nolan (1999). At the heart of this breakthrough was the inspiration that descriptively segmental phonemes are encoded in articulation (Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). That is to say, a perceived phoneme sequence is restructured rather than produced as a simple sequential cipher or articulation alphabet. The encoding occurs because the vocal articulators are intrinsically separable into controllable parts, and the expression of a sequence of phonemes is thereby reassembled as an imbricated pattern of

constituent acts that unfold concurrently and asynchronously. This approach explained well the inexhaustible variety of articulatory and acoustic correlates of each phoneme, or, rather, the lack of a consistent physical manifestation of a phoneme, because whichever segments preceded and followed it shaped its articulation by contributing to the encoding; and, no phoneme is ever expressed in isolation of coarticulatory influence. Liberman et al. explain that such recoding achieves high rates of segmental transmission with sluggish anatomy. The cost is to obscure the relation between an intended or perceived phoneme and its articulatory and acoustic form. Accordingly, speech gives phonetically encoded expression to an intended if abstract phoneme series.

From this premise, a characterization of perception follows as directly as night follows day. If the acoustic speech stream is an encoded version of phonemes, due to the articulatory restructuring of an abstract segmental series, then an inverse operation required to apprehend the segmental series obliges a perceiver to reciprocate the motor encoding in some fashion. A variety of specific technical hypotheses about this kind of perception was ventured in different versions of the *motor theory*, including reliance on learned articulatory correlates of the auditory forms of speech; covert efferent mimesis; and, imagined surrogates of proprioception that accompanied a talker's speech.

Challenged by evidence, the motor theory looked terrific at a distance, from the perspective of neuropsychology or studies of human evolution (Galantucci, Fowler, & Turvey, in press). At close range, the disconfirming proofs of its technical claims emerged steadily from detailed research on the relation between perception and production. Crucially, studies of extremely young infants showed that perceptual sensitivity develops in advance of articulation, and is not a consequence of it (Jusczyk, 1997). In adults, the invariant characteristics presupposed of the articulation of individual phonemes was falsified in studies of articulatory motion and electromyography (MacNeilage, 1970). This is an enormously intriguing literature impossible to gloss. Yet, acknowledging exceptions, the fair preponderance of evidence showed that every phoneme takes many anatomical forms, and invariance in the correspondence of phoneme to motor expression was found neither in an aggregate of  $\alpha$ -efferent activity, nor in the precise motion or configuration of articulators, nor in the shapes of the vocal tract achieved by articulation. In a revision of the motor theory proposed to answer research that it had motivated, perception was held to resolve the invariant phonemic intentions of a talker rather than the acts of articulation, varying without limit, as they are executed (Liberman, & Mattingly, 1985). This version represents spoken communication as a transaction composed of deeply encoded phonemic intentions, aligning the revised motor theory with the symbolic emphasis of a linguistic view of speech perception.

A pair of conjoined accounts of more recent vintage aims to span the gulf between intention and action while retaining the emphasis on production of the motor theory: *articulatory phonology* and *direct realism* (Goldstein & Fowler, 2003). Articulatory phonology offers a description of linguistic contrast set in abstract articulatory primitives, and direct realism describes a perceiver's sensitivity to articulatory contrasts by attention to the visible, audible and palpable effects of speech. The crucial contribution of this

proposal is a representation of lexical contrasts in a repertoire of *gestures*, not of phonemic segments. Building on the characterization of articulation given by Liberman et al. (1967), a contrastive gesture is designated as: (1) a movement of a particular set of articulators; (2) toward a location in the vocal tract where a constriction occurs; (3) with a specific degree of constriction; and, (4) occurring in a characteristic dynamic manner. In this perspective, a word is indexed by a gestural score describing its production as the coupled asynchronous action of lips, tongue tip, tongue body, velum and larynx. Such gestural components are understood as quasi-independent actions of vocal articulators. The pattern with which gestures impose and release constrictions creates the contrasts customarily described in a segmental phoneme series. The traditional separation of phonemic contrast and phonetic expression theoretically collapses in this account into an equivalence between linguistic properties and vocal acts. With respect to the principle at the core of the motor theory, this asserted equivalence of linguistic contrast and manifest articulation denies the encoding that supplied the articulatory character of the inverse function purportedly applied by a perceiver to a speech stream. In complementary function, the account describes the perceiver distinguishing words in the same gestural components that the talker employs to create speech. The phonemic properties of speech are apprehended perceptually without decoding them, according to this argument, because the acoustic pattern and its sensory effects are transparent to the articulatory components that index spoken words across the lexicon (see Figure 2).

Part of the attraction of this theoretical gambit is the potential of an articulatory phonology to explain allophonic variation in a natural way. That is, the production of a

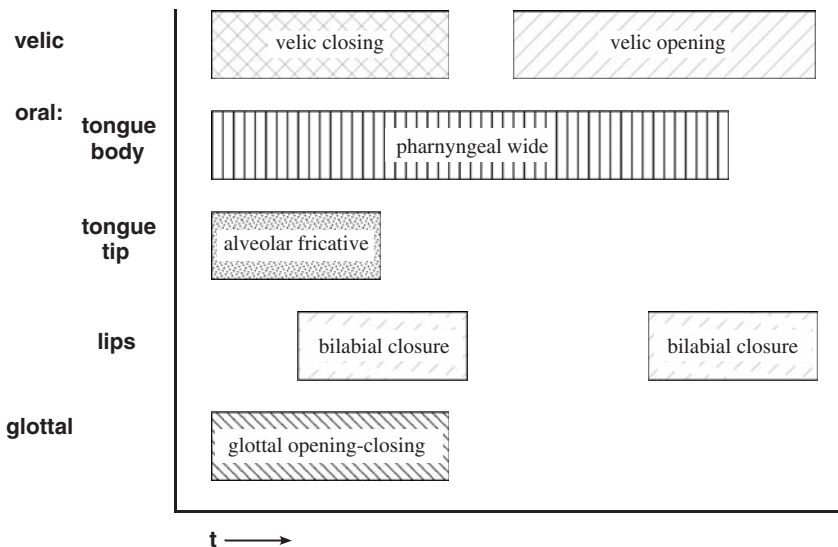


Figure 2. A gestural score representing the coupled actions of quasi-independent vocal articulators in the production of the word SPAM. (Browman & Goldstein, 1991, p. 318).

canonical phoneme sequence actually varies in exact phonetic, or expressed, detail. The word SECURITY, for example, is produced as these variants, among others: [sək<sup>h</sup>ju.ɹɪ<sup>h</sup>i], [sɪk<sup>h</sup>ju.ɹɪ] and [sk<sup>h</sup>ʒ.ɹɪ]. Under an articulatory phonology, many variants are potentially rationalized as consequences of minimally different task dynamics of the same gestural components given in a lexical representation. Variation attributable to differences in speech-rate, reductions, lenitions and apparent deletions are likewise described as natural variants of the same gestural form, and this sameness remains available to perception, in principle. A listener who resolves the gestural components in a speech stream might notice but pay little attention to the effects of slight phase differences in the expression of constituents of an intended contrast. Even in casual speech, in which the canonical forms of words can be compromised by the expressive aims of a talker, articulatory phonology promises to explain the variation without representing the phonetic form of the articulation differently from the phonemic form distinguishing a word from all others. Admittedly, in some synchronic and diachronic cases, it seems that talkers do express different contrasts than the lexicon employs, yet the different representations warranted by these facts can nonetheless be described by postulating no more than minimal changes in the components of a gestural score. Of course, there are some phonological phenomena in some languages that defy simple characterization – Nature provides her own exceptions – and it is not clear how these will be resolvable to general principle in relating the phonemic and expressed forms (Browman & Goldstein, 1991). But, the perceptual claims of this account are readily evaluated.

Two critical axioms are assumed in the perceptual account given by articulatory phonology and direct realism, and if they are not exactly false, they are less true than the account demands. The first is an asserted *isomorphism* between the components used in language to create contrast and the components of spoken acts; they are designated as a single set of gestures. The second is a state of *parity* such that talker and listener match; the expressed forms and the perceived forms are claimed to be the same. Of course, these axioms are related. People who speak the same language use the same canonical contrasts. If they express them differently, or if a talker nonaccidentally expresses the same form in gestural variants on different occasions, then the relation between canonical and expressed forms can be regulated, adjusted or reshaped; phonemic and phonetic form are not identical in this circumstance. Instead, some of the degrees of freedom in articulation would be reserved for expression beyond those that are committed to the canonical form of the word. If articulation varies with a talker's communicative aims, then canonical and expressed forms do not match, and parity must be achieved, not simply fulfilled.

Is articulation isomorphic to phonemic contrast? A recent set of examples described by Hawkins (2003) shows the extent of the mismatch in ordinary circumstances. The graded expressions of the utterance, "I do not know," that she discusses range widely: [ai du nɔt nɔu], [ai dɔnt nɔu], [dɔnɔ], and [ʒʒʒ]. If this shows the varieties of phonetic form under different expressive demands, consider an example of the same expressive demand imposed on different individuals. Three adult talkers from Southern New England contributed to a corpus of sentence-list reading (Remez, Fellowes, & Rubin, 1997), each producing the word DROWNING in consistent yet distinct form [d.ɹɔʊnɪŋ], [dʒ.ɹɔʊnɪŋ] and

[ds.ɹʊnɪŋ]. Even a single individual in a stable pragmatic condition can generate such variation. In a recent observation of a plausibly commonplace incident, we witnessed a taunting child calling her father: [dæ:ri]... [diæ:ri]... [dædi]... [dæ:ri]... [dæ? di]... [dæ:ri]... etc. In each of these examples, it is difficult to maintain that phonetic form is well described as an obligatory isomorphic projection of canonically given gestural form. While these cases obviously express phonemically contrastive gestures under mutual coarticulatory influence, they also show the converging influence on phonetic form of long-term effects of dialect and idiolect, and the immediate effect of pragmatic pressure and opportunity. In the causes of departure from isomorphism, there is more to explain than the effects of coarticulation and rate variation.

Can a phonology indicate the aspects of the phonetic realization that stem from canonical lexical form and those that are controlled by consistent non-lexical aspects of expression? Surely, dialect and idiolect are components of language, and the phonetic means of marking contrasts between communities and among talkers within them shares the phonetic grain of production with the expression of lexical contrasts. Perhaps an ultimate model would include the influence of long-standing or momentary expressive aims on the articulatory parameters in the task dynamic organization of production, explaining the regular expressive variants of canonical lexical form that supersede straightforward coarticulatory and rate effects. But, the consistent lapses in isomorphism between canonical and expressed form deprives the perceptual account of its central claim: the transparency of lexical contrast sensed via public aspects of speech. In the absence of isomorphism, the perceiver is challenged to resolve the canonical gestures from a speech stream that marks dialect, idiolect, affect and attitude in phonetic form as well. Variation in phonetic form that disrupts simple isomorphism with phonemic form is often conceptualized as noise, the consequence of undeniable but unpredictable deflection from canonical expression, but this is pessimistic and mistaken (for instance, Pisoni, 1997). Perceivers are sensitive to dialect variation even if intuitive dialect geography is no more accurate than intuitive geography. Attributes of idiolect independent of vocal quality are evidently tracked and integrated perceptually, and are effective for identifying individuals by discerning their phonetic habits (Remez et al., 1997). The phonetic basis for perceiving affect and attitude are harder to pin down, empirically, but in tractable cases like smiling – a gesture of the lips – the phonetic effects are sensed readily (Tartter, 1980). In other words, a perceiver is often capable of resolving the aspects of phonetic expression that derive from its multiple causes, and a complete account of perception would aim to describe the facility with which we attend to the properties of articulated linguistic form that count lexically and the collateral resolution of those that are consistent and informative but not distinctive lexically.

The axiom of parity denotes sameness in language forms shared by a talker in composing an expression and by a listener in perceiving it. Here, the intended sense of parity applies only to the gestural components of language, and this is completely apt. After all, lexical parity is commonly breached: conversations are fine with people who say LIGHTNING BUG when we say FIREFLY, SACK for BAG, TRASH for GARBAGE, that is, under conditions in which the communicative function matches while the lexical form does not.



The claim of parity states that the forms of perception and production are the same, and the claim at some level of resolution cannot be false (Liberman, 1996). If the perceiver knows what the talker said well enough to repeat it, lexical and phonemic parity occurred, at least. But, because isomorphism is suspicious, expressed forms can be understood to differ from abstract phonemic forms. The axiom becomes harder to sustain in that case because phonemic parity can occur without phonetic parity; and, because phonetic parity is so unlikely, even in monozygotic twins reared in the same household (Johnson & Azara, 2000; Nolan & Oh, 1996; cf. Gedda, Bianchi, & Bianchi-Neroni, 1955).

Critical data on this topic indicate that perception is bistable, permitting attention to be drawn to superficial and canonical form concurrently. The study (Goldinger, 1998) used an original measure of perceptual resolution. An experimenter presented a recorded utterance for a subject to repeat immediately or after a brief imposed delay. Comparing the elicited speech samples to the eliciting sample showed that utterances produced immediately were more similar to the eliciting sample than were those produced after even a brief delay. Despite all, a similar utterance was far from a faithful replica of the model. This is expected, to be precise, for not even nightclub impressionists achieve their characterizations by exact replication of the speech of John Wayne and Cary Grant – and they rehearse. The difference in the two conditions of lag must be attributed to phonetic attributes inasmuch as the words were the same, hence, the contrastive phonemic properties were the same. With respect to the parity axiom, though, the result is troubling, because the finding of only rough similarity insinuates that if parity is fostered it is unattained; and, that the faint shadow of parity that actually is manifest lasts only a moment, and once the impulse toward parity subsides the default state of disparity returns. Moreover, studies of deliberate imitation show that an individual typically provides an erroneous imitation of a self-produced speech sample (Vallabha & Tuller, 2004). If parity does not occur in this limiting case, when would it?

A plausible description of these phenomena is possible without invoking a disposition to isomorphism and parity. In the moment when a spoken word is perceived, phonemic and phonetic forms are resolvably different from each other. The salient differences often include aspects of a talker's speech that differ greatly or minimally from a perceiver's own characteristic articulation. Speech initiated in this state can be nudged toward the form of the immediate phonetic model and away from the habitual expression of canonical form. At a greater delay, though, the vividness of the phonetic impression of the eliciting utterance has faded in its contrast with long established articulatory habits, and production is free from the adulterating pressure of a phonetic form distinct from the talker's intrinsic dynamic. It is as if talker and listener express lexical and phonemic parity by means of their phonetic differences. With sustained exposure to an individual talker, a perceiver is likely to form an impression of the talker's characteristic articulatory variation, sufficient to imagine speech produced in the voice and style of the talker, and perhaps to adopt the phonetic characteristics in a deliberate imitation (Johnson, Foley, & Leach, 1988). There is some evidence that such vicarious experience of the speech of familiar others can influence a talker's production in detail (Sancier & Fowler, 1997). But, the listener and talker need not match phonetically for any of this to occur. Indeed, in order for the phonetic similarity of two talkers to wax and wane, they cannot match.

The assertions of isomorphism and parity mask a significant aspect of the perception and the production of speech, namely, the ubiquity of mismatching form. Whether the discrepancy occurs in the visible and audible properties of speech, as in audiovisual speech perception, or in the phonetic realization of phoneme contrasts, as occurs whenever two individuals speak to each other, it seems that you neither expect nor require your conversational partners to use the identical expressive forms that you use. Or, more precisely, the sharing of words apparently licenses variegation in articulation, both in groups – as dialect unless the group also possesses an army and a navy, in which case it is a language – and in individuals – as idiolect.

### 1.3.3. *Perceiving speech linguistically*

A linguistic emphasis in explanations of speech perception is familiar. The basic notion deriving from Jakobson and Halle (1956) identifies phoneme contrasts as symbolic and linguistic, and neither articulatory nor auditory. In this regard, they assert symbolic status to the phoneme and the word alike. This is subtle, for it warrants a distinction between the form of words (“I said PIN, not PEN”) and their meanings (“I meant PIN, not PEN”). The relation between sound and meaning is arbitrary notwithstanding the contrary claim of *phonesthesia*, a perennial topic of romantic symbolists (Aman, 1980). In order for the listener to know what the talker meant, the listener must resolve the form of the talker’s utterance; without grasping the form of a talker’s speech, a listener has merely guessed the talker’s meaning. It is this juncture that is critical for this conceptualization, because of the complexity in the relation between the canonical form regulated by the language and the expressed form regulated in compromise between linguistic and personal expression.

Initially, accounts of this genre offered a well-defended description of perception as a process of increasing abstraction (cf. Halle, 1985). The difference between phonetic form and canonical phonemic form decreed the initial conditions. Perception began with a sensory pattern, and the perceiver was obliged to transform it in order to resolve its phonemic attributes. The asynchronous distribution of acoustic correlates of a phoneme in a speech stream precludes a simple alignment of the sensory attributes and a canonical segmental series. In this model, several influences on the expressed form of speech must be undone before the segments can be exposed: the effects on the acoustic correlates of phoneme contrasts due to variation in the rate of production, the effects attributable to anatomical scale differences among talkers, the effects due to differential placement of emphasis, to variation in articulatory clarity, to foreign accent, and, of course, the effects due to co-production of sequential phonemes, syllables and words. In short, the characterization depicted a perceiver wielding stable standards – schemas – of the typical sensory presentation of the phonemes in the language, and applying a perceptual function to strip the instance-specific detail from an impinging sensory stream. Once a sensory sample was recast with sufficient abstractness, it was fit to match a stable linguistically-determined form.

Evidence from the listening lab had calibrated a perceiver’s suppleness in adapting to the properties that drive the expressed form of speech to depart from a hypothetical abstract form. If some proposals relied on a dynamic that operated feature by feature

(Stevens, 1990), others described the comparison of segmental instances to prototypes (Samuel, 1982), and, in contrast to principles of likelihood, other accounts invoked a standard of segmental goodness independent of typicality yet still subject to the influence of experience (Iverson & Kuhl, 1995). The shared premise of these accounts is the use of progressive abstraction for the perceptual accommodation to variability. Categories of phonemic experience are rightly understood as commutable markers of contrast independent of talker or circumstance. After all, there is no pair of words that depends for its contrast on production by a specific talker, at a specific speech rate, paralinguistic expression of affect, vocal pitch, etc. But, the actual phonetic form of speech is too bound to the local conditions of production to be simply redeemed as an abstract phoneme series composing a word. The view that the incommensurate phonetic and phonemic forms are harmonized by reshaping the phonetic form into a less specific and more general version has been called *abstractionist* (Richardson-Klavehn & Bjork, 1988).

In such accounts, to appraise an unanalyzed bit of speech a perceiver must reconstruct the incoming sensory form to permit contact with a schematic idealization, or so an abstracting account could claim until critical studies of priming with spoken words. In a priming paradigm, the effect of a collateral probe (called “the prime”) on the performance of a perceptual task is generally taken as evidence of relatedness. The closer the relation of a prime and a target, the greater the facilitation by the prime of a test subject’s act concerning the target. This description of perception as the recognition of an abstract form warranted equivalence of the detailed phonetic variants of a spoken word used as prime and target because the point of contact inherent to identification was allegedly indifferent to the disparity among spoken instances of the same canonical phonemic form. But, in a series of studies that dislodged abstraction as the orthodox formula in speech perception, test subjects proved to be acutely sensitive to the exact phonetic similarity of prime and target, as if the specific phonetic attributes were preserved, and not simply registered as a preliminary to the process of abstraction requisite to identification (Goldinger, Luce, Pisoni, & Marcario, 1992; Luce, Goldinger, Auer, & Vitevitch, 2000).

In a description of perception by abstraction, the set of contact points is given by the number of resolvable types. The set is potentially small if the segmental phoneme inventory is used. If legal pairs or triads of phonetic segments are used, the set is larger, perhaps tens of thousands for English in comparison to the three dozen phoneme segments, but this set size can hardly be taxing on a nervous system capable of impressive feats of rote learning. But, imagine an indexing scheme representing instance-specific variation: it expands without limit. In contrast to the notion of the infinite use of finite means at the heart of every generative system, long-term knowledge that only encoded every raw instance is simply not compatible with the componential nature of phonology and morphology not to mention parity at any level. And, this consideration cannot apply solely to perceived form, for some studies had shown that we track the differential likelihood associated with the modality of the instances (Gaygen & Luce, 1998). That is, a spoken instance is encoded in a form distinct from a heard instance; a typed instance is marked in memory to distinguish it from a read instance. So far, there has been general agreement that these varied instances coalesce into types that match the abstract forms, preserving

the linguistic drivers of differentiation of lexical items through highly varied realization of canonical form. But, how are instances encoded?

In some descriptions of the adaptive resolution of superordinate phonemic types and subordinate phonetic instances, each level is treated as a linguistic representation derived from a raw sensory sample (Goldinger & Azuma, 2003). The instance is preserved as an unelaborated residue of stimulation. A literal understanding of an instance specific memory of utterances warrants a sensory encoding, for this is the only kind of representation that does not oblige the perceiver to an interpretation that substitutes for the direct experience of the instance. Yet, this notion can only be sustained in disregard of the psychoacoustic benchmarks of speech sounds (for instance, Pisoni & Tash, 1974). The unelaborated impression is gone in a tick of the clock. Indeed, the fleeting trace of an utterance arguably forces the retention of instance specific attributes while precluding an encoding of raw auditory experience.

We do not know the form of instance specific attributes yet, though some studies show that a perceiver is exquisitely sensitive to subtle phonetic variants, those that are far more detailed than simple categorization requires (McLennan, Luce, & Charles-Luce, 2003). Some phonetic variants are obviously due to chance—speech produced with food in the mouth, for instance, includes concurrent acts that compromise the expression of linguistic and paralinguistic properties with the accidental moment by moment acts to retain the bolus of food in the mouth. Other subtle phonetic variants are regulated, such as those that distinguish dialects and idiolects, and from their consistency we can infer that their production is perceptually monitored, and that phonetic perception incorporates dialectal and idiolectal dispositions, at least some of the time. Such sensitivity to varieties of phonetic expression at large in a language community might have played a role in findings that subphonemic discrimination of speech sounds always exceeded a prediction based on phoneme identification (see Liberman, 1957). Although these reports had been explained as an expression of auditory sensitivity, fine grain phonetic differences exist at a parallel level of resolution, and it is likely that perceivers attend to this detail because at this grain the linguistic and paralinguistic drivers of expression converge. A finding of instance specificity is potentially reducible to allophonic specificity, at least in linguistic dimensions of this phenomenon. But, not all specificity will be reducible to linguistically regulated properties of speech. In order to explain episodic properties of utterances – you were standing in the moonlight, the breeze was lightly rustling the leaves and a firefly twinkled just as you whispered, “Jazz and swing fans like fast music” – a state-dependent form of inscription might be required, but this is unlikely to be central to language. If this approach to speech perception holds potential for explaining the core problems that motivate research, perhaps because it is the most freewheeling of the accounts we have considered. Others lack the suppleness required by the accumulated evidence of the perception of speech as a cognitive function that finds linguistically specified contrasts under conditions that defy simple acoustic, articulatory, visual and tactile designation.

A listener who attends to subtle varieties of phonetic expression in speech is obliged to do so by the lack of uniformity in speech production. In accommodating this aspect of

variation, a listener meets a challenge created by language communities. The individuals who compose our communities vary in anatomical scale, dialect and idiolect, age, social role and attitude, and these dimensions are expressed in each utterance along with the linguistic message. If the sensory samples reflect these converging influences on expression, it is not surprising that a listener's attention to the attributes of a spoken event include features of the talker and the conditions in which an utterance occurred. In perceiving speech, a listener attends to personal attributes of the talker, and research on the perception of individuals, though it has sometimes run parallel to studies of linguistic perception, ultimately converges with it.

## 2. PERCEPTUAL IDENTIFICATION OF THE TALKER

Research on the perceptual organization of speech establishes that no set of auditory qualities is necessary for the resolution of linguistic form. A listener experiences linguistic impressions evoked by an indefinitely wide variety of acoustic causes. But, the quality of the voice, varying as widely as the acoustic causes of phonetic impressions, is undeniably salient. It is consistent with this intuition that some models have apportioned the perception of voice quality to a separate analyzer, one that is concerned with nonlinguistic attributes, including indexical properties of the talker producing the utterance. Such a clean separation of linguistic and indexical perceptual organization is a ready solution to the problem entailed in between-talker perturbations of linguistic form, the effects of which have been charted in numerous studies (see reviews in Johnson & Mullennix, 1997). A corollary question provoked by the separation of linguistic and indexical perception is whether auditory quality is necessary for the discrimination and identification of different talkers. Although the apportionment of functions warrants it, both forensic and laboratory investigations of talker identification and discrimination yield little hope of delineating a set of acoustic attributes or auditory qualities that designate individual talkers. Recent research on the effects of variation across individuals and instances in the perception of phonetic form indicate that indexical and linguistic processes are likely to be concurrent organizations of the same underlying cause.

The first empirical study of individual identification from speech noted the prevalence of voice recognition testimony by earwitnesses in court cases, despite a lack of scientific evidence attesting to a listener's ability to perform such a task (McGehee, 1937). Twenty years after Bell Laboratories created the sound spectrograph for speech analysis, Kersta (1962) developed a talker verification technique that involved visual analysis of spectrograms by trained experts, and coined the term *voiceprint* in hopeful analogy to the fingerprint. The contentious use of this technique in court cases and by government investigative agencies motivated much of the research on talker identification in the 1960s and 1970s (Hollien & Klepper, 1984; Kreiman, 1997); despite the controversy, earwitness evidence and expert analyst testimony are admissible in courts on a case-by-case basis – but not in Maryland. From the start, talker identification by ear was found to vary widely in accuracy. This reflects the fact that the acoustic products of an individual talker, while unique, are not distinctive. Arguably, the qualitative ways in which talkers differ

from each other and the way the speech of a single talker fluctuates in different situations derive from the same source: variability in phonetic repertoire. Without a dependable set of voice quality attributes to allocate to a separate stream, concurrent resolution of a word and a talker can be viewed as a form of multistability – obliging a listener’s attentional finesse rather than segregation of analytical functions.

## 2.1. Talker Identification by Human and Machine

It seems effortless in ordinary circumstances to identify a talker by listening to speech, especially in a context of other commonplace events. Speech often occurs in situations in which conversants see one another, and even without visual contact, the uncertainty of a talker’s identity can be constrained by other situational factors. For relatively long stretches of speech (>1200 ms) recorded under optimal conditions with a closed set of familiar alternative talkers, talker identification is nearly perfect, unsurprisingly (see reviews by Hollien & Klepper, 1984; Kreiman, 1997). Compromising any one of those ideals leads to a predictable decline in identification performance, and familiarity with the talker set and language are particularly important for accurate identification (Hollien, Majewski, & Doherty, 1982). Talker identification from speech can be difficult to explain because, unlike physical residues such as fingerprints, a talker’s acoustic realization of the same word varies from instance to instance; and, no single acoustic attribute varies less across an individual’s utterances than between individuals, precluding an account of identification by a simple acoustic feature. Talkers differ in multidimensional aspects of voice quality, and researchers have expended considerable effort attempting to find reliable psychoacoustic models of these differences (see Kreiman, Vanlancker-Sidtis, & Gerratt, 2005; Laver, 1980; Nolan, 1983). The partial success of this enterprise has yielded machine implementations of talker identification routines that are used in security applications.

The developers of automatic methods of talker identification might promise very high accuracy rates for their systems, similar to those found when human listeners recognize a familiar talker from speech. These methods benefit from the fact that a cooperative and motivated individual provides the standard and comparison samples, and the bank of standards is limited to a finite number of individuals. The main challenges for such systems are the possibility that an intruder could use a recording of the target individual to trick the recognizer, or that the circumstances during collection of a comparison sample deviate too far from those of the standard. For example, IBM’s Voice Identification and Verification Agent boasts a 1% false alarm rate with only a 3% false rejection rate for a 20 s automated interview procedure (Navratil, Kleindienst, & Maes, 2000). Similar claims are made by the current industry leaders in voice identification technology, Nuance and SpeechWorks. The apparent success of these products arises as much from an individual’s private knowledge as from a faithful acoustic rendering of the voice – these products rely on proprietary spectral analysis routines for acoustic voice verification, yet the system architecture employs an error-prone speech recognizer to verify answers to the interview questions. Although these systems can be found

in many corporate settings, companies have not completely forgone the use of live agents to handle false rejections from the routines, because automatic recognizers are so labile to subtle acoustic differences within the same talker. Despite claims to the contrary by the developers of the systems, the lack of more widespread application of such devices attests to the impracticality of their use.

### 2.1.1. *Talker identification in forensic settings*

In forensic applications of talker identification, an earwitness to a crime attempts to match a target talker from memory, or an expert analyst attempts to match a recorded sample of an unidentified talker to a set of known alternative talkers using a combination of auditory and visual spectrogram inspection. Under these conditions, it is impossible to calibrate accuracy because the set of alternative suspects might happen not to include the actual perpetrator. When a listener is not familiar with a talker, identification can be quite poor, especially with a potentially uncooperative suspect who might adopt a vocal disguise – although identification accuracy in laboratory settings is above chance across a set of listeners who are familiar with the talker (Hollien et al., 1982). Often, the target speech sample might not have been produced or recorded under quiet conditions, and the analyst must rely on degraded acoustic samples and subsequently poor spectrographic representations.

From its humble beginnings in the 1960s, there has been an effort to establish the admissibility of voiceprint analysis into court proceedings, with some success. At best, an earwitness or voiceprint examiner can choose the member of a voice line-up who sounds most similar to the original talker, or come to no decision. Expert voiceprint analysts often opt for no decision. However, for the victim of a crime providing earwitness testimony, the motivation to identify a perpetrator might override the uncertainty inherent in such a task, leading to a mistaken identification. Moreover, any decision based on relative similarity is extremely sensitive to the set of alternatives. To construct a fair test, a talker identification line-up should include a single target suspect paired against a set of talkers known to be innocent who are similar to the target in sex, gender, dialect, height, age, weight, socio-economic status and level of education. All of these factors have been found to form the basis for discrimination among talkers, and some are known to be identifiable with a moderate level of accuracy from speech alone (for a recent casting of this contentious issue, see Krauss, Freyberg, & Morsella, 2002; Morsella, Romero-Canyas, Halim, & Krauss, 2002).

### 2.1.2. *Acoustic basis for talker identification*

Do any specific acoustic parameters elicit a consistent impression of a talker's identity? Ideally, an acoustic attribute for talker differentiation would be a correlate of fixed anatomical differences between talkers, impervious to speaking habits and situational corruption. In order to identify an individual talker, not only must the parameter distinguish different talkers, but it must also be consistently produced by a given talker, or at least the variation in the production of this hypothetical attribute by a single talker should be smaller than

that which occurs between talkers. Previous reviews of talker identification evaluate the utility of many acoustic parameters, ranging from average phonatory frequency ( $F_0$ ) to more derived measures of laryngeal and supralaryngeal acoustic effects (Hollien & Klepper, 1984; Kent & Chial, 2002; Laver, 1980; Nolan 1983). Most of these measures require about a minute of fluent speech to provide stable estimates of a single talker, the reliability of which could be due to the increase in sampled phonetic repertoire as well as the absolute amount of acoustic data (Bricker & Pruzansky, 1966).

Perhaps the most obvious differences between talkers are due to sex (male in contrast to female) and age (children in contrast to adults), which despite great overlap throughout the range are conspicuously expressed as differences in average  $F_0$  and formant frequencies, presumably due to differences in vocal anatomy. The large differences in average  $F_0$  among males, females and children (132 Hz, 223 Hz and 264 Hz; Peterson & Barney, 1952) have commonly been treated as differences in scale (see Fant, 1966). However, because an individual talker produces a wide range of variation in  $F_0$  that overlaps with other talkers'  $F_0$  distributions, a single measure alone is not enough to differentiate even males from females. Moreover, average  $F_0$  is not reliable over time for the same talker, so any difference that is large enough to differentiate talkers of the same sex is likely to typify the same talker at different times (Markel & Davis, 1979). An automatic routine that only exploited long-term properties of  $F_0$  would be vulnerable to the use of a different recording of the same talker at another time or under different circumstances, leading to a false rejection. Although including  $F_0$  aids perceptual identification relative to filtering or manipulating speech, it is not clear whether the improvement in identifying a talker is based on static long-term properties or dynamic short-term fluctuations of  $F_0$  (Nolan, 1983). Furthermore, because pitch is perceptually salient, it can be readily manipulated by a talker attempting to disguise the voice.

In supralaryngeal measures, female vowel formants are roughly 20% higher on average than male formants, and child formants are about 20% higher than adult female formants. However, the magnitude of these differences varies widely across both the individual formants ( $F_1 < F_2$  or  $F_3$ ), the vowel inventory ( $[ɔ] < [æ]$ ), and different languages, therefore, they can not be a simple function of talker anatomy (Fant, 1966; Johnson, 2005). There is no characteristic formant profile for an individual talker that is stable over time. In order to eliminate the sample bias of variation caused by different assortments of individual segments on the measurement of formant frequencies, long-term spectra have been studied. These aim to characterize a talker independent of phonetic samples by averaging the short-term power spectrum over a long span of speech, generally at least 10 s of continuous speech (Furui, Itakura, & Sato, 1972). Like average  $F_0$  and formant frequency estimates, long-term spectra are not reliable for indexing an individual talker over different speaking contexts. Compared to identification by human listeners hearing familiar talkers, a talker identification routine based on long-term spectra is more severely affected by distress and disguise (Hollien et al., 1982). Long-term spectra do reflect differences in voice quality, but as correlates of individual talkers they are not stable, failing to track the same talker under different speaking conditions. Figure 3 illustrates long-term spectral measures adapted from



Nolan (1983). For example, one talker adopting two different supralaryngeal postures, neutral or pharyngealized, shows as much change between postures as another talker adopting the same contrast; and, the global change is not uniformly reflected in the long-term spectral consequences for each talker. As shown in the bottom plot, additional derived measures of long-term spectra fail to distinguish an individual talker consistently across different contexts of speaking – for some articulatory postures, the two talkers overlap in the log-transformed slope of the long-term spectra, while for other postures, the talkers are very distinctive.

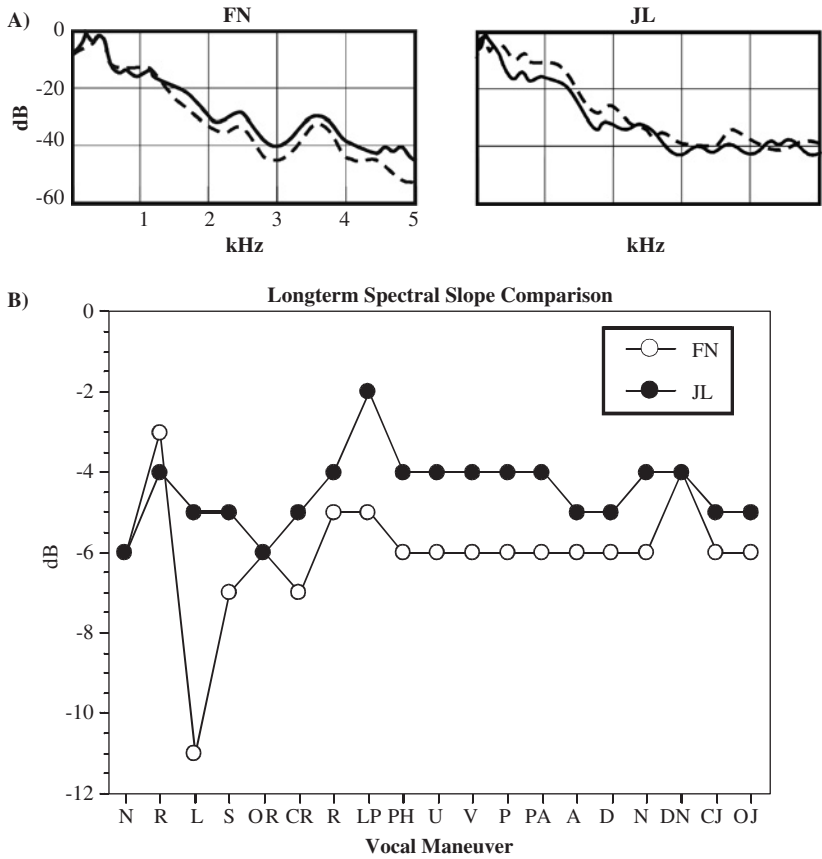


Figure 3. The effect of selected vocal postures on long-term speech spectra (Nolan, 1983). (A) Neutral and pharyngealized spectra for talkers FN and JL; (B) Effect of vocal maneuvers on the slope of the speech spectrum for talker FN (open bullet) and JL (filled bullet): n=neutral; r=raised larynx; l=lowered larynx; s=labial/spread lips; or=labial/open rounding; cd=labial/close rounding; r=retroflex; lp=laryngo-pharyngealized; ph=pharyngealized; u=uvularized; v=velarized; p=palatalized; pa=palato-alveolarized; a=alveolarized; d=dentalized; n=nasalized; dn=denasalized; cj=close jaw; and oj=open jaw.

### 2.1.3. *Phonetic identification of talkers*

The differences in long-term spectral measures of an individual talker across different postures are roughly equivalent to those between talkers. Despite an impression that such complex acoustic measures provide a context-free representation of a talker's unique vocal profile, they are not characteristic of the talker because they do not designate the talker independently of different vocal postures. Furthermore, many qualitative supralaryngeal vocal postures involve anatomical structures used in phonetic expression, influencing the production of phonetic forms. Laver (1980) argues that it might be possible to derive the postures from patterns of variability in phonetic expression, making voice quality a function of phonetic form.

Because of the versatility of the speech production system, Nolan suggests that voice quality measures may best be used to classify rather than to identify talkers absolutely. Indeed, many empirical projects point out the lack of reliability in judgments of talker identity based on impressions of voice quality (see review in Kreiman, 1997). Voice quality is not an independent perceptual stream, divorced from phonetic attributes, that can be used to identify the speaker of a message. Even nasal resonances, long considered good cues for talker identification due to the relative anatomical immutability of the nasal cavities (Glenn & Kleiner, 1968; Sambur, 1975; Su, Li, & Fu, 1974), are susceptible to many situational factors, including everything from colds to stylistic variation.

Some empirical studies of perception provided exclusively dynamic acoustic attributes of speech, using a synthesis technique to exclude a contribution of F<sub>0</sub> and familiar voice quality to perception. Under such extreme conditions, listeners still resolved phonetic attributes and talker identity (Remez et al. 1997). In these studies, the three lowest formants of speech were replaced with time-varying sinewaves, with a fourth tone following the fricative poles. These signals lack the natural acoustic correlates of vocal sound production, yet the linguistic message was readily resolved; the acoustic transformation to sinewave tones evoked fine grain phonetic properties. Listeners could identify both familiar and unfamiliar talkers in closed set sinewave tests, and the performance by listeners who knew the talkers was similar to those who simply matched a natural sample to a sinewave utterance. Arguably, a listener who recognized a sinewave version of a familiar talker must have relied on longstanding knowledge of a particular talker's articulatory habits, or idiolect, which comprise linguistic, not qualitative, attributes. It appears that a listener will use any parameter, even a phonetic one, that can identify a talker. No acoustic attribute is necessary, and each parameter varies in its sufficiency for contributing to the distinction across different situations.

As with identification of linguistic attributes of speech, talker identification is not reliably cued by a particular set of acoustic attributes. Instead, listeners are evidently not particular about the acoustic attributes or the psychoacoustic effects that are used to identify specific individual talkers. Neither the pitch or quality of the voice, nor the short- or long-term spectrum, nor any specific set of acoustic correlates of a talker is required for identification. Apparently, a listener identifies attributes of a talker as if the commitment

to the particular sensory manifestation of the individual is adaptable. In like manner to the perception of phonetic attributes, this inclination to find functional contrasts in unexpected acoustic or auditory form opposes the fixity of a normative conceptualization of talker identification.

## 2.2. Perception of Words and Talkers

A significant portion of the literature on talker differences has considered this source of variation a kind of noise in the correspondence between phonetic and acoustic attributes introduced during the transformation from phoneme to phonetic event. From this perspective, talker differences obscure the acoustic correlates of phonemes that permit a perceiver to resolve words. Many accounts of speech perception propose a normalization function that removes this noise from the signal, arriving at abstract phonemes (Pisoni, 1997; Magnuson & Nusbaum, 2004; Johnson, 2005). This presumes that a listener does not use phonetic attributes to identify a talker, which are stripped from the sensory properties of speech by such functions. A phoneme representation of speech lacks all trace of the circumstances of its uttering, including those phonetic attributes that are distinctive of dialect and idiolect. The abstraction obliged in such accounts leaves only voice quality available for talker identification. However, not only is voice quality an unreliable metric for describing or identifying talkers, it is not necessary for identification of talkers, as shown in studies using sinewave speech. Laver (1980) proposed a set of physiological parameters that can be used to distinguish voice qualities, and many of these overlap with the gestural/articulatory systems used to differentiate phonemes as well (Goldstein & Fowler, 2003; Studdert-Kennedy & Goldstein, 2003). Indeed, a talker's idiosyncratic phonetic repertoire converges on the same series that a listener apprehends as a word.

Preliminary attempts to characterize talker differences aimed to quantify the effect of a particular vocal tract on production of a phone class. Gauging the large acoustic phonetic differences between men, women and children was a good place to start, but such inventories failed to yield a uniform function attributable to vocal tract size (Fant, 1966; Johnson, 2005). In order to remove the effect on a phoneme or word of the sex of the talker, for example, a different scaling function would be necessary for words containing a rounded vowel like /u/ than for an unrounded vowel like /a/, and for individual formants themselves. A hypothetical scaling function that varies with the phonemic class of the segment is of limited value, and consequently it appears unlikely that perceptual rescaling with the sex of a talker plays a role in identifying words, if at all. Acknowledging that talkers also differ in habitual vocal tract postures, such as larynx lowering or raising, retroflexion, or velarization, presents a similar barrier to deriving a talker normalization function. Once again, phonemes differ in their susceptibility to such postures, roughly depending on the relative difference between the anatomical focus of the phonemic and qualitative attributes (Laver, 1980). Perception of words and talkers alike might depend on resolution of detailed phonetic form, which can evoke both phonological and idiosyncratic indexical impressions.

Some recent empirical projects demonstrate that these idiosyncrasies are not always discarded early in the projection of sensory samples to phonemic impressions – they

survive transduction to affect perception of the message. To consider several examples, both implicit and explicit memory of spoken words in a list are affected by whether a single talker or several talkers utter the items (Goldinger, 1996; Nygaard, Sommers, & Pisoni, 1992). Familiarity with a talker differentially facilitates resolution of the words spoken by the talker independent of likelihood or familiarity (Clarke & Garrett, 2004; Goldinger, 1998; Nygaard, Sommers, & Pisoni, 1994; cf. Lieberman, 1963). Failure to notice a change in the talker during a shadowing task has no effect on speed of shadowing, while noticing the change does (Vitevich, 2003). Listeners track differences in voicing, exhibiting graded rather than uniform perceptual categorization of consonants, and the category structures converge on talker differences (Allen & Miller, 2004; Miller & Volaitis, 1989). A listener will even track subcategorical variation in the precise timing of voicing contrasts, and reciprocate them by approximating the temporal expression of voicing in shadowed responses (Fowler, Brown, Sabadini, & Weihing, 2003).

A talker is characterized by more than a collection of simple acoustic attributes imposed upon a word or phoneme, whether such attributes derive from anatomy or from habit, and listener sensitivity to these effects does not take the form of a normalization function yielding only linguistic categories. Talker conditioned within-category variation in phonetic form influences speech perception and memory, at the same time that variants in phonetic form index the talker.

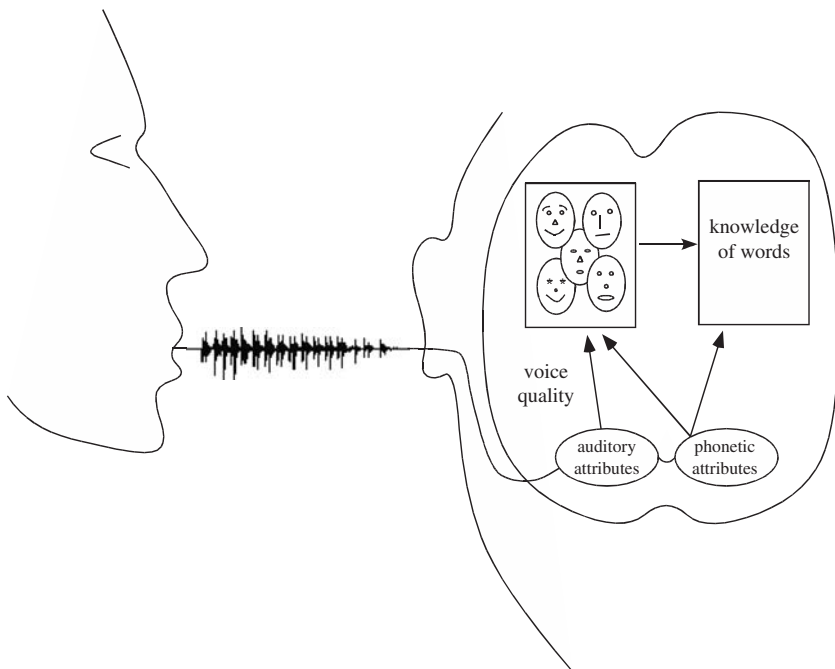


Figure 4. Hypothetical system architecture for the perception of individual talkers from qualitative and phonetic attributes (Remez, Fellowes, & Rubin, 1996).

### 2.3. Individual Stylistic Variation

The attempt to delineate the acoustic attributes correlated with anatomical differences was reasonable because anatomical differences among talkers were thought to be a kind of scalar property limiting the range of other factors that might influence speech production. Changes in a talker's anatomy surely occur – due to growth and maturation, tooth gain or loss, dueling injury – but, roughly, we are stable over long spans. If the dimensions of anatomy are relatively secure, why does the speech of an individual vary across different contexts? Notwithstanding the effects of the compliance of the pharyngeal cavity and the cheeks, tongue and lips on acoustics, much of the moment-to-moment variability that an individual talker exhibits also reflects changing goals, situations and addressees. A talker might produce more or less characteristic attributes of a regional dialect on different occasions (Bourhis & Giles, 1977; Giles, 1973; Labov, 1966, 1986). A talker's emotional states are conveyed in speech, although the acoustic effects are exceedingly complex (Banse & Scherer, 1996; Scherer, 1986; Williams & Stevens, 1972). Talkers shorten the duration of repeated referents in discourse, and listeners attribute such shortening to *givenness*, that is, to a topic already introduced in a narrative (Fowler & Housum, 1987). A large component of research on within-talker differences attempts to characterize the acoustic changes under conditions of clear versus casual speech production (Picheny, Durlach, & Braida, 1985).

The ability to produce clear speech under appropriate conditions demonstrates a talker's ability to control detailed aspects of phonetic form. As surveyed by Uchanski (2005), the main acoustic-phonetic changes from conversational to clear speech include increases in amplitude, duration and F<sub>0</sub>; changes in the slope of the long-term spectrum; phonological effects such as decreased vowel reduction, more prominent stop burst production and elimination of alveolar flapping in coronal consonants; and increases in F1 and F2 frequency and exaggerated voicelessness. Once again, the phonetic differences between the casual and clear speech of an individual deploy the same set of parameters that distinguish different talkers. Of course, when a talker speaks clearly in order to indicate sincerity, or authority, or to reduce the hazard of misunderstanding, this ultimately finds expression in an individual talker's idiolectal phonetic repertoire. And, clear speech for one talker might not be clear for another (Bradlow & Bent, 2003).

Taken together, these phenomena demonstrate the inherent flexibility of language – listeners readily adapt to changing talkers, while still perceiving differences among them. Although comprehension requires relative stability in linguistic form, not all variability must be discarded in service of parity. A listener must track the talker's use of linguistic currency in order to calibrate the tokens available for exchange and to determine what the talker means from what the talker says. With a detailed resolution of phonetic form, speech perception buys both linguistic types and indexical properties. Attention plays this role in speech perception, whether of linguistic or indexical attributes, and the next section describes a self-regulatory process that sets the grain of speech perception with an ear toward speech production.

### 3. PERCEPTUAL SELF-REGULATION

In dexterity, complexity and effectiveness, there is little that we do that surpasses the production of speech. This is not to deny that each of us is occasionally tongue-tied, incomprehensible or wrong. But, apart from the challenge of imagining something useful or clever to say, articulation is normally fast and easy, and on this dimension the role of perception in production was initially misunderstood. The origin is in Lashley's analysis of acts that express a sequential order, described in a founding document of psycholinguistics (Lashley, 1951). He had turned his attention to language because of the empirical opening it offered to oppose the prevailing models of action, which evaded the problem of serial order in a conceptualization of coordinated movement restricted to peripheral sensory-motor chains. Among the arguments in his analysis entailing plans for sequential acts, Lashley contemplated the fluency with which the articulatory components of speech succeed each other. At the phonemic grain, the procession of expressed elements occurs so rapidly that the sensory consequences of the first cannot be responsible for triggering the production of the next, and so on. The second component in a series is initiated, in Lashley's line of reasoning, well in advance of the conduction of the afferent consequences of the preceding component, whether taken as auditory, orofacial tactile or muscle sense. In this circumstance, he proposed that a series of acts was composed and controlled centrifugally, without slowing to the crawl obliged by a sequencing mechanism of sensory triggering. He was right in part.

Phonetic expression is incredibly stable. Over decades, habits of articulation are consistent (House & Stevens, 1999), and this stability reflects the constancy of anatomical and functional constraints on vocal acts. That is to say, skeletal actions must be adaptable to an environment of displaceable masses and extrinsic forces. In contrast, the topography of vocal landmarks – teeth, palate, tongue – is relatively fixed, and the masses intrinsic to the vocal tract vary only gradually over the lifespan and not at all during an act. If the microenvironment composed by a vocal tract is durable, the vocal acts committed within it need not be especially adaptable. From this perspective, it is not surprising that research pursuing Lashley's speculation found ample evidence that speech production is hampered very little by the absence of sensation.

The research corroborating this viewpoint about articulatory control aimed to assess the consequences of disrupted sensation on speech, and the findings were consistent. To interfere with auditory experience of speech, a noise load is imposed, masking a listener's self-productions (Lane, Catania, & Stevens, 1961); to interfere with orofacial taction, the lingual, labial and pharyngeal surfaces are anesthetized (Borden, Harris, & Oliver, 1973); to interfere with muscle sense, the intrafusal muscle fibers supplied by the mandibular branch of the trigeminal nerve are selectively albeit reversibly blockaded (Abbs, 1973). None of these sensory streams is essential to the control of speech production. It is easy to see that such studies would warrant an account of the articulation of speech in central open-loop control, given this juxtaposition of prodigious expression in both variety and fluency, and relative proficiency of articulation in conditions that preclude sensory monitoring.

If an adult can succeed in producing speech without sensory supply, this does not mean that language development can occur without the means to align self-produced vocalization with the speech of others, whether in the early home setting of infancy or in the schoolyard society of juveniles. After all, though the internal environment of the vocal tract is set, a conversational environment is often distressingly unpredictable. The deterioration of spoken language that can follow deafening in childhood (Binnie, Daniloff, & Buckingham, 1982) is evidence of the importance of auditory reafference; in the classic contrast, the causes of sensory states are dichotomized as *reafferent*, owing to the consequences of self-produced action, and *exafferent*, owing to extrinsic objects and events independent of self-produced action (von Holst & Mittelstaedt, 1950). Aside from developmental functions in which the exafferent effects of the speech of a linguistic community are reprised in the reafference of a young talker, there is ample evidence that self-regulation of phonetic production depends on reafference which, though inessential in the short-term, is exploited nonetheless when it is available. This theme in speech perception research is evolving, though it is possible to see the principles emerging in studies of detailed phonetic production. Throughout, the mark of reafferent control of speech is the constraining effect of linguistic repertoire on sensitivity and production alike.

### 3.1. Vocal Effort

The original report by Étienne Lombard (1911) of reafferent control of speech production pertained to vocal effort. A talker adjusts the power of speech as if to maintain a constant difference between the sensory effects of the voice and the momentary extrinsic noisemakers obscuring conversation. Reviewing studies of this *Lombard sign*, Lane and Tranel (1971) reported that in addition to reafferent control, the functions exhibited an intriguing communicative contingency, namely, compensation in production for noise load occurred only when a listener was present. If compensation were purely an egocentric self-regulatory function, it would appear regardless of the presence of an audience. However, talkers did not regulate vocal power when reading a script into a microphone with noise presented over headphones; only conversational settings induced the critical adaptive pattern. Moreover, vocal effort is also regulated by a crossing factor, the perceived distance between the talker and listener (Liénard & Di Benedetto, 1999). The standards for regulating appropriate vocal effort involve power in relation to noise corrected by the talker's implicit goal of maintaining the sound level at the listener's ear. Some portion of this regulatory ability is not restricted to humans producing speech: Zebra finches also adapt vocal effort to ambient noise (Cynx, Lewis, Tavel, & Tse, 1998).

### 3.2. Phonation

Reafferent control of vocal effort is a relatively gross if complex parameter, and not the only evidence of reafferent regulation. Studies of more detailed control indicate that the pitch of the voice falls under reafferent control. In one line of research on this aspect of production (Jones & Munhall, 2000), a subject articulated a long syllable under

conditions in which the acoustic experience of self-produced  $F_0$  provided by the experimenters was veridical or altered electronically. When phonatory frequency was modified, it was transposed up or down. Participants in the study compensated for the displacement, opposing the perturbations established electronically. For instance, when pitch was shifted upward in frequency, a talker lowered vocal pitch to bring reafferent experience to the internal standard.

### 3.3. Phonetic Production

Critical findings about the regulation of fine segmental structure are also reported. In one, Houde and Jordan (1998) created an acoustic-phonetic analogy to studies of visual perceptual adaptation using displacing prism spectacles. In the visual circumstance, viewers gradually recalibrate their reaching movements to accommodate the illusory displacement of the visible world created by the prisms. Moreover, immediately after removing the prisms, reaching shows an adaptive rebound of the discontinued visual displacement. In auditory displacement during speech production, Houde and Jordan provided listeners with online acoustic perturbations of their own utterances. Over the course of a prompted, whispered syllable production task, talkers heard the natural acoustic consequences of vocal sound production, but the speech spectrum was modified electronically; the method induced a tolerable 16 ms delay. The modification entailed shifting formant frequency so that a talker heard a different vowel in response to the produced vowel: / $\epsilon$ / formants were shifted either up to / $i$ / or down to / $\text{æ}$ /, always produced in a /bVb/ syllable. In the context of other consonants, / $\epsilon$ / formants were unaltered, and formants from a talker's production of other vowels were likewise unaltered. While receiving perturbed feedback, most talkers shifted their productions of the altered vowels to compensate for the distortion, resulting in production of lowered or raised vowels, as appropriate to counter the formant frequency displacement. These altered productions, once shifted in frequency by the electronic apparatus, matched the reafference typical of the intended vowel produced in the clear. Furthermore, compensation generalized to production of the same vowel in different consonant contexts from the training set and to different vowels, and talkers persisted in shifted productions when the altered reafference was replaced with masking noise.

Reafferent control is not limited to auditory samples, an aspect of the multimodal sensory expression of phonetic attributes. In one instance of the effectiveness of tactile and muscle sense, a talker was outfitted with a dental prosthesis effectively lengthening the maxillary incisors (Jones & Munhall, 2003). A variety of changes in phonetic expression would be expected following a derangement of familiar dimensions, of course, and this study focused on instances of /t/. A change in a stable feature of the vocal anatomy can be expected to disrupt articulation, and to impose a requirement to adapt specific aspects of sound production. A subject in the study improved in producing natural-sounding /t/ as a result of experience with the dentures over a testing hour when auditory reafference was blocked; once auditory reafference was allowed, there was little benefit beyond that which somatic reafference established. In its time course, the slow adaptation to an abrupt



change in a fixed vocal structure is familiar from our own experience as dental patients. It can take a while to get used to speaking with a bite splint or new choppers, and auditory reafference of the altered production that ensues is motivating but inadequate to elicit immediate compensation.

Adaptation of speech production to the presence of a pseudopalate that changes the shape of the roof of the mouth exhibits a similar time course and impact on fine placement (Baum & McFarland, 1997; McFarland, Baum, & Chabot, 1996). In contrast, a functional constraint, such as occurs when the motion of the jaw is fixed during speech production by the concurrent requirement to hold an object between the teeth, is readily assimilated. Although some individuals fare better than others in adapting to functional limits, in general reafference rapidly drives production toward typical phonetic expression. Indeed, although canonical form might be difficult to realize with the jaw height or the lip aperture fixed and motionless, a talker is not inexperienced in reconciling the functions of ingestion, deglutition and respiration with sound production, and either practice or endowment are exploitable to minimize the effect of this sort of limit in action.

Of the studies examining the effect of altered auditory reafference, the introduction of systematic departures from veridical samples provided evidence about the contribution of sensory states to phonetic production. The alterations created by researchers are subtle, though, in comparison to the drastic departure from veridicality experienced by a user of a cochlear implant. An implant user typically becomes accustomed to the anomalous quality delivered by the stimulator, an electrode that uses a coarse grain place-code to evoke a rough correlate of incident spectrotemporal acoustic properties in the activity of the auditory afferents. Pitch experience is evoked only poorly if at all, and experience of melody and harmony is meager in contrast to meter and rhythm. Nevertheless, adults who rely on such recurrent sensory qualities to control speech production can perform well (Vick et al., 2001). After a year of use the regulation of production extends throughout the English phone classes, and despite variation in success with adventitiously deafened linguistically competent adults the preponderant outcome observed in one survey establishes the value of reafference in sharpening phonetic production (Gould et al., 2001). Intuitively, an adult who already expresses and comprehends language might seem to have an advantage of experience in exploiting an impoverished sensory sample of speech, whether to perceive the speech of others or to produce speech by reafferent regulation. But, astonishingly, some children whose deafness is profound by age 3 can employ the reafference available through an implant to speak and to comprehend speech well enough to master English phoneme contrasts (Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000). It is unclear why so many children fail, though the success of many who can learn language this way defines the problem as linguistic rather than one specifically of sensory veridicality (Watson, Qiu, Chamberlain, & Li, 1996).

#### **3.4. Self-Regulation in Conversation**

One prominent claim about the production of speech in conversational conditions holds that production is regulated allocentrically as well as egocentrically. That is to say,

the talker regulates production with respect to the listener's state, and not simply with regard to the refference associated with self-produced speech. In the elaboration by Lindblom (1990), conversation is described as a competition, in which a talker aims to minimize the effort in articulation by neutralizing contrasts under conditions in which the cost of miscommunication is offset by a listener's vigilance and acuity. The listener, complementarily, aims to minimize the cognitive load of vigilance and acuity by relying on the talker to produce salient versions of the linguistic contrasts. This talking version of prisoner's dilemma characterizes the self-regulatory options for conversations, and views the participants as not entirely cooperative, to its great merit. But, it does discount the evident sharing that occurs when people talk.

Studies of interacting talkers have found fairly consistent patterns of linguistic change over the course of conversational interaction, and such changes are variously termed coordination (Clark, 1996), alignment (Pickering & Garrod, 2004) or accommodation (Shepard, Giles, & Le Poire, 2001). Most of these projects examined the increase in similarity (convergence) of diverse aspects of interlocutor's speech, from the schematic (Garrod & Doherty, 1994), to the syntactic (Branigan, Pickering, & Cleland, 2000), to the lexical/semantic levels (Krauss & Weinheimer, 1964; Wilkes-Gibbs & Clark, 1992). Research on convergence at sub-lexical levels includes measures of acoustic-phonetic attributes such as perceived accentedness, fundamental frequency covariation, and voice amplitude (Giles, 1973; Gregory, 1990; Natale, 1975). Convergence in such parameters appears to be particularly susceptible to the effects of social factors that are confined to communication exchanges, such as interlocutors' relative dominance or perceived prestige (Gregory, Dagan, & Webster, 1997; Gregory & Webster, 1996).

Some reports also describe a pattern opposite to convergence under some circumstances (Giles, Coupland, & Coupland, 1991; Shepard et al., 2001). Although it is tempting to attribute convergence to an automatic imitative function facilitating increased intelligibility (e.g., Pickering & Garrod, 2004), divergence often does not preclude intelligibility, but serves a communicative need for the diverging party (Bourhis & Giles, 1977; Labov, 1974). Indeed, a recent study found that interacting talkers differed in the extent to which they converged in phonetic repertoire over the course of a single cooperative task (Pardo, 2006). The study employed a modified version of the Map Task (Anderson et al., 1991), in which paired talkers are given matched sets of iconic maps with labeled landmarks; in one set, the maps contain paths around various landmarks, while the other set of maps have only labeled landmarks. The talkers cannot see each other or their partner's maps, yet they must communicate effectively enough that the instruction receiver can duplicate the paths on the instruction giver's maps. The talkers in this study were able to perform the task well, and they also converged in phonetic repertoire over the course of the short interaction, but instruction givers converged to instruction receivers more than the reverse. Furthermore, the shifts in phonetic repertoire persisted to a recording session conducted immediately after the conversational setting. Because some talkers converged more than others, phonetic assimilation cannot be attributed to an automatic function in which perception of an addressee's phonetic variants primes recurrent production of the variants. Rather, to account for the asymmetry,

consider that perception yields phonetic variants available for selection, depending on paralinguistic functions operating in parallel with linguistic functions. In this case, those functions related to the talker's role in the conversational setting. Given that the phonetically variable lexical tokens were perfectly intelligible to both parties, it is arguable that phonetic convergence serves communicative purposes beyond that required for intelligibility – a listener tolerates a wide variety of acoustic–phonetic forms when perceiving speech produced in conversational interaction.

The inherent tolerance of variability licenses each talker to sample a broad contrast space to compose a unique phonetic repertoire. Conceivably, once an idiolect forms, some parameters remain relatively free to assimilate in social exchange. An elaboration of the circumstances that promote or preclude such compliance would provide a clearer view of the processes shaping the phonetic landscape from its linguistic and paralinguistic bases. Overall, the consideration of perceptual self-regulation in production exposes the accommodating relationship of perception and production despite the difference in grain. This asymmetry of range and detail – because we can do so little in comparison to what we discern – precludes a rigid coupling of perceived and produced speech, mirror neurons notwithstanding (Rizzolati & Cragheiro, 2004). Instead, attention at the phonetic grain informs both linguistic and paralinguistic functions, evoking linguistic forms that are regulated to fulfill the aims of communicative exchanges.

#### **4. A CONCLUDING WORD**

In this discussion of the theoretical and empirical study of the perception of speech, we have emphasized the principles motivating classic and recent explanations. Instead of a comprehensive review of research paradigms within this lively domain, we have adopted a functional focus on speech as an expression of language. As a consequence of this commitment, speech perception is cast as the means by which spoken expressions are resolved as coherent audible and visible events, as strings of familiar linguistic forms, as the product of an individual talker, and as a model for recurrent self-expression. Ostensibly, these related aspects of speech perception are the ordinary experience of talkers and listeners, and do not depend for their justification on specific research methods or puzzles. New investigations will determine whether the perspective articulated here proves to be useful in the search for new evidence to refine the descriptive and explanatory inquiry into the perception of speech.

#### **ACKNOWLEDGEMENT**

We are grateful for the generosity of our colleagues whose instruction and advice were helpful in setting the focus of our discussion: Carol Fowler, Louis Goldstein, Michael Studdert-Kennedy, David Pisoni, and Philip Rubin. We also offer special thanks to Daria Ferro for her steady hand and clear eye in drafting the artwork. This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (DC00308) and a grant from the National Science Foundation (0545133) to Barnard College.

**REFERENCES**

- Abbs, J. (1973). The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations. *Journal of Speech and Hearing Research*, *16*, 175–200.
- Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *115*, 3171–3183.
- Aman, R. (1980). Clean up your fexing language! Or, how to swear violently without offending anyone. *Maledicta*, *4*, 5–14.
- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task corpus. *Language and speech*, *34*, 351–366.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636.
- Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America*, *102*, 2353–2359.
- Bernstein, L. E., Auer, E. T., Jr., & Moore, J. K. (2004). Audiovisual speech binding: Convergence or association. In: G. A. Calvert, C. Spence, & B. E. Stein (Eds), *Handbook of multisensory processing* (pp. 203–223). Cambridge, MA: MIT Press.
- Bertelson, P., Vroomen, J., & de Gelder, B. (1997). Auditory-visual interaction in voice localization and in bimodal speech recognition: The effects of desynchronization. In: C. Benoît, & R. Campbell (Eds), *Proceedings of the workshop on audio-visual speech processing: Cognitive and computational approaches*, Rhodes, Greece, ESCA (pp. 97–100).
- Binnie, C. A., Daniloff, R. G., & Buckingham, H. W. (1982). Phonetic disintegration in a five-year-old following sudden hearing loss. *Journal of Speech and Hearing Disorders*, *47*, 181–189.
- Borden, G. J., Harris, K. S., & Oliver, W. (1973). Oral feedback, I. Variability of the effect of nerve-block anesthesia upon speech. *Journal of Phonetics*, *1*, 289–295.
- Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In: H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119–135). London: Academic Press.
- Bradlow, A. R., & Bent, T. (2003). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, *112*, 272–284.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–B25.
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Bricker, P., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America*, *40*, 1441–1450.

- Browman, C., & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes and historical change. In: I. G. Mattingly, & M. Studdert-Kennedy (Eds), *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman* (pp. 313–338). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation, *Journal of Experimental Psychology: Human Perception and Performance*, 27, 115–127.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25, 975–979.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, 116, 3647–3658.
- Cynx, J., Lewis, R., Tavel, B., & Tse, H. (1998). Amplitude regulation of vocalizations in noise by a songbird, *Taeniopygia guttata*. *Animal Behaviour*, 56, 107–113.
- Diehl, R. L., Kluender, K. R., Walsh, M. A., & Parker, E. M. (1991). Auditory enhancement in speech perception and phonology. In: R. R. Hoffman, & D. S. Palermo (Eds), *Cognition and the symbolic processes: Applied and ecological perspectives* (pp. 59–76). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179.
- Eimas, P. D., & Miller, J. L. (1992). Organization in the perception of speech by young infants. *Psychological Science*, 3, 340–345.
- Elliott, L. L. (1962). Backward and forward masking of probe tones of different frequencies. *Journal of the Acoustical Society of America*, 34, 1116–1117.
- Fant, C. G. M. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos*, 5, 3–17.
- Fant, G. (1966). A note on vocal tract size factors and nonuniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 4, 22–30.
- Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory & Language*, 49, 396–413.
- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory & Language*, 26, 489–504.
- Furui, S., Itakura, F., & Sato, S. (1972). Talker recognition by longtime averaged speech spectrum. *Electronics and Communications in Japan*, 55-A (10), 54–61.

- Galantucci, B., Fowler, C. A., & Turvey, M. T. (in press). The Motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 00, 000–000.
- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53, 181–215.
- Gaygen, D. E., & Luce, P. A. (1998). Effects of modality on subjective frequency estimates and processing of spoken and printed words. *Perceptions & Psychophysics*, 60, 465–483.
- Gedda, L., Bianchi, A., & Bianchi-Neroni, L. (1955). La voce dei gemelli: I. Prova di identificazione intrageminale della voce in 104 coppie (58 MZ e 46 DZ). [The voice of twins: A test of intratwin identification of the voice in 104 pairs (58 MZ and 46 DZ).] *Acta Geneticae Medicae et Gemellologiae*, 4, 121–130.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87–109.
- Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. London: Cambridge University Press.
- Giles, H., Scherer, K. R., & Taylor, D. M. (1979). Speech markers in social interaction. In: K. R. Scherer, & H. Giles (Eds), *Social markers in speech* (pp. 343–375). Cambridge: Cambridge University Press.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 1–16.
- Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1211–1238.
- Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In: A. S. Meyer, & N. O. Schiller (Eds), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Berlin: Mouton de Gruyter.
- Gould, J., Lane, H., Vick, J. C., Perkell, J. S., Matthies, M. L., & Zandipour, M. (2001). Changes in speech intelligibility of postlingually deaf adults after cochlear implantation. *Ear & Hearing*, 22, 453–460.
- Gregory, D., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversational partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, 21, 23–43.

Gregory, D., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status predictions. *Journal of Personality & Social Psychology*, *70*, 1231–1240.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior*, *14*, 237–251.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*, 423–445.

Halle, M. (1985). Speculations about the representation of words in memory. In: V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 101–114). New York: Academic Press.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405.

Hirsh, I. J. (1988). Auditory perception and speech. In: R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds), *Stevens' handbook of experimental psychology, volume I: Perception and motivation* (pp. 377–408). New York: Wiley.

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*, 598–613.

Hollien, H., & Klepper, B. (1984). The speaker identification problem. In: R. E. Rieber (Ed.), *Advances in forensic psychology and psychiatry* (Vol. 1, pp. 87–111). Norwood, NJ: Ablex.

Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics*, *10*, 139–148.

Holst, E. von, & Mittelstaedt, H. (1950). Das reafferenzprinzip. Wechselwirkung zwischen zentralnervensystem und peripherie. [The reafference principle. Interaction between the central nervous system and the periphery.] *Naturwissenschaften*, *37*, 464–476.

Honda, K. (1996). Organization of tongue articulation for vowels. *Journal of Phonetics*, *24*, 39–52.

Houde, J. F., & Jordan M. I. (1998). Sensorimotor adaptation in speech perception. *Science*, *279*, 1213–1216.

House, A. S., & Stevens, K. N. (1999). A longitudinal study of speech production, I: General findings. *Speech Communication Group Working Papers*, *XI*, 21–41.

Howell, P., & Darwin, C. J. (1977). Some properties of auditory memory for rapid formant transitions. *Memory & Cognition*, *5*, 700–708.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*, 553–562.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton & Co. Printers.

- Johnson, K. (2005). Speaker normalization in speech perception. In: D. B. Pisoni, & R. E. Remez (Eds), *The handbook of speech perception* (pp. 363–389). Malden, MA: Blackwell.
- Johnson, K., & Azara, M. (2000). The perception of personal identity in speech: Evidence from the perception of twins' speech. Unpublished manuscript.
- Johnson, K., & Mullenix, J. W. (1997). *Talker variability in speech processing*. San Diego: Academic Press.
- Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, 16, 337–342.
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of *F0* production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, 108, 1246–1251.
- Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America*, 113, 532–543.
- Joos, M. (1948). Acoustic phonetics. *Language*, 24 (supplement), 1–137.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kent, R. D., & Chial, M. R. (2002). The scientific basis of expert testimony on talker identification. In: D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds), *Modern scientific evidence* (Vol. 2, pp. 590–630). Eagan, MN : West.
- Kersta, L. G. (1962). Voiceprint identification. *Nature*, 196, 1253–1257.
- Klatt, D. H. (1989). Review of selected models of speech perception. In: W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.
- Kohler, W. (1910). Akustische Untersuchungen, II. [Acoustic investigations]. *Zeitschrift fur Psychologie mit Zeitschrift fur Angewandte Psychologie und Charakterkunde*, 58, 59–140.
- Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, 38, 618–625.
- Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1, 113–114.
- Kreiman, J. (1997). Listening to voices: Theory and practice in voice perception research. In: K. Johnson, & J. R. Mullenix (Eds), *Talker variability in speech* (pp. 85–108). New York: Academic Press.
- Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In: D. B. Pisoni, & R. E. Remez (Eds), *Handbook of speech perception* (pp. 228–362). Oxford: Blackwell.
- Kuhl, P. K. (1991). Human adult and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–107.



- Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In: W. J. Hardcastle, & N. Hewitt (Eds), *Coarticulation: Theory, data and techniques* (pp. 7–30). Cambridge: Cambridge University Press.
- Labov, W. (1966). *The social stratification of english in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1974). Linguistic change as a form of communication. In: A. Silverstein (Ed.), *Human communication: Theoretical explorations* (pp. 221–256). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Labov, W. (1986). Sources of inherent variation in the speech process. In: J. S. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 402–425). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lachs, L., & Pisoni, D. B. (2004). Cross-modal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, 30, 378–396.
- Lane, H. L., Catania, A. C., & Stevens, S. S. (1961). Voice level: Autophonic scale, perceived loudness and effects of sidetone. *Journal of the Acoustical Society of America*, 33, 160–167.
- Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14, 677–709.
- Lashley, K. S. (1951). The problem of serial order in behavior. In: L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112–136). New York: Wiley
- Laver, J., (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117–123.
- Liberman, A. M. (1996). Introduction: Some assumptions about speech and how they changed. In: A. M. Liberman (Ed.), *Speech: A special code* (pp. 1–44). Cambridge, MA: MIT Press.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 421–461.
- Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*, 65, 497–516.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience*, 15, 549–558.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172–187.
- Liénard, J. -S., & Di Benedetto, M. -G. (1999). Effect of vocal effort on spectral properties of vowels. *Journal of the Acoustical Society of America*, 106, 411–422.

- Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America*, 99, 1683–1692.
- Lombard, E. (1911). Le signe de l'élevation de la voix. [The sign of a rising voice.] *Annales des Maladies de L'oreille et du Larynx*, 37, 101–119.
- Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62, 615–625.
- MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review*, 77, 182–196.
- Magnuson, J. S., & Nusbaum, H. C. (2004). Acoustic differences, listener expectations, and the perception of talker differences. Unpublished manuscript.
- Markel, J. D., & Davis, S. B. (1979). Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base. *IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-27*, 74–82.
- Massaro, D., & Stork, D. (1998). Speech recognition and sensory integration: A 240 year old theorem helps explain how people and machines can integrate auditory and visual information to understand speech. *American Scientist*, 86, 236–244.
- McFarland, D. H., Baum, S. R., & Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *Journal of the Acoustical Society of America*, 100, 1093–11104.
- McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology*, 17, 249–271.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29, 539–553.
- Miller, G. A. (1951). *Language and communication*. New York: McGraw-Hill.
- Miller, G. A. (1965). Some preliminaries to psycholinguistics. *American Psychologist*, 20, 15–20.
- Miller, J. L., & Volaitis, L. E. (1989). Effects of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505–512.
- Modell, J. D., & Rich, G. J. (1915). A preliminary study of vowel qualities. *American Journal of Psychology*, 26, 453–456.
- Morsella, E., Romero-Canyas, R., Halim, J., & Krauss, R. M. (2002). Judging social identity from voice and photograph. Poster presented at meetings of Society for Personality and Social Psychology.
- Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, 58, 351–362.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality & Social Psychology*, 32, 790–804.

Navratil, J., Kleindienst, J., & Maes, S. (2000). An instantiable speech biometrics module with natural language interface: Implementation in the telephony environment. *Proceedings of the IEEE ICASSP-2000*, Turkey, June, 2000.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Nolan, F., & Oh, T. (1996). Identical twins, different voices. *Forensic Linguistics*, 3, 39–49.

Nygaard, L., Sommers, M., & Pisoni, D. (1992). Effects of speaking rate and talker variability on the representation of spoken words in memory. In: J. Ohala (Ed.), *Proceedings of the international conference on spoken language processing*, Edmonton, Alberta, University of Alberta Press (pp. 591–594).

Nygaard, L., Sommers, M., & Pisoni, D. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42–46.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382–2393.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.

Picheny, M. (2004). Achieving superhuman speech recognition: Via heredity or environment? *Minutes of the Columbia University Seminar on Language & Cognition*, January 29, 2004.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing, In: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28, 96–103.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, 27, 169–226.

Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In: K. Johnson, & J. W. Mullenix (Eds), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15, 285–290.

Remez, R. E. (2001). The interplay of phonology and perception considered from the perspective of organization. In: E. V. Hume, & K. A. Johnson (Eds), *The role of speech perception phenomena in phonology* (pp. 27–52). New York: Academic Press.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1996). Phonetic sensitivity and individual recognition: Notes on system architecture. Paper presented at the symposium on speaker characteristics: Effects of gestural variability on speech perception. Third Joint Meeting, Acoustical Societies of America and Japan, Honolulu, HI, December 2.

- Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651–666.
- Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sinewave analogues of speech. *Psychological Science*, 12, 24–29.
- Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, 101, 129–156.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947–950.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. In: M. R. Rosenzweig, & L. W. Porter (Eds), *Annual review of psychology*, (Vol. 39, pp. 475–543). Palo Alto: Annual Reviews.
- Rizzolati, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Rosen, S. M., Fourcin, A. J., & Moore, B. C. J. (1981). Voice pitch as an aid to lipreading. *Nature*, 291, 150–152.
- Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In: D. B. Pisoni, & R. E. Remez (Eds), *The handbook of speech perception* (pp. 51–78). Oxford: Blackwell.
- Rosenblum, L. D., & Gordon, M. S. (2001). The generality of specificity: Some lessons from audiovisual speech. *Behavioral and Brain Sciences*, 24, 239–240.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, 31, 307–314.
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421–436.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- Shannon, R. V., Zeng, F. -G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.
- Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication accommodation theory. In: W. P. Robinson, & H. Giles (Eds), *The new handbook of language & social psychology* (pp. 33–56). New York: Wiley.
- Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416, 87–90.

Stevens, K. N. (1990). Lexical access from features. A paper presented at the workshop on speech technology for man-machine interaction, Bombay, India.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In: P. D. Eimas, & J. L. Miller (Eds), *Perspectives on the study of speech* (pp. 1–38). Hillsdale: Lawrence Erlbaum Associates.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In: M. H. Christiansen, & S. Kirby (Eds), *Language evolution: The states of the art* (pp. 235–254). Oxford: Oxford University Press.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science*, 11, 153–158.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, 27, 24–27.

Uchanski, R. M. (2005). Clear speech. In: D. B. Pisoni, & R. E. Remez (Eds), *The handbook of speech perception* (pp. 207–235). Oxford: Blackwell.

Vallabha, G. K., & Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America*, 116, 1184–1197.

Vick, J. C., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., & Zandipour, M. (2001). Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing. *Journal of Speech, Language and Hearing Research*, 44, 1257–1267.

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception & Performance*, 29, 333–342.

Watson, C. S., Qiu, W. W., Chamberlain, M. M., & Li, S. (1996). Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. *Journal of the Acoustical Society of America*, 100, 1153–1162.

Weijer, J. Van de. (1997). Language input to a prelingual infant. In: A. Sorace, C. Heycock, & R. Shillcock (Eds). *Proceedings of the GALA '97 Conference on Language Acquisition* (pp. 290–293). Edinburgh, Scotland: GALA.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory & Language*, 31, 183–194.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, 52, 1238–1250.

## Chapter 8

# Spoken Word Recognition

*Delphine Dahan and James S. Magnuson*

### 1. INTRODUCTION

We solve an astounding array of information-processing challenges when we perceive a speaker's intended message. Apparently effortlessly, we accommodate variability in talker characteristics, dialect, speaking rate, and acoustic environment, all of which perturb the mapping between speech and linguistic categories. Without the aid of invariant cues to phonetic categories or word boundaries, we map acoustics onto phonetic categories, phonetic categories onto words in memory, words onto phrases and syntactic structures, words and syntax onto semantics, etc. Or do we?

On this view of language understanding, spoken word recognition is a distinct subsystem providing the interface between low-level perception and cognitive processes of retrieval, parsing, and interpretation. The narrowest conception of the process of recognizing a spoken word is that it starts from a string of phonemes, establishes how these phonemes should be grouped to form words, and passes these words onto the next level of processing. Some theories, though, take a broader view and blur the distinctions between speech perception, spoken word recognition, and sentence processing (e.g., Elman, 2004; Gaskell & Marslen-Wilson, 1997; Klatt, 1979; McClelland, St John, & Taraban, 1989).

What motivates the narrow and broad conceptions? There are empirical, pragmatic, and theoretical motivations for the narrow view. Empirically, psycholinguistic levels of processing map roughly onto linguistic levels of description. The fact that linguistic knowledge can be described as a hierarchically structured set of levels leads to the reasonable hypothesis that speakers (or signers) and perceivers may represent and operate on those structures. Indeed, this hypothesis is given face validity by the fact that humans can make decisions about levels like phonemes and words and that perception can be influenced by manipulations at those levels (though there is a long history of debate over their psychological reality; see Pisoni & Luce, 1987, for a review).

The pragmatic motivation for the narrow view stems from the fact that over a century of concerted study of speech perception has led to a catalog of complex empirical phenomena and candidate cues for speech perception, but little understanding of the specific components of the speech signal that humans use to decode speech and achieve phonetic constancy (Nusbaum & Magnuson, 1997; Remez, 2005). Rather than wait for a complete understanding of early perceptual processes, psycholinguists have made significant progress in understanding the processing of words and sentences by making the simplifying assumption that a string of phonemes makes a reasonable proxy for the results of initial perception, and that a series of sound forms associated with lexical entries makes a reasonable proxy for the input to sentence processing.

Theoretically, the narrow view is motivated in part by the assumption that the division of labor in staged systems affords significant processing efficiencies (Fodor, 1983; Norris, 1994; Norris, McQueen, & Cutler, 2000). Breaking the problem into distinct stages is argued to provide cognitive economy if the result is a series of mappings that are straightforward relative to the complexity of the full mapping from lowest to highest level (restrictions on the information available to each level are also key to the *interaction vs. autonomy* debate discussed below).

The broader view of spoken word recognition (in the extreme, as the mapping from speech to meaningful units that may be larger than words) has empirical and theoretical motivations. One consideration is that by assuming that the input to spoken word recognition is a string of abstract, phonemic category labels, one implicitly assumes that the nonphonemic variability carried on the speech signal is not relevant for spoken word recognition and higher levels of processing. However, if this variability and detail is not random but is lawfully related (even partially) to linguistic categories, the simplifying assumption that the output of speech perception is a string of phonemes may actually be a complicating assumption. Indeed, there is growing evidence that spoken word recognition is influenced by information in the signal that cannot be captured in a string of phonemes. For example, misleading coarticulatory cues caused by splicing the onset and most of the vowel of one consonant–vowel–consonant (CVC) word or nonword onto the last consonant of another CVC word (creating “subcategorical mismatches”; Whalen, 1984) changes the time course of lexical activation and competition (Dahan, Magnuson, Tanenhaus, & Hogan, 2001a; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999).

What purpose might this fine-grained sensitivity serve? One challenge posed by assuming that words are identified from a string of phonemes is the *embedding problem*; most long words have multiple shorter words embedded within their phonemic transcriptions (e.g., depending on dialect, and neglecting all subphonemic cues, *unitary* contains *you*, *unit*, *knit*, *it*, *tarry*, *air*, and *airy*) and conversely, many short words embed in one or more other words (McQueen, Cutler, Briscoe, & Norris, 1995). Successful spoken word recognition depends on distinguishing intended words from embeddings. However, the embedding problem is significantly mitigated when subphonemic information in the input is considered. For example, listeners are sensitive to very subtle

durational differences (in the range of 15–20 ms) that distinguish *phonemically identical* syllables that occur in short words (*ham*) from those embedded in longer words (*hamster*) (Salverda, Dahan, & McQueen, 2003; see also Davis, Marslen-Wilson, & Gaskell, 2002).

Thus, the bottom-up signal contains vital information that simplifies the mapping from speech to words that would be lost were words identified from a string of phonemes. Might the same be true for subsequent processes? There is increasing evidence that the construction of syntactic and semantic structures relies on more than just a sequence of words. Indeed, a sequence of words is almost always temporarily compatible with multiple structures. For example, the structure associated with the word sequence *John knew the answer* differs whether it is followed by *was wrong* or *to the question* (e.g., Altmann, 1999). A growing body of work has documented the role played by the prosodic structure of an utterance (marked by prosodic breaks and intonational prominences) in favoring some structures over others (e.g., Kjelgaard & Speer, 1999; for a review see Cutler, Dahan, & van Donselaar, 1997; Speer & Blodgett, 2006, this volume). This literature indicates that information from the speech signal is passed onto higher levels of processing. This supports an integrated view of phonetic, lexical, and sentential processing.

Sentence-level top-down constraints on lexical activation have received some attention in spoken word recognition, but chiefly with respect to how top-down information might constrain the set of activated lexical items (e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987, 1990). Immediate access to syntactic, semantic, and nonlinguistic context could provide significant constraints on spoken word recognition, by influencing the activation of homophones, semantic associates, or context-appropriate lexical items (Shillcock & Bard, 1993), helping resolve lexical ambiguity resulting from phonological assimilations (Gaskell & Marslen-Wilson, 2001), or by restricting the set of possible referents (Brown-Schmidt, Campana, & Tanenhaus, 2004).

Throughout this chapter, as we describe the central themes of current research on the recognition of spoken words, we will adopt the more prevalent, narrow view (except when noted) that most current work assumes. As we will discuss at the end of the chapter, the growing evidence for subcategorical specificity may herald a dramatic shift in theories of spoken word recognition. Taking the broad view – confronting the speech signal itself and considering how higher levels of representation might constrain lexical access – may be the key to significant progress in understanding spoken word recognition.

The recognition of a spoken word can be viewed as the process of classifying an auditory stimulus as belonging to one “word-form” category, chosen from many alternatives. As this description stresses, this process requires matching the spoken input with mental representations associated with word candidates, and selecting one among several candidates that are at least partially consistent with the input. Frauenfelder and Tyler (1987) classified the functions required of any theory of spoken word recognition into three stages. *Initial contact* is how input interfaces with and activates lexical representations. *Selection* describes how the set of activated lexical alternatives is evaluated with respect to the sensory input. *Integration* refers to how candidates are evaluated with



respect to the linguistic and nonlinguistic context, in order to identify which is the likeliest candidate for recognition as well as to build larger linguistic structures.

Early models viewed these processes as discrete, only partially overlapping stages, in particular predicting a temporal delay between access and selection (e.g., Marslen-Wilson, 1987; Zwitserlood, 1989). More recent models allow for continuous uptake and evaluation of the input, thus blurring functional and temporal distinctions between access, selection, and integration (for behavioral evidence supporting the continuous view, see Dahan & Tanenhaus, 2004). Nonetheless, the theoretical distinctions are useful as each process poses challenges of a different nature.

In the course of this chapter, we will review the central issues pertaining to contact and selection in turn, and how they have been conceptualized in different models and theories of spoken word recognition. We will also address whether categorizing the input as a member of a word category changes listeners' percept of the input. This question hinges on the architecture of the processing system, i.e., whether higher levels of representations (such as words) can affect lower levels, such as speech sounds or phonemes. Finally, we will briefly review integration, and close with a discussion of what we see as the most crucial challenges to theories of spoken word recognition and spoken language generally, and the approaches we find most promising and most likely to lead to solutions.

## 2. INITIAL CONTACT

When someone speaks, the linguistic content and speaker characteristics (e.g., physiology of the vocal tract, gender, regional origin, emotions, identity) simultaneously influence the acoustics of the resulting spoken output. Additional sources of variability include rate of elocution, prosodic prominence, and the phonetic context in which each word is pronounced. Nonetheless, listeners are able to recognize acoustically different stimuli as instances of the same word, thus extracting the similarity that exists between these different tokens, and perceiving them as members of the same category. How are words mentally represented to allow for this complex categorization?

The traditional (and dominant) view assumes that people represent the form of words as categories that abstract away from variability. Drawing on linguistic theories, the mental representation of a word form is usually conceived as a sequence of phonemes (sometimes themselves decomposed into a bundle of contrastive features). Within this framework, the ease with which a given pronunciation is categorized as a token of a given word is assumed to depend upon the degree to which its components have characteristics typically associated with the word's phonemes. Speaker-specific information is often viewed as a source of noise which does not contribute to the process of identifying the linguistic units present in the signal.

This view has not gone uncontested. An episodic view, most forcefully argued for by Goldinger (1996, 1998), conceptualizes lexical representations as ensembles of detailed

memory traces (or episodes) of word instances. Several recognition memory studies have shown that people implicitly retain in memory nonlinguistic aspects of spoken words (e.g., Hintzman, Block, & Inskip, 1972; Goldinger, 1996, 1998; Palmeri, Goldinger, & Pisoni, 1993). The question at stake is whether these memory traces of words constitute the knowledge that people access and use when identifying spoken words. Goldinger (1998) applied (Hintzman's) (1986) MINERVA2 model of episodic memory to spoken word recognition. In this model, a speech episode (a word) is simultaneously compared to all memory traces. Activation of a trace is proportional to its acoustic similarity with the stimulus. The aggregate of all activated traces (the so-called *echo*) is sent to working memory and corresponds to the listener's percept. Because the echo consists of a blend of the memory traces that resemble the stimulus, it tends to capture the aspects that are common among the traces but not the aspects that differ. This principle enables the model to make generalizations and categorize new tokens without assuming the existence of abstract mental categories. A critical challenge to the episodic view is how the similarity between an actual speech stimulus and memory traces would be computed, if no normalization or other data-reducing process abstracting from surface variability is assumed. Goldinger's model has thus far assumed word-length episodes and remains agnostic about how words would be isolated from the utterances they are embedded in, which is problematic given the challenges posed by word segmentation (see below). Given its radical departure from classical approaches, this theory may well have the potential to bring new leverage to problems of speech perception and spoken word recognition. However, until similarity mapping and segmentation are spelled out, the episodic view faces the same challenges as the traditional, abstract view.<sup>1</sup>

The traditional view has influenced much of the research on spoken word recognition. Thus, the recognition of a spoken word is generally viewed as the mapping of the speech input onto abstract lexical representations, with abstract units standing for the word's subcomponents, the phonemes, mediating this mapping. An extended line of research has documented how listeners accommodate the variability inherent to speech rate and phonetic context in the perception and recognition of individual phonemes (Miller & Liberman, 1979). We will not review this literature here, but rather will focus on how theories of spoken word recognition have embodied, or sometimes departed from, the classical approach to spoken word recognition.

---

<sup>1</sup> Goldinger's (1998) simulations have two critical problems. The model assumes the input is pre-segmented into word-length episodes (the primitive unit), which are represented as vectors of units (with values of  $-1$ ,  $0$ , or  $+1$ ), with some units representing the word type, and others representing talker and context information. While Goldinger claimed such a model can achieve phonetic constancy without normalization of talker differences, the solution depends on this unrealistic assumption about the input. In real speech, talker variability conditions phonetic realization. In Goldinger's simulations, the input is in effect pre-normalized. The episodic model's promise to solve phonetic constancy without normalization may be possible, but tests with more realistic input are needed to evaluate it (see Goldinger & Azuma, 2003, for a discussion of how adaptive resonance theory may provide the means to test the theory with more realistic inputs and mechanisms).

## 2.1. Initial Contact and Similarity Metrics

The first question for any model is the nature of the input representation: How do the products of sensory information interface with the lexicon? As mentioned earlier, the input to word recognition has traditionally been assumed to be a string of phonemes, output by a speech perception system (as in the original COHORT model; Marslen-Wilson & Welsh, 1978). This representation was also adopted by the model SHORTLIST (Norris, 1994), although mainly for practical reasons. However, the string-of-phonemes encoding of the speech input assumes that subphonemic variation in the signal is lost, while such variation has been shown to affect listeners' word recognition. For example, Andruski, Blumstein, and Burton (1994) demonstrated that, as the realization of the initial segment of a word like *king* is modified as to differ from a prototypical /k/ (by shortening the duration of the stop voice onset time) but not enough to change the ultimate categorization of this segment, people are nonetheless less likely to categorize the word as an instance of *king* (see McMurray, Tanenhaus, & Aslin, 2002, for converging evidence; for demonstrations of listeners' sensitivity to subcategorical cues in vowels, see (Dahan et al., 2001b); Marslen-Wilson & Warren, 1994; Whalen, 1984). Evidence for graded activation of words based on subphonemic similarity requires a finer representational grain than phonemes.

Another issue related to the string-of-phonemes assumption is that it imposes a dissociation between the process of recognizing words from that of recognizing its components. A recent attempt to add an automatic phone recognizer to the SHORTLIST model exposed the limitations of this assumption (Scharenborg, ten Bosch, Boves, & Norris, 2003). In these simulations, the automatic phone recognizer took real speech (naturalistic speech samples from telephone conversations) as input and generated a sequence of phonemes. From this string of phonemes, the activation and competition mechanisms implemented in the SHORTLIST model yielded the best matching word candidate. Word recognition accuracy performance was poor (around 25%), but improved considerably when one of the model's parameters, the penalty assigned to candidates that mismatch the phonemic input, was set to zero. This result may be interpreted as evidence that SHORTLIST, originally tested on an unrealistically accurate phonemic input, must be revised to accommodate likely erroneous input from a phone recognizer. On the other hand, this result can be taken as reflecting the shortcoming of a phonemic string as input. If "hard" phonemic decisions are made by the phone recognizer, the fact that other phonemic interpretations were substantially supported by the signal is lost. Most of all, these simulations illustrate how much the modeling of spoken word recognition hinges on assumptions about the input representation.

The next simplest solution is to assume that the input takes the form of localist phoneme activation units (as in the REVISED COHORT model (Marslen-Wilson, 1987, 1989), and MERGE (Norris et al., 2000)). Subphonemic detail can be approximated in the distributed representation afforded by the entire set of phonemes. Thus, a segment ambiguous between /k/ and /g/ can be represented by partial activation of both units. A slightly more fine-grained representation can be achieved with units representing a set of (usually binary) acoustic-phonetic features (as in the DISTRIBUTED COHORT MODEL

(Gaskell & Marslen-Wilson, 1997)). However, the realism of these schemes is limited, as they fail to incorporate a critical aspect of speech, coarticulation.

As demonstrated by the seminal work by Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) and contrary to a listener's subjective impression, a spoken utterance is not a concatenated sequence of discrete speech sounds. The gestures involved in the articulation of each sound overlap temporally with the gestures that generate adjacent sounds. One of the consequences of this temporal overlap has been coined the "segmentation" problem. A spoken utterance cannot be divided into smaller portions, each one representing a single segment.<sup>2</sup> If the recognition of a spoken word involves the mapping of the input onto word representations where segments are in temporal order, the listener must assign the presence of a given acoustic feature in the input to a given segment.

The TRACE model (McClelland & Elman, 1986) uses the most complex input representations of any current model of speech perception and spoken word recognition. The input is a "pseudo-spectral" representation based on seven acoustic-phonetic features, each represented with a nine-unit continuous vector, which encode the degree to which the feature is represented in the input. Features spread over time by ramping up to a phoneme center and then ramping off. Phoneme centers are close enough together and features spread far enough that there is substantial overlap between phonemes, creating a rough analog to coarticulation.

TRACE's architecture is also a critical aspect in the way it accounts for the processing of coarticulated speech. In TRACE, units that stand for hypotheses at the featural, phonemic, or word level, are replicated every three time slices. Each unit stands for a linguistic unit potentially present in the input at a different point in time. The extensive unit reduplication has often been criticized as an implausible feature of the model (beginning with McClelland & Elman, 1986). However, this is central to solving the segmentation issue, as it accommodates the fact that features that result from the overlap of articulatory gestures coincide in time. A given time slice may provide evidence supporting different phonemes, thus activating several incompatible phoneme units. However, within each level, units that span the same portion of the input inhibit each other. Consequently, the phoneme hypothesis for which the evidence is the strongest can win the competition. Thus, TRACE's architecture allows the segmentation of coarticulated speech into a sequence of discrete segments.

However, TRACE's input scheme provides a very rough approximation of coarticulation in real speech. While it accommodates the temporal overlap of gestures, it fails to accommodate the fact that this temporal overlap affects the articulatory (and therefore acoustic) realization of segments (i.e., the "lack of invariance" issue, cf. Liberman et al., 1967). There have been very few even moderately successful attempts to devise

---

<sup>2</sup> This is true despite the fact that the duration of single segments, such as consonants or vowels, are often reported. Such segmentation is based on conventions on how to define boundaries between segments based on their relative prominence (see Fowler, 1984).

psychologically tractable models that work directly on the actual speech signal or a minimally transformed speech signal. (The hidden-Markov models and similar mechanisms used in automatic speech recognition systems arguably substitute the black box of the brain with largely opaque statistical approximations; see Nusbaum & Magnuson, 1997, for discussion.) Klatt's (1979) LEXICAL ACCESS FROM SPECTRA (LAFS) model is perhaps the best known, but the mapping from spectra to lexical items is at least as variable as the mapping from speech to phonemes. Work in the adaptive resonance framework has grappled with real speech signals (the ARTSTREAM model; Grossberg, Govindarajan, Wyse, & Cohen, 2004) but has yet to be extended to the recognition of phonemic or lexical forms. The strategy of Plaut and Kello (1999) may well be the best hope for progress toward more realistic input. They use a collection of articulatory and acoustic cues that *might* turn out to be tractable to extract from speech (auditory and visual cues to jaw movements, changes in formants, etc.), and in combination, *might* prove a sufficient basis for speech perception and spoken word recognition.

## 2.2. Initial Constraints on Activation

Theories differ on the patterns of activation that follow initial contact. More specifically, they differ in the theories of similarity they assume. The ORIGINAL (Marslen-Wilson & Welsh, 1978), REVISED (Marslen-Wilson, 1987, 1989) and DISTRIBUTED COHORT models (Gaskell & Marslen-Wilson, 1997, 1999, 2002) place great emphasis on word onsets. The real-time constraints of the speech signal motivate an emphasis on optimal use of bottom-up information as it becomes available. Since a word's onset is heard first, it should determine which lexical items are first activated. Thus, in the original COHORT model, the set of activated lexical alternatives was constrained to a *word-initial cohort* of items that matched perfectly the phonemic representation of the first approximately 150 ms of a word's onset. In light of evidence that a word might be recognized even when its first sounds are altered (for example, due to mispronunciation, cf. Cole, 1973), the revised and DISTRIBUTED COHORT models abandon the strict, all-or-none match constraint. Instead, lexical representations are activated as a function of their similarity to a spoken word, with this similarity being continuously evaluated rather than limited to the initial portion of the spoken word. Nonetheless, the models' emphasis on real-time processing maintains a special status to the spoken word's initial sounds, as they contribute to the activation of some words, and thereby the interpretation of subsequent spoken material will be biased in favor of these words (see the discussion of *Selection* below for a full description of how these biases might be implemented).

The NEIGHBORHOOD ACTIVATION MODEL (NAM; Luce, 1986; Luce, Pisoni, & Goldinger, 1990; Luce & Pisoni, 1998) differs from any instantiation of the COHORT model by predicting activation of words that reflects their global similarity with the spoken word.<sup>3</sup>

---

<sup>3</sup> NAM is not a processing model *per se* – it is more properly considered a formal similarity model. However, its similarity metric imposes significant constraints on an underlying processing mechanism, and as such, it is appropriate to consider what NAM predicts in terms of lexical activation.

Two similarity metrics were developed within the model. The more complex one is derived from observed similarity measures, such as position-specific diphone confusion probabilities. Similarity between the spoken word and other words is computed by comparing the confusability of each of its segments with other words' segments in the same position within the word. Similarity is gradient, although limited to words that have the same number of segments. The simpler metric, sometimes called the *one-phoneme shortcut metric*, distinguishes words that are predicted to become activated during the perception of a spoken word (i.e., its neighbors) from those that are not, with no gradiency in the degree of activation of the former. Activated words (i.e., neighbors of the spoken words) are defined as words that differ from the spoken word by no more than one phoneme, whether by substitution, deletion, or addition, in any position. Thus, neighbors of *cat* include *bat*, *kit* and *cap* (substitutions), *at* (deletion), and *scat* and *cast* (additions).

The COHORT and NEIGHBORHOOD models make different predictions about what items may be activated by a spoken word. The COHORT model predicts that hearing *cat* also activates *castle* but should activate *bat* to a negligible degree. NAM predicts that *cat* will activate *bat* but not *castle*, as it differs by too many phonemes. There is empirical support for each prediction. Marslen-Wilson (1993) reported a series of studies in which the auditory presentation of a word primes visual lexical decisions to semantic associates of words overlapping in onset, but not in rhyme (e.g., *beaker* would prime *insect*, an associate of *beetle*, but not *stereo*, an associate of *speaker*). But Luce and Pisoni (1998) reported that neighborhoods based on global similarity provide the best prediction of processing time for large sets of words in tasks like lexical decision and naming, although they did not separate out the contribution of the cohort-type neighbors from that of non-cohort ones (we discuss this result further in the *selection* section).

TRACE makes an intermediate prediction: It activates both onset- and rhyme-overlapping words, because, as in the NEIGHBORHOOD model, words can be activated even if they mismatch at onset. However, unlike the NEIGHBORHOOD model, TRACE represents time: Words that become activated early in the spoken input have an advantage over words that become activated later, because more of the spoken word has been heard and selection mechanisms are then more effective at favoring the best matching candidate. Thus, TRACE predicts activation of both onset- and rhyme-overlapping candidates, although at different times and of different amplitude. Allopenna, Magnuson, and Tanenhaus (1998) provided behavioral data supporting this prediction. They estimated lexical activation to word candidates by monitoring eye movements to pictures as participants followed verbal instructions to move an item on a computer screen. Fixations were closely time-locked to the speech (with a lag only slightly larger than that required to plan and launch an eye movement), and mapped closely onto phonetic similarity over time (with higher and earlier fixation proportions to onset-overlapping competitor than rhyme-overlapping competitor) as well as response probabilities generated by TRACE. This study highlights the importance of a measure of lexical activation over time, given the rapid evolution of lexical activation as the spoken input is heard.

The Allopenna et al. (1998) study highlights one shortcoming of the similarity model embodied in NAM to the study of spoken word recognition. The temporal distribution of similarity is not considered; *dab* and *bad* are assumed to be equally active upon hearing *dad* (ignoring frequency for the sake of the example). NAM fails to capture the temporal dimension of speech and the special status that the initial sounds have due to their temporal precedence (Marslen-Wilson & Zwitserlood, 1989). It also gives too much weight to the match in the number of segments or syllabic structure by entirely excluding the contribution of words that are more than one phoneme longer than the word to be recognized, despite evidence suggesting that words of different lengths affect the processing of a given word (Marslen-Wilson, 1984, 1987). The algorithm cannot be easily extended to the computation of competition environment for polysyllabic words, as most of these words have very few, if any, competitors under the one-phoneme difference definition.<sup>4</sup> Finally, the one-phoneme shortcut metric, which has been most widely used by researchers and has proven useful in stimulus selection and experimental control, treats any phoneme deviation equally, regardless of its phonetic nature. Confusion between two words differing by one-phoneme addition or substitution, or confusion between two words differing by a vowel or a consonant, are all assumed to be equivalent, despite empirical evidence that the nature of the phonetic feature(s) that differ between two words is an important factor in accounting for word confusions (e.g., Bailey & Hahn, 2005; Hahn & Bailey, 2005; see also van Ooijen, 1996).<sup>5</sup>

### 2.3. Plasticity in Mapping the Speech Signal onto the Lexicon

As pointed out in the introduction to this section, the acoustic form that a given word takes can vary greatly. Nonetheless, listeners have little difficulty accommodating this variability, which has sometimes been interpreted as reflecting plasticity in the mapping of the speech signal onto the lexicon. Here we review some of this work.

A substantial number of studies have examined the processing of spoken words that have undergone phonological assimilation. In connected speech, the value of a segment's feature (i.e., place of articulation or voicing) may assimilate to that of the same feature from its surrounding segments. For instance, the place of articulation of the final alveolar sound of the word *green*, may be altered to become (or approach) the bilabial place of articulation of the initial sound of the subsequent word *boat*, so that the sequence may sound a little like *gream boat*. The conditions under which assimilation may occur are dictated by the phonology of the language. Research on the perception of assimilated words

---

<sup>4</sup> Cluff and Luce (1990) used the one-phoneme difference algorithm to compute the competition environment of bisyllabic words composed of two monosyllabic words (e.g., *jigsaw*) by establishing competitors for each syllable independently, thereby considering only monosyllabic competitors.

<sup>5</sup> Luce, Goldinger, Auer, and Vitevitch (2000) report examples of cases where a more complex metric, based on positional similarity ratings, makes distinctly different predictions than the one-phoneme shortcut metric, e.g., predicting competition between *veer* and *bull* due to high similarity at every segment despite no complete phoneme matches.

has shown that this deviation does not preclude the identification of the assimilated token as an instance of the intended word. Gaskell and colleagues (Gaskell, 2003; Gaskell, Hare, & Marslen-Wilson, 1995; Gaskell & Marslen-Wilson, 1996, 1998) have suggested that listeners have learned to accept the assimilated form as a token of the intended word in the appropriate context, especially if the assimilation was only partial (thus, maintaining some of the acoustic characteristics of the original segment). This proposal was supported by simulations from a connectionist model that was trained to learn to map acoustically variable (but arguably simplified) input onto canonical, fixed representation of words (see Gow, 2001, 2002, 2003a, 2003b, for a critique of Gaskell and colleagues' proposal and for a competing account of the perception of assimilated words).

Other rule-based variations do not involve a phonemic or subphonemic alteration. These pronunciations are characteristic of casual (as opposed to careful) speech, and often described as including an atypical or reduced realization of some of the segments of words. For example, the final consonant of the word *flute* can be realized with an alveolar closure and an audible release (the typical realization of the phoneme /t/), or realized as a glottal stop, with no release. Similarly, the vowel of the unstressed syllable of a polysyllabic word can be so drastically reduced that it is not acoustically present in the signal (e.g., *police*, pronounced roughly as [plis]). How can two fairly different realizations be interpreted as instances of the same word? Do people represent the multiple forms that a word can take in order to accommodate such variation? And if so, at what level of abstraction are they represented, and does frequency of occurrence of variants determine whether a variant is represented or not?

Recent work has addressed these questions by examining whether variations are equally effective at making contact with the lexical representation of the intended word (or, put slightly differently, whether variations are equally categorized as members of the intended word category). Some studies have probed the degree to which the meaning of the word becomes available (e.g., Deelman & Connine, 2001; Sumner & Samuel, 2005). Other studies have examined the degree to which variants map onto the same or different representations by assessing whether having heard one variant facilitates the subsequent processing of the alternative (e.g., LoCasto & Connine, 2002; McLennan, Luce, & Charles-Luce, 2003; Sumner & Samuel, 2005; Utman, Blumstein, & Burton, 2000). The findings that emerge from these studies are complex and often conflicting. Some results suggest that any variant facilitates the processing of any alternative, which is sometimes interpreted as evidence for a single, abstract and general representation; other results argue for specificity. Some researchers found evidence for a special status of the most frequent variant (Connine, 2004), while others did not (Sumner & Samuel, 2005). This line of research has only begun, and it is too early to draw definite conclusions. Nonetheless, one aspect that has been little considered is the relevance of the context in which these variants occur. Listeners may be sensitive to how likely and expected a given variation is, given what is known of the talker's speaking style, speed of elocution, and perhaps geographic or dialectal origin. Such expectations (or use of context) may determine the degree to which a token will be mapped onto or activate representation(s) associated with the intended word.



Indeed, we know that listeners do adapt to the characteristics of the talker or speech that they hear. Evidence of such adaptation comes from studies showing that word identification is impaired by trial-to-trial changes in the voice of the talker (Mullennix, Pisoni, & Martin, 1989; Nusbaum & Morin, 1992) and/or in his/her speaking rate (Sommers, Nygaard, & Pisoni, 1994), and from studies showing advantage for the identification of words spoken in a familiar vs. unfamiliar voice ((Nygaard), Sommers, & Pisoni, 1994; although see Luce & Lyons, 1998). This suggests plasticity in the process of perceiving and interpreting speech. Listeners' ability to adapt to the characteristics of the speech or the talker they are exposed to has long been acknowledged (e.g., Joos, 1948; Ladefoged & Broadbent, 1957; Peterson & Barney, 1952). More recently, a number of studies have documented how adaptation to distorted or foreign-accented speech proceeds. The process appears to operate quite rapidly, with measurable improvement in comprehension observed after as little as two to four sentences (Clarke & Garrett, 2004). Importantly, from relatively short exposure to distorted speech, people acquire knowledge that can generalize to sentences containing unheard words (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Greenspan, Nusbaum, & Pisoni, 1988), or to similarly distorted speech from a different talker (e.g., Dupoux & Green, 1997). Furthermore, listeners' perceptual adaptation to unusual speech or talker characteristics seems to be (at least largely) mediated by lexical knowledge. Listeners who were exposed to 20 distorted nonsense sentences prior to testing on sensible sentences fared no better than people with no prior exposure to distorted speech (Davis et al., 2005; for similar conclusions, see Eisner & McQueen, 2005; Norris, McQueen, & Cutler, 2003).

Evidence for plasticity in the mapping of spoken input onto lexical representations may help explain how listeners cope with the extreme variability found in speech. So long as this variability is context-dependent, and thus lawful, prior (even brief) exposure to speech from a new talker may trigger the learning of a new mapping between speech input and linguistic units.

#### **2.4. The Interaction Debate: Is the Interface Bidirectional?**

Is the architecture underlying spoken word recognition autonomous (feedforward only) or interactive (lexical representations feed information back over the interface to the sublexical representations)? Bottom-up and top-down information *is* integrated: The literature is full of examples of lexical effects on tasks that tap sublexical representations. Phonemes are detected more quickly in words than nonwords (the word superiority effect; Rubin, Turvey, & Van Gelder, 1976). Listeners report hearing phonemes consistent with lexical or sentential context in locations completely replaced with noise (the phoneme restoration effect; e.g., Warren, 1970; Samuel, 1981, 1997). If a phoneme continuum is attached to a context that makes one endpoint a word and the other a nonword (e.g., /t/-/d/ attached to *-ash* or *-ask*), categorical perception boundaries shift such that more steps are identified as consistent with the lexical endpoint (Ganong, 1980; a bias is also found in word-word contexts with a frequency differential; Fox, 1984). Helpful visual contexts are integrated quickly to resolve ambiguities in sentence processing

(Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The crux of the interaction debate is *when* integration occurs.

This debate has recently taken center stage in spoken word recognition research, having been energized by forceful empirical and theoretical arguments for autonomous models of spoken word recognition by Norris et al. (2000, 2003). In brief, the *autonomous view* is that processing stages can be optimized by allowing them access only to bottom-up information (disallowing interaction of top-down information). This view of stages within a processing system is related to arguments for modularity between processing systems (Fodor, 1983). In both cases, the idea is that veridical perception depends upon transparent processing of the incoming signal. On this view, if top-down information is integrated directly with sensory information, an organism *ipso facto* loses the possibility of veridical perception, as there is no distinction between information in the environment and information in the organism. Autonomous models account for lexical effects on sublexical tasks by proposing parallel, competing lexical and sublexical routes (as in the Race model; Cutler & Norris, 1979), or that the locus of sublexical decisions is, counter-intuitively, post-lexical. In the Merge model (Norris et al., 2000), for example, there are two banks of phoneme units. One is the source of bottom-up input to the lexical layer. The second receives input from the bottom-up phoneme nodes and the lexical nodes. This decision layer can thus integrate lexical and phonological knowledge without changing the prelexical interpretation of the sensory input. The separate bank of decision nodes is justified on the grounds that phonemic awareness is a late-developing artifact of learning to read, based on evidence that phonemic awareness does not develop if one does not learn to read (see Norris et al., 2000; Marslen-Wilson & Warren, 1994, for discussion; but there is evidence that sublexical awareness (if not precisely phonemic) does emerge in preliterate children (see Liberman, Shankweiler, Fischer, & Carter, 1974) and illiterate adults (Bertelson & de Gelder, 1989); see Shankweiler & Fowler, 2004, for a review). Falsifying this position would require showing that top-down lexical effects have a perceptual, rather than decisional, locus.

On the *interactive view*, if top-down information *can* usefully constrain interpretation of bottom-up information, it should be used, and veridical perception can be maintained by properly weighting bottom-up and top-down information. Falsifying this position is more difficult. Alternative explanations for lexical effects must be proposed, and evidence must show that when those explanations make predictions that are different from lexical feedback predictions, the lexical feedback predictions are incorrect. Over the past two decades, the debate appeared to be settled at least two or three times, with alternative apparent falsifications of autonomous and interactive positions.

Elman and McClelland (1988) seemingly falsified the autonomous position, by showing lexical effects on sublexical processing rather than sublexical decisions. They conducted a study designed to demonstrate *lexically mediated compensation for coarticulation*. Compensation for coarticulation (Mann & Repp, 1981) refers to the fact that in normal production, if a segment with a front place of articulation follows one further back (or vice versa), physical and temporal constraints may prevent the articulation from

reaching its ideal location, with the result that in this context, the front segment will have a place of articulation further back than normal. When a front-back continuum (e.g., /t/-/k/) is presented following a back segment (e.g., /ʃ/) the category boundary shifts toward the back (i.e., more steps on the continuum are identified as the front segment /t/), and the opposite happens after a front segment (e.g., /s/). In Elman and McClelland's (1988) study, this low-level perceptual phenomenon was coupled with the Ganong (1980) effect. The Ganong effect shows that the interpretation of an ambiguous sound (symbolized by ?, intermediate between, e.g., *p* and *b*) embedded in a larger spoken stimulus (e.g., ?*ea*ce) is biased toward the interpretation that turns the spoken stimulus into a real word (e.g., *pe*ace). Elman and McClelland (1988) reasoned that if the basis for the Ganong effect is feedback to the perceptual level, a restored phoneme in that paradigm should have similar *consequences* as an intact phoneme, and in particular, it should drive compensation for coarticulation. They found exactly that result: the boundary of a *tapes-capes* continuum shifted following a segment ambiguous between /s/ and /ʃ/ as a function of the lexical bias preceding the ambiguous segment (e.g., *Christma-* or *fooli-*). For the next decade, many regarded this as strong evidence in support of interaction.

However, Pitt and McQueen (1998) explored the hypothesis that the basis for the effect was diphone transitional probabilities (TPs), based on an analysis by Cairns, Shillcock, Chater, and Levy (1995), purportedly showing that Elman and McClelland's lexical contexts were confounded with TP. Under the TP hypothesis, compensation for coarticulation after a segment ambiguous between /s/ and /ʃ/ is driven by the higher probability of /s/ after the final vowel of *Christma*, /ə/, than after the final vowel of *fooli*, /i/, and, conversely, the higher probability of /ʃ/ after /i/ than after /ə/. Because these transitional probabilities can be viewed as involving sublexical knowledge only, Elman and McClelland's (1986) results would not be proof of lexical influence on sublexical processing. Pitt and McQueen directly tested this hypothesis and found compensation for coarticulation with nonword contexts as a function of TP, but failed to find it in lexical contexts where TP was controlled. For the next several years, this was regarded by many as strong evidence that TP was the basis for "lexically" mediated compensation for coarticulation.

Samuel and Pitt (2003) provided a thorough empirical and acoustic analysis of the paradigm. They reported new studies in which they found lexically mediated compensation for coarticulation with several contexts with opposite lexical and diphone TP biases. They also provided plausible perceptual explanations for the minority of cases where lexically mediated compensation for coarticulation has not been found (e.g., Pitt & McQueen, 1998; and some contexts tested by Samuel and Pitt themselves). Magnuson, McMurray, Tanenhaus, and Aslin (2003a) reported converging evidence as well as a new corpus analysis of transitional probabilities in American English that revealed that not all of Elman and McClelland's lexical contexts were confounded with diphone TP. They also used corpus analyses to show that no particular *n*-phone TP could predict observed lexical effects. Instead, the appropriate TP context seems to be an *n*-phone of dynamic length, where *n* resolves to word length, and thus the knowledge driving mediated compensation for coarticulation seems to be lexical.

Further evidence for feedback comes from *selective adaptation to restored phonemes*. Samuel (1997, 2001a, b) has shown that “restored” phonemes (phonemes replaced with noise, but which subjects report hearing in a manner consistent with lexical or larger contexts) can drive the selective adaptation found with fully articulated phonemes. If a segment at one end of a categorical perception continuum is repeated many times, the boundary shifts toward that stimulus, such that a smaller step toward the opposite end of the continuum leads to a change in perception. Restored phonemes have similar (though weaker) effects, suggesting the locus is prelexical.

Norris et al. (2003) added a new wrinkle to the debate. Based on evidence for short-term changes in phonemic categories based on implicit perceptual learning, they acknowledged the need for feedback, but argued that it need not occur on-line. Instead, they make a distinction between on-line feedback (as in interactive models) and *feedback for learning*, although without specifying how feedback for learning is triggered or timed; if it is not to happen during processing, the learning signal must be stored until some opportune “down-time” during which the learning signal may be transmitted. The idea is that since (according to their arguments) on-line feedback can serve no useful role, and since a principled division can be made between on-line and “for learning” feedback in computational models, the most parsimonious account remains an autonomous model with feedback for learning. Norris et al. acknowledge the possibility that feedback might be implemented in such a way that it simultaneously provides on-line and for-learning feedback (see (Mirman), McClelland, & Holt, in press, for just such an implementation, which incorporates Hebbian learning into TRACE), but again, that such an architecture is not necessary; on this view, on-line feedback might exist, but only because it either allows a convenient medium for or is an epiphenomenon of feedback-for-learning.

One might argue that in light of the added complexity of post-perceptual decision units in Merge (cf. Samuel, 2001a), the need for feedback to account for perceptual learning, and the ability of a single feedback system to incorporate on-line feedback (accounting for lexical effects on phonemes) and feedback for learning, interaction provides the more parsimonious account. However, given the alternative explanations for the empirical record provided by Norris et al. (2000, 2003), along with their evolving theoretical perspective, there remains room for reasonable disagreement on this debate. Stronger theoretical and empirical cases are required to settle it.

### **3. SELECTION: HOW IS ACTIVATION REGULATED AND RECOGNITION ACHIEVED?**

Once the activation set is specified, a mechanism is needed to evaluate the items in the set and eventually *select* an item for lexical access (and a comprehensive theory must also specify under what conditions selection will fail to occur, e.g., in the case of a nonword input). All current theories assume that a form of competition is required for selection. As a spoken word is heard, multiple lexical items are considered as a function of their phonological similarity to the input and of their frequency of occurrence, or prior

probability, and activated lexical items compete for selection. The two key factors we will discuss here are the role of frequency and a sampling of the competition mechanisms proposed under different theories. We will also include a discussion on the issue related to recognizing words in utterances (i.e., the word segmentation issue), as it requires competition among incompatible hypotheses (those that claim the same portion of the input).

### 3.1. Frequency

It has long been established that words that occur frequently in the language (as reflected by counts of large text corpora) are recognized faster, and more accurately under noisy conditions, than words that occur rarely (e.g., Howes & Solomon, 1951; Savin, 1963). This frequency effect can be couched in Bayesian terms as the impact on perceptual decisions of the prior probability of encountering a given word. The influence of frequency has been instantiated in various ways within theories and models of spoken-word recognition. In search models (e.g., the AUTONOMOUS SEARCH model (Forster, 1989)), word forms are mentally organized into bins, arranged by frequency of occurrence within each bin, with the result that *initial contact* with the lexicon is ordered by frequency. The recognition of a spoken word is viewed as a self-terminating search. The search terminates sooner for high-frequency words, for which a match between the input and a word form can be established early in the search, than for low-frequency words. In localist activation models, which characterize the dominant view in the field, word forms are conceived as independent processing units that accumulate activation proportionally to their match with the incoming signal. In such models, word frequency can directly influence the activation of word units by modulating the units' threshold for response (e.g., the LOGOGEN model (Morton, 1969)), the units' resting (i.e., default) activation (e.g., the COHORT model (Marslen-Wilson, 1987)), the strength of connections between sublexical and lexical representations (MacKay, 1982, 1987), or can act as a post-activation, decision bias, thus acting on *selection* (as in the NAM (Luce, 1986; Luce & Pisoni, 1998; Luce et al., 1990)).

In an attempt to contrast the *initial contact* and *selection* instantiations of frequency, some researchers hypothesized that frequency operating as a decision bias should be observed late, with respect to the onset of spoken input (e.g., Connine, Titone, & Wang, 1993; Goldinger, Luce, & Pisoni, 1989). Although such delay was reported in some studies (Connine et al., 1993), Dahan, Magnuson, and Tanenhaus (2001b) showed that frequency effects could be observed in the earliest moments of lexical processing. They monitored participants' eye movements as they followed spoken instructions to interact with items in a visual display. When fixation proportions over time to low-frequency targets, low-frequency cohorts, high-frequency cohorts and unrelated distractors were compared, Dahan et al. found frequency effects in the earliest signal-driven changes in fixation proportions (within about 200 ms of word onset) – although the magnitude of frequency effects grew as more of a word was heard. Dahan et al. added three frequency mechanisms to TRACE to compare predictions of different proposals for how frequency might be instantiated. Resting level and post-activation bias mechanisms yielded virtually identical predictions (when the post-activation bias was applied continuously, though

for it to have a “late” locus it would have to be applied suddenly after a certain amount of bottom-up evidence accrued). A bottom-up connection strength instantiation (in which connections between phonemes and high-frequency words were stronger than those between phonemes and low-frequency words) provided the best fit to the data. This account predicts a continuous effect of frequency, but a gradual one, since the frequency effect depends on the strength of the bottom-up input. The bottom-up connection strength account would also be consistent with learning models in which connection strengths are tuned to prior probabilities through experience.

### 3.2. Competition

There is now considerable evidence that the recognition of a spoken word is affected by the set of lexical alternatives that are partially compatible with the input. A word that is phonetically similar to few and/or rare other words is recognized more easily than a word similar to many and/or frequent other words, above and beyond effects of the frequency of the word itself (Luce, 1986; Luce & Pisoni, 1998). This indicates that the recognition process does not solely depend on the degree to which the spoken input matches the representation of a given word, but also on the degree to which the input matches the representations of alternative words. All current theories of spoken word recognition acknowledge the need for competition, but differ in the mechanisms they assume accomplishes it. The primary mechanisms are *decision rules* and *direct competition*. We will focus on these, and then turn to a third alternative, *emergent competition*.

*Decision rule competition.* The original COHORT model (Marslen-Wilson & Welsh, 1978) predicted that the recognition of a spoken word depends on the activation of multiple candidates (the *word-initial cohort*) but only indirectly; the cohort determines the *uniqueness point* of the target word – the point at which the target is the last lexical candidate compatible with the input. The model assumed that the onset of a spoken word activates all word candidates sharing that onset. As more input becomes available, candidates are pruned from the competitor set as soon as they mismatch (e.g., *cat* is removed from *castle*’s cohort when /s/ is heard), until only one candidate remains. Inclusion or exclusion of a candidate from the competitor set was viewed as an all-or-none and frequency-insensitive process. Revisions to the model, prompted by theoretical and empirical arguments (Marslen-Wilson, 1987), changed the mechanism for cohort inclusion and exclusion into a gradient activation process reflecting the degree of evidence for a candidate in the input and its frequency. In this revised model, candidates cannot be described as simply in or out of the cohort. Instead, they are more or less activated, and the criterion for recognition was changed into a decision rule that evaluates a unit’s activation level with respect to the activation level of all other units (Marslen-Wilson, 1987, 1993). This, in effect, allows the recognition of a given word to be affected by other candidates’ match to the input, but without direct competition between units; any lexical item’s activation reflects its goodness of fit to the input. Competition only exists at the level of the decision rule.

A similar mechanism was proposed earlier as part of the NAM developed by Luce and colleagues (Luce, 1986; Luce et al., 1990; Luce & Pisoni, 1998). The model states that the probability of recognizing a given word can be approximated by the ratio of the target word's log frequency to the summed log frequencies of all items in its neighborhood, including the target word; in other words, ease of recognition is predicted to be proportional to the amount of frequency the target contributes to the total frequency of its neighborhood. Computed over large sets of words, this probability rule was shown to account for more unique variance in tasks like lexical decision or naming (about 15%) than any other factor (the next best was target frequency alone, which only accounted for 5%). The NEIGHBORHOOD model stands out among current theories in that it is a formal mathematical model of activation and competition, but not a processing model. It also stands out for its power and simplicity. The frequency-weighted probability rule compactly embodies general principles shared by current theories, as well as the specifics of the neighborhood conception of competitors, and generates precise, testable predictions. Nonetheless, as noted above, the NAM fails to incorporate the dynamics of a spoken word's competition environment.

*Direct competition.* Connectionist models like TRACE (McClelland & Elman, 1986), SHORTLIST (Norris, 1994), and more recently PARSYN (Luce et al., 2000) assume competition among lexical units via lateral inhibition. Units within the lexical layer (and the phoneme layer, in the case of TRACE and PARSYN) send each other inhibition as a function of their respective activation, which depends on their similarity to the input. For example, upon hearing the input /kat/ (*cot*), the units *cat* and *cap* would also both be activated; *cat* is more similar to the input than *cap*, and so would be activated more strongly, and send more inhibition to *cap* than viceversa (assuming equal word frequency). The end result is that a lexical item with an activation advantage will eventually suppress its competitors. The recurrent loops created by lateral inhibition in these sorts of models give them temporal dynamics, which allow fine-grained predictions of the activations of targets and competitors over time.

Distinguishing between an implementation of lexical competition in terms of decision rule or lateral inhibition has proven difficult, as they make very similar predictions (Marslen-Wilson, Moss, & van Halen, 1996; see also Bard, 1990). Similar debates are taking place among models of perceptual choice (Usher & McClelland, 2001). Decision-rule competition is arguably a simpler computational mechanism than lateral inhibition. In the decision-rule implementation, the temporal dynamics of candidates' activation can only reflect changes in the evidence supporting each candidate, as the spoken input unfolds over time. By contrast, competition via lateral inhibition predicts temporal dynamics that reflect both the impact of evidence from the input and recurrent loops on candidates' activation. Distinguishing between these two implementations is thus likely to require consideration of lexical activation over time.

*Emergent competition.* Gaskell and Marslen-Wilson (1997, 1999, 2002) have proposed a distributed architecture, where words are represented by overlapping, distributed patterns of node activation. One portion of these nodes stands for phonological features,

while another stands for semantic features. A given word is represented as a pattern of activation among phonological and semantic feature nodes, thus capturing the form and the meaning of that word. When the initial portion of a word is presented to the model, patterns learned by the network that are consistent with the input are simultaneously activated. However, because there is only one substrate for activation—the same set of distributed nodes—the outcome is an activation pattern that *blends* the consistent patterns. Thus, competition takes the form of interference between the patterns associated with candidates consistent with partial input. The activation pattern resulting from processing partial input may be more or less coherent depending on the nature of the information that the nodes encode (phonological vs. semantic) and the number of compatible hypotheses simultaneously considered.

We refer to this as *emergent competition* because the competition dynamics arise from a complex combination of interacting causes. These include intricate patterns of excitatory and inhibitory weights that emerge as a function of the corpus on which a recurrent network is trained, the attractors that form for phonological, semantic, and possibly combinations of inputs and outputs.

The model's distributed architecture makes an intriguing prediction. Although the model assumes the simultaneous activation of all the word candidates that match the input, it also predicts that the resulting pattern of activation does not represent the form or the meaning of any of these candidates individually. Rather, because this activation pattern is a blend, their common features (most often, their shared sounds) are faithfully represented, whereas their divergent features (such as their semantic features, as words that are phonologically similar are not typically semantically related) have been blended; reconstructing the divergent features of word candidates would depend, among other things, on the number of word candidates involved.

Gaskell and Marslen-Wilson (2002) reported data supporting the model's prediction. In particular, they showed that the presentation of a spoken prime that is compatible with several possible candidates (e.g., /kæpt ɪ/, compatible with a number of candidates, including *captain* and *captive*) does not boost participants' speed at making a lexical decision on a word semantically related to one of the candidates (e.g., *commander*), suggesting that the semantic representations of the activated phonological forms were blended and not sufficiently distinctive to allow detectable priming. By contrast, the presentation of a spoken prime that is compatible with only one possible candidate (e.g., /gʊmə/, only compatible with *garment* [British English pronunciation]) did facilitate processing of a word semantically related to this candidate (e.g., *attire*). This result can be accounted for by the distributed architecture assumed by Gaskell and Marslen-Wilson's model because the pattern of activation in the semantic feature nodes becomes less coherent as more candidates are considered and more heterogeneous patterns (associated with form-overlapping candidates with unrelated meanings) participate in the blend.

Models with localist representations could also account for this result. We are unaware of any current, implemented model that could do so without modification, but the general



principles of, e.g., interactive activation are consistent with the result. An explanation parallel to that of Gaskell's and Marslen-Wilson's is that the larger the phonological competitor set is, the weaker the activation that each of their semantic representation receives. The phonological competitors initially receive equivalent support from the phonological input (*mutatis mutandis* for differences in frequency, etc.). As long as no phonological representations are strongly favored by the bottom-up input, however, their corresponding semantic representations receive too little activation to be detected via priming. An analogous mechanism exists in ARTWORD (Grossberg & Myers, 2000), where a perceptual resonance (assumed to lead to conscious perception) is established only once the level of activation of one candidate (or "chunk") has sufficiently overcome that of its competitors.

### 3.3. Word Segmentation in Continuous Speech: Competition Across Word Boundaries

A spoken utterance cannot easily be segmented into the words that compose it because boundaries between words are not reliably marked in the acoustic signal, and have often been blurred through phonological phenomena such as coarticulation and resyllabification. This is not to say that word boundaries are never acoustically marked. For instance, silent pauses between phrases mark the boundaries of the words that appear at the edges of these phrases. In fact, an extensive literature has demonstrated that listeners make use of word-boundary cues when present (phonotactic cues: McQueen, 1998; prosodic cues: Salverda et al., 2003; phonetic cues: Quené, 1992, 1993; Gow & Gordon, 1995). What this literature has shown is that word-boundary cues are used as a source of evidence supporting word candidates that are consistent with the hypothesized word boundary, and not used prelexically, to chunk the signal into words *before* initiating contact with the lexicon, as had been previously proposed (e.g., Cutler, 1990).

Because word boundary cues are probabilistic at best, and because words tend to share many of their components with other words, multiple words are consistent with virtually any portion of an utterance. For example, McQueen et al. (1995) established that 84% of English polysyllabic words contain at least one shorter embedded word (e.g., *ham* in *hamster*, or *bone* in *trombone*). This lexical ambiguity sometimes applies across word boundaries, as in *ship inquiry*, where (in British English) *shipping* matches *ship* and the initial portion of *inquiry*. Thus, competition among word candidates that start at different points in time is required. As mentioned earlier, TRACE models inter-word competition by assuming that all word units that overlap in time, i.e., competing for the same portion of the input, inhibit one another. Because a unit representing the same word is replicated many times over time/space, a given word unit can become activated as soon as the input provides some evidence supporting it, regardless of where in time the information appears. For instance, after the sequence /ʃɪpɪŋ/ (the initial portion of the phrase *ship inquiry*), *inquiry* can start receiving activation from the input, and eventually be recognized, even though *shipping* is already strongly activated. Note that some words can compete even when they do not share any segments. In the example above, the candidate

*shipment* competes with *inquiry* because both are competing for the same portion of the input. Thus, TRACE solves the problem of segmenting words out of a continuous spoken input by using the same mechanism it uses to segment a coarticulated signal into a sequence of phonemic units.

Alternatives to TRACE's solution to word segmentation and recognition have been proposed. Norris (1994) criticized the multiple replications of the lexical network in TRACE. He developed SHORTLIST, a model in which a limited set of candidates that are most activated by (i.e., consistent with) the input is compiled. The model consists of two components. A lexical search network, implemented as a simple dictionary lookup, provides a list of the best matches to the input at each phoneme position. The second component is a competition network including as many as the top 30 candidates aligned with each input position (SHORTLIST is often described as allowing a maximum of 30 words to enter the competition network, but this is inaccurate; D. Norris, personal communication). Items selected for each shortlist compete with one another proportionally to the number of sounds they share in an interactive activation network. Items in different shortlists also compete *if* they overlap. For example, given the input *ship inquiry*, *ship* and *shipping* will enter the shortlist aligned with the first phoneme. *Inquiry* will eventually dominate the shortlist aligned with the fourth phoneme, i.e., after *ship*, and will inhibit *shipping*, because the two overlap in input positions 4 and 5, but it will not inhibit *ship*, since it does not overlap with *ship*. Thus, *ship* and *inquiry* create pressure for a parse into nonoverlapping words, and eventually inhibit *shipping* sufficiently to allow *ship* to be recognized. The selection–competition cycle repeats itself as input is presented to the model. At each time step, a new lexical search is done for *every position encountered so far*. The composition of the shortlist changes dynamically as spoken input becomes available, with some candidates dropping and being replaced by new candidates, depending on bottom-up match/mismatch scores from the lexical search network and inhibition within the competition network.

Despite the important computational economy offered by establishing the competitor set in a dynamical fashion, compared to a hard-wired manner as in TRACE, SHORTLIST also has several limitations. First, the lexical search mechanism is called recursively—a new search is done at each position as each new phoneme is heard. If the lexical search were implemented as a recurrent network, this would require one copy of the lexical network for each phoneme position, and so the model would require the same number of nodes as TRACE, plus those used in the shortlists (but would use many fewer *connections*). Second, the biological plausibility of the dynamic programming required by SHORTLIST must be addressed (cf. Protopappas, 1999). Finally, it has yet to be shown that SHORTLIST can account for the broad range of data TRACE can.

ARTWORD (Grossberg & Myers, 2000) is a model specifically designed to account for the dynamics of inter-word competition and how later-arriving information can modulate the perception of earlier occurring speech. In this model, the spoken input activates sensory features. Activation of these features is transformed into a sequence of items in working memory. The sequential order of these items is encoded by a gradient of activity

within the representation (with the most active item representations corresponding to the most recent event). The activity pattern in working memory in turn activates “list chunks” that match the active items *and* their order. List chunks consist of unitized linguistic units (e.g., phonemes, syllables, words). Activated chunks compete with one another, proportionally to their level of activation and to the number of items they compete for. Once an activated list chunk reaches an activation threshold, it sends back activation to the consistent items in working memory, and inhibition to inconsistent items.

The excitatory loop between list chunks and items in working memory corresponds to a process known as *resonance*. In Grossberg and Myers’s (2000) own words, “when listeners perceive fluent speech, a wave of resonant activity plays across the working memory, binding the phonemic items into larger language units and raising them into the listener’s conscious perception” (p. 738). Thus, in this model, recognizing a spoken word can be described as having associated a given linguistic interpretation to a portion of speech represented in working memory, where time is encoded.

The dynamics of the resonance wave is the major factor that determines how continuous speech is perceived as a succession of segmented and unitized word units. First, the model includes two reset mechanisms that can terminate one resonance to allow for the next one to be initiated (see Grossberg, Boardman, & Cohen, 1997, for more details). Thus, the perception of a multi-word utterance can be described as a sequence of resonance waves. Second, because of competition among activated chunks, ARTWORD accounts for recognition despite the activation of multiple candidates at various points in the signal. Third, the model allows for later-arriving information to modify the resonance wave by *resonant transfer*: The resonance associated with a short word (e.g., *ham*) can be transferred to a longer one (e.g., *hamster*) as the second syllable of the word *hamster* is processed. Finally and critically, ARTWORD can account for the impact of some word-boundary cues (such as segmental lengthening, e.g., Salverda et al., 2003) without invoking additional mechanisms. Indeed, a resonance transfer can only occur within a very limited, speech-rate-dependent time window. Thus, if the first sounds of the second syllable of *hamster* are delayed (because of lengthening of the last sounds of *ham*, a silent pause, or lengthening of the sound following *ham*), the resonance established between the word chunk *ham* and items in working memory may have been reset, and the items’ activation fallen to low activation levels. No resonance transfer is then possible, and listeners will perceive the word *ham* followed by another word starting with the sounds /st/. This is consistent with Salverda et al.’s results, showing that long /ham/ syllables tend to be interpreted as monosyllabic words.

#### 4. INTEGRATION: WHEN AND HOW IS CONTEXT INTEGRATED?

Words occur embedded in a larger context, most often in a sentence. There exists a tight interdependency between a given word and its sentential context. A word contributes to the meaning of the sentence, but the contribution of a word to the meaning of the sentence also rests on the sentence itself.

Most of the empirical work examining the interaction between a word and its sentential context has focused on the possible constraint that the context may impose on the set of word candidates compatible with the spoken input. Initial studies suggested a late impact of context. For example, Tanenhaus, Leiman, and Seidenberg (1979; see also Swinney, 1979) presented listeners with auditory sentences that were biased toward one sense of a homophone (e.g., *she held the rose* vs. *they all rose*), and then used visual lexical decision to probe semantic activation. They found statistically equivalent priming for associates of both senses (e.g., *flower* and *stand*) immediately after homophone offset, but only found reliable priming for the context-appropriate sense 250 ms later. This was interpreted as evidence for context-free initial lexical activation, quickly followed by an integration stage where word interpretations incompatible with the context are rejected. Similar conclusions were reached by Zwitserlood (1989), who reported evidence for the early activation of the meaning of all words compatible with the initial sounds of a spoken word, regardless of the context.

However, Shillcock and Bard (1993) tested the hypothesis that the Tanenhaus et al. contexts contained very weak biases (other form classes besides nouns or verbs could have been heard at the homophone position, and the contexts at best biased listeners toward thousands of nouns vs. thousands of verbs). They used contexts that had been experimentally established as biased towards a single item – the closed class word, *would* (*John said he didn't want to do the job but his brother would, as I later found out*) – or towards a large number of items: (*John said he didn't want to do the job with his brother's wood, as I later found out*). In the closed-class case, they found no evidence of priming of *wood*; its associate, *timber*, was not primed even if they probed prior to the offset of *would*. This suggests that top-down context *can* affect early stages of word recognition, but that top-down information is generally given much less weight than bottom-up, and is proportional to prior probability: the more narrowly constraining the top-down information is, the greater the impact it may have on early moments of processing (see Dahan, Swingley, Tanenhaus, & Magnuson, 2000, who report evidence for the early impact of determiners marked for grammatical gender on the recognition of subsequent spoken nouns in French).

Generally speaking, theories of spoken word recognition have remained agnostic about the integration of sensory information with higher level context. Notable exceptions to this are the three versions of the COHORT model. In the original COHORT model, top-down knowledge (e.g., semantic context) played an active role throughout selection, allowing recognition prior to the uniqueness point for words strongly supported by context. It also had the potential to guide initial contact, by preventing a highly inconsistent item from entering the recognition cohort. In the revised COHORT model, in light of intuitive and empirical evidence that clearly articulated words that have low probability in a particular context are still clearly perceived, context no longer affected initial contact (i.e., could no longer exclude an item from entering the cohort despite strong bottom-up support). Instead, context was viewed as acting on a set of candidates first established on the basis of sensory information only. The model argued in favor of a context-free, initial activation

stage. The most recent version of the model, the *DISTRIBUTED COHORT* model, departs from this stance by assuming no division between initial contact and selection. Semantic features are an integral part of lexical representations, and thus semantic and phonological knowledge are simultaneously activated by bottom-up input. This last instantiation, by renouncing the theoretical processing division between form and meaning, is compatible with findings of a continuous integration of different sources of evidence in order to ultimately derive an interpretation of the spoken input.

## 5. AVENUES FOR PROGRESS

The three most crucial developments for theories of spoken-word recognition, as argued throughout this chapter, are (1) increasing evidence that the input to spoken word recognition retains much if not all of the surface detail of utterances; (2) evidence that language representations are not static but instead are subject to constant change; and (3) the emergence of theoretical frameworks that deny the existence of distinct stages corresponding to speech perception, spoken word recognition, sentence processing, and beyond – and empirical support for these theories. These developments may herald a radical reconceptualization of spoken word recognition and language processing in general, if not an all-out paradigm shift.

There are two sets of findings that compellingly demonstrate that the input to lexical access is not limited to an abstract phonemic code. The first (reviewed briefly in our introduction) is evidence that fine-grained phonetic detail affects the time course of lexical activation and competition (Andruski et al., 1994; Davis et al., 2002; Salverda et al., 2003). The second (reviewed in Section 2) is evidence that even (putatively) non-linguistic surface detail, such as talker sex or even more fine-grained talker characteristics, is preserved in memory for spoken language (Goldinger, 1996). The fact that such detail not only affects memory but also word recognition motivates exemplar theories like Goldinger's (1998) episodic lexicon theory, in which the basis for lexical (and potentially lower and higher levels of representation) categories are clusters of memory traces of, essentially, raw speech "episodes" that preserve all surface detail. On such a view, each new memory trace has the potential to change the "category" with which it is clustered, making exemplar theories compatible with recent evidence that short-term changes in phonotactic probabilities quickly influence production (Dell, Reed, Adams, & Meyer, 2000) and comprehension (Onishi, Chambers, & Fisher, 2002). These rapid changes in lexical production and processing challenge the frequent, if implicit, assumption that the adult phonological and lexical knowledge is more or less fixed.

These developments pose significant challenges to theories of spoken word recognition and spoken language processing in general. They point to a system in which there may be distinct levels of representation (given the cognitive economies of compositionality and generativity afforded by, e.g., phonemes and words), but also parallel episodic representations that are less abstract, and without discrete stages corresponding to the descriptive levels of speech perception, word recognition, sentence processing, and so on.

As mentioned earlier, Goldinger and Azuma's (2003) appeal to adaptive resonance (e.g., Grossberg, 2003) as a potentially unifying framework capable of incorporating learning, sublexical and lexical effects as well as the principles of episodic lexicon theory, appears to hold substantial promise.

However, integrating this view with the processing of actual speech or a close analog, remains a significant challenge. While the ARTSTREAM model (Grossberg et al., 2004) has demonstrated the potential of the ART framework to process the speech signal itself, it has not yet been extended to contact with phonemic or lexical forms. Plaut and Kello (1999) provided another framework with significant promise, in which close analogs of the speech signal are used, and phonological and semantic representations are treated within perception and production, as well as development.

Integrating (descriptive) levels of speech perception and word recognition upwards also remains as a significant challenge. Theories of sentence processing in the constraint-based framework have long blurred the boundary between lexical access and sentence processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Kim, 1998), assuming that lexical representations include not just phonological and semantic knowledge, but also specify the syntactic relations in which a lexical item can participate. Evidence that lexical access and sentence processing are constrained in an immediate and continuous fashion by nonlinguistic context – such as the actions afforded to the listener by combinations of objects and instruments (Chambers, Magnuson, & Tanenhaus, 2004), or even affordances available to interlocutors (Hanna & Tanenhaus, 2003) – demands that we scale our theories up and integrate them with sentence, discourse, and general cognitive processes.

We began this chapter by contrasting the modal, narrow view of spoken word recognition (as the mapping from phonemes to sound forms that provide access to the lexicon) with a broad view, encompassing the speech signal, the word level, and higher levels of structure and representation. The broad view is supported by growing evidence for continuous effects of subphonemic information at the lexical level and beyond on the one hand, and immediate integration and interaction between descriptively low and high levels of linguistic representation and even non-linguistic affordances of physical objects (Chambers et al., 2004) on the other. Our view is that significant progress in understanding spoken word recognition, and language processing more generally, will require stretching (or possibly abandoning) current theories and models to accommodate the broad view of language processing.

## ACKNOWLEDGEMENTS

We thank Dan Swingley and Dan Mirman for comments and discussions that greatly improved this chapter. Preparation of this chapter was supported in part by the National Science Foundation under Grant No. 0433567 to DD, and by National Institute on Deafness and Other Communication Disorders Grant DC-005765 to JSM.

## REFERENCES

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, *38*, 419–439.
- Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory & Language*, *41*, 124–145.
- Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163–187.
- Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*, 339–362.
- Bard, E. G. (1990). Competition, lateral inhibition, and frequency: Comments on the papers of Frauenfelder & Peeters, Marslen-Wilson, and others. In: G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 185–210). Cambridge, MA: MIT Press.
- Bertelson, P., & de Gelder, B. (1989). Learning about reading from illiterates. In: M. Galaburda, & M. Albert (Eds), *From reading to neurons* (pp. 1–23). Cambridge, MA: MIT Press.
- Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2004). Reference resolution in the wild: Circumscription of referential domains by naive participants during an interactive problem solving task. In: J. C. Trueswell & M. K. Tanenhaus (Eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
- Cairns, P., Shillock, R., Chater, N. & Levy, J. (1995) Bottom-up connectionist modelling of speech. In: J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds). *Connectionist models of memory and language*, London, UK: University College London Press.
- Chambers, C. G., Magnuson, J. S., & Tanenhaus, M. K. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *30*, 687–696.
- Clarke, C. M., and Garrett, M. F. (2004) Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*, 3647–3658.
- Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance*, *16*, 551–563.
- Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, *13*, 153–156.
- Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, *11*, 1084–1089.

- Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 81–94.
- Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In: G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 105–121). Cambridge: MIT Press.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language & Speech*, *40*, 141–201.
- Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In: W. E. Cooper, & E. C. T. Walker (Eds), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale: Erlbaum.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.
- Dahan, D., Swingle, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in french. *Journal of Memory and Language*, *42*, 465–480.
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 498–513.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222–241.
- Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 218–244.
- Deelman, T., & Connine, C. M. (2001). Missing information in spoken word recognition: Nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 656–663.
- Dell, G.S., Reed, K.D., Adams, D.R. & Meyer, A.S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1355–1367.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception & Performance*, *23*, 914–927.



- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*, 224–238.
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*, 301–306.
- Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech wave. In: J. S. Perkell, & D. H. Klatt (Eds), *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, *27*, 143–165.
- Fodor J (1983) *Modularity of mind*. Cambridge, MA: MIT Press.
- Forster, K. I. (1989). Basic issues in lexical processing. In: W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.
- Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics*, *36*, 359–368.
- Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 526–540.
- Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition*, *25*, 1–20.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, *6*, 110–125.
- Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, *31*, 447–463.
- Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, *19*, 407–439.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, *22*, 144–158.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language & Cognitive Processes. Special Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives on the Lexicon*, *12*, 613–656.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 380–396.

- Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, 23, 439–462.
- Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory & Language*, 44, 325–349.
- Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, 45, 220–266.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Goldinger, S.D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 305–320.
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language*, 28, 501–518.
- Gow, D. W., Jr. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory & Language*, 45, 133–159.
- Gow, D. W., Jr. (2002). Does english coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception & Performance*, 28, 163–179.
- Gow, D. W., Jr. (2003a). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65, 575–590.
- Gow, D. W., Jr. (2003b). How representations help define computational problems: Commentary on Grossberg, Gaskell and Greenberg. *Journal of Phonetics*, 31, 487–493.
- Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception & Performance*, 21, 344–359.
- Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14, 421–433.
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, 423–445.
- Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception & Performance*, 23, 481–503.

- Grossberg, S., Govindarajan, K.K., Wyse, L.L., & Cohen, M.A. (2004) ARTSTREAM: A neural network model of auditory scene analysis and source segregation". *Neural Networks*, 17, 511–536.
- Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, 107, 735–767.
- Hahn, U., & Bailey, T. M. (2005). What makes words sound similar. *Cognition*, 97, 227–267.
- Hanna, J.E. & Tanenhaus, M.K. (2003). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, 28, 105–115.
- Hintzman, D. L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411–428.
- Hintzman, D. L., Block, R. A., & Inskip, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, 11, 741–749.
- Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41, 401–410.
- Joos, M. (1948). *Acoustic phonetics*. Baltimore, MD: Linguistic Society of America.
- Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory & Language*, 40, 153–194.
- Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279–312.
- Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.
- Liberman, A., Cooper, F. S., Shankweiler, D., P. & Studdert-Kennedy, M. G. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201–212.
- LoCasto, P., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception and Psychophysics*, 64, 208–219.
- Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39, 155–158.
- Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and parsyn. *Perception and Psychophysics*, 62, 615–625.
- Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, 26, 708–715.

- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.
- Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In: G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge, MA: MIT Press.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.
- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89*, 483–506.
- MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer.
- Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003a). Lexical effects on compensation for coarticulation: The ghost of Christmas past. *Cognitive Science, 27*, 285–298.
- Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003b). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science, 27*, 801–805.
- Mann, V. A., & Repp, B. H. (1981). Influence on preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America, 69*, 548–558.
- Marslen-Wilson, W. (1984). Function and process in spoken word recognition. In: H. Bouma & D. G. Bouwhuis (Eds), *Attention and performance x: Control of language processes* (pp. 125–150). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71–102.
- Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In: W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.
- Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In: G.T.M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp.148–172). Cambridge, MA: MIT Press.
- Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In: G. T. M. Altmann, & R. Shillcock (Eds), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting*, pp. 187–210. Mahwah: Erlbaum.
- Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 1376–1392.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representations and process in lexical access: Words, phonemes, features. *Psychological Review, 101*, 653–675.

- Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance*, *15*, 576–585.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- McClelland, J. L., St John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language & Cognitive Processes. Special Parsing and Interpretation*, *4*, SI287–SI335.
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 539–553.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*, B33–B42.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory & Language*, *39*, 21–46.
- McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, *10*, 309–331.
- McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 1363–1389.
- Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*, 457–465.
- Mirman, D., McClelland, J. L., & Holt, L. L. (in press). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin and Review*.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*, 365–378.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences*, *23*, 299–370.

- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.
- Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. In: K. Johnson, & J. W. Mullennix (Eds), *Talker variability in speech processing* (pp. 109–132). San Diego, CA: Academic Press.
- Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In: Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds), *Speech perception, speech production, and linguistic structure* (pp. 113–134). Tokyo: OHM.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, *83*, B13–B23.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 309–328.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.
- Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*, 21–52.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory & Language*, *39*, 347–370.
- Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In: B. MacWhinney (Ed.), *The emergence of language* (pp. 381–416). Mahwah, NJ: Lawrence Erlbaum Associates.
- Protopapas, A. (1999). Connectionist modeling of speech perception. *Psychological Bulletin*, *125*, 410–436.
- Quené, H. (1992). Durational cues for word segmentation in dutch. *Journal of Phonetics*, *20*, 331–350.
- Quené H. (1993). Segment durations and accent as cues to word segmentation in dutch. *Journal of the Acoustical Society of America*, *94*, 2027–2035.
- Remez, R. E. (2005). The perceptual organization of speech. In: D. B. Pisoni, R. E. Remez (Eds), *The Handbook of Speech Perception*, (pp. 28–50). Oxford: Blackwell.
- Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, *19*, 394–398.

- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474–494.
- Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*, 97–127.
- Samuel, A. G. (2001a). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*, 348–351.
- Samuel, A. G. (2001b). Some empirical tests of merge's architecture. *Language & Cognitive Processes*, *16*, 709–714.
- Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory & Language*, *48*, 416–434.
- Savin, H.B. (1963). Word frequency effects and errors in the perception of speech. *Journal of the Acoustical Society of America*, *35*, 200–206.
- Scharenborg, O., ten Bosch, L., Boves, L., & Norris, D. (2003). Bridging automatic speech recognition and psycholinguistics: Extending shortlist to an end-to-end model of human speech recognition (I). *Journal of the Acoustical Society of America*, *114*, 3032–3035.
- Shankweiler, D., & Fowler, A. E. (2004). Questions people ask about the role of phonological processes in learning to read. *Reading and Writing: An Interdisciplinary Journal*, *17*, 483–515.
- Shillcock, R. C., & Bard, E. G. (1993). Modularity and the processing of closed-class words. In: G. T. M. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The second sperlonga meeting* (pp. 163–183). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, *96*, 1314–1324.
- Speer, S., & Blodgett, A. (2006). Prosody. In: M. A. Gernsbacher, & M. Traxler (Eds), *Handbook of Psycholinguistics*. Amsterdam: Elsevier.
- Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*, 322–338.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior*, *18*, 645–659.
- Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal Behavior*, *18*, 427–440.

- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Trueswell, J.C., & Kim, A.E. (1998). How to prune a garden-path by nipping it in the bud: Fast-priming of verb argument structures. *Journal of Memory and Language*, 39, 102–123.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108, 550–592.
- Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception and Psychophysics*, 62, 1297–1311.
- van Ooijen, B. (1996). Vowel mutability and lexical selection in english: Evidence from a word reconstruction task. *Memory and Cognition*, 24(5), 573–583.
- Warren, R. M. (1970). Restoration of missing speech sounds. *Science*, 167, 392–393.
- Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, 35, 49–64.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.



This page intentionally left blank

## Chapter 9

### **Visual Word Recognition: The Journey from Features to Meaning (A Travel Update)**

*David A. Balota, Melvin J. Yap, and Michael J. Cortese*

#### **1. INTRODUCTION**

##### **1.1. The Word**

Well, it has been more than a decade since the literature on visual word recognition has been reviewed for the Gernsbacher (1994) *Handbook of Psycholinguistics*, and there continues to be considerable interest in understanding the processes tied to the “word.” Understanding the journey from features to meaning has clearly made progress, but there is still some distance to go. In the present chapter, we will provide an update on the major issues that were covered in the 1994 chapter, and introduce many new issues that have arisen during the interim. This chapter will focus on isolated visual word recognition research; other chapters in this volume are devoted to auditory word recognition, and recognizing words in sentential context. The goal of the present review is not to provide in-depth reviews of every area addressed by word recognition researchers. This would far exceed space limitations. Rather, we will attempt to acquaint the reader with the richness and diversity of the empirical and theoretical issues that have been uncovered in this literature.

The organization of the chapter is as follows: First, we will briefly outline why word recognition research has been central to a number of quite distinct developments in cognitive psychology, psycholinguistics, and cognitive neuroscience. Second, we will review the evidence regarding letter recognition, sublexical organization, and lexical-level influences on word recognition. Interspersed within each of these sections is a discussion of some of the current theoretical developments and controversies. Third, we will review the literature on context and priming effects in word recognition; again, highlighting major theoretical developments and controversies. Fourth, we will discuss some limitations regarding inferences that are possible based on the available data, and highlight some recent developments that have provided additional leverage on such issues.

## 1.2. Why the Word?

In order to provide a framework for understanding the breadth of word recognition research, it is useful to list a few of the basic research issues that the word recognition literature has touched upon. For example, word recognition research has been central to notions regarding different levels/codes of analysis in language processing, attention, and memory (e.g., Craik & Lockhart, 1972; Posner, 1986). The lexical unit is ideally suited for such work because words can be analyzed at many different levels, e.g., features, letters, graphemes, phonemes, morphemes, semantics, among others. As we shall see below, much of the work in visual word recognition has been devoted to identifying the functional roles of these different levels.

Second, word recognition research has been central in the development of theories of automatic and attentional processes (e.g., Healy & Drewnowski, 1983; LaBerge & Samuels, 1974; Neely, 1977; Posner & Snyder, 1975). Part of the reason for this emphasis is the natural relation between the development of reading skills and the development of automaticity. Here, one can see the extra impetus from education circles regarding the development of word recognition skills. Moreover, the notion that aspects of word recognition have been automatized and are no longer under the conscious control of the reader has historically provided some of the major fuel for arguments regarding self-encapsulated linguistic processing modules (see Fodor, 1983). As we shall see, the issue of how attentional control signals might modulate processes involved in word recognition has received renewed interest recently, and hence, notions of automaticity and modularity have been reevaluated.

Third, word recognition research has also been central to developments regarding basic pattern recognition processes. One of the most difficult problems in pattern recognition research has been identifying the underlying subordinate critical features of a given pattern (e.g., Neisser, 1967). Written words are relatively well-defined patterns. Historically, because words have been the central unit of analysis in much of the verbal learning and memory research that dominated experimental psychology between the 1950s and 1960s, there was considerable interest in developing norms that quantify different components of words (e.g., Kučera & Francis', 1967, word frequency norms; Noble's, 1952, meaningfulness norms; Osgood, Suci, & Tannenbaum's, 1957, semantic differential). As we shall see, there has been a resurgent interest in developing norms that help quantify the different characteristics of words (see for example, the chapter by Burgess, this volume). Clearly, the importance of the lexical unit in developing models of pattern recognition is due in part to the efforts devoted to defining the stimulus.

Finally, because words are relatively well-characterized patterns, they have been the focus of development of formal mathematical models of pattern recognition. For example, one of the first formal models in cognitive psychology was the Selfridge and Neisser (1960) Pandemonium model of letter recognition. Moreover, the interactive activation framework developed by McClelland and Rumelhart (1981) was central to nurturing the current widespread interest in formal connectionist models of cognitive performance (for example, see Seidenberg & McClelland, 1989). As we shall see, word-level analyses

appear to be an ideal battleground for pitting symbolic, rule-based models against connectionist models of cognition.

In sum, word recognition research has been central to work in cognitive psychology and psycholinguistics because words are relatively well-defined minimal units that carry many of the interesting codes of analysis (i.e., orthography, phonology, semantics, syntax), and processing distinctions (e.g., automatic vs. attentional) that have driven much of the work in cognitive psychology and psycholinguistics. Thus, although it would seem that the more important goal would be to pursue how individuals process language at higher levels such as clauses, sentences, and paragraphs, many researchers have pursued research at the level of the word because of its inherent tractability. In fact, word-level analysis was the initial focus of neuroimaging studies (see Petersen, Fox, Posner, Mintun, & Raichle, 1989), and continues to be central to the efforts in the burgeoning field of cognitive neuroscience. As we shall see in the following review, although progress is being made, the ease of tracking the processes involved in word recognition may be more apparent than real.

## **2. FEATURES, LETTERS, AND MODELING CONSTRAINTS**

We shall now review some of the variables that have been pursued in word recognition research. First, we shall attempt to break the word down into smaller, more tractable bits. Second, we will discuss work that addresses how orthography maps onto phonology in English. Third, we will discuss the influence of variables that can be quantified at the whole word level, e.g., frequency, familiarity, age of acquisition, orthographic neighborhood size, along with a set of additional semantic variables. Fourth, we will provide an overview of the influence of single word context on isolated word recognition, via a review of the priming literature. Sprinkled within each of these sections will be discussion of the major theoretical models and issues.

### **2.1. Features**

A common approach to understanding pattern recognition is that a given pattern must first be broken down into features that are common to the set of patterns that one is interested in modeling. Some of the initial work in this area was developed by Gibson and Gibson (1955), who forcefully argued that feature-level analyses were an essential aspect of pattern recognition and, more generally, perceptual learning. These primitive features were the building blocks for pattern recognition. This provided researchers with a well-specified problem: what are the primitive features used in letter recognition? The hunt was on!

Fortunately, the feature analytic approach is ideally suited for letter recognition. Although there are differences across fonts, English orthography can be relatively well-described by a limited set of features, such as horizontal lines, vertical lines, closed curves, open curves, intersections, cyclic redundancy, and others (see, for example, Gibson, Osser, Schiff, & Smith, 1963). Once researchers proposed such primitive

features, both behavioral and neurological evidence began to accumulate that documented the role of such features in visual perception. On the behavioral side, there were studies of confusion matrices indicating that letters that shared features were more likely to be confused in degraded perceptual conditions, compared to letters that did not share many features (e.g., Kinney, Marssetta, & Showman, 1966). In addition, visual search studies by Neisser (1967), among others, indicated that subjects were relatively faster to find a given target letter (e.g., Z) when it was embedded in a set of letters that did not share many features with the target (e.g., O, J, U, D), compared to a set of letters that did share many features with the target (e.g., F, N, K, X).

There was also exciting evidence accumulating during the same period that appeared to identify neural substrates that might subserve feature-like detection processes. Consider, for example, the pioneering (and Nobel Prize winning) work by Hubel and Wiesel (1962, 1968). These researchers used single cell recording techniques to investigate neural activity in areas of the striate cortex in alert cats. When different stimuli were presented to the retina of the cat, there were increases in neural activity in specific cortical areas. Hubel and Wiesel found evidence that there were cells that appeared to be especially sensitive to visual stimuli that mapped onto such things as vertical lines, horizontal lines, angles, and even motion. The importance of this work is very simple: it provided the neurological evidence that converged with the notion that pattern recognition ultimately depends upon primitive feature analytic processes. More recent work by Petersen, Fox, Snyder, and Raichle (1990) using positron emission tomography has extended this work to humans in demonstrating significant blood flow changes in specific areas of the striate cortex corresponding to feature-like detection systems for letter fonts in humans.

At the same time behavioral and neural evidence was accumulating in support of features being used in pattern recognition, one of the first computational models of pattern recognition was developed. This was a model of letter recognition developed by Selfridge (1959; Selfridge and Neisser, 1960). The model initially coded the stimulus into a set of 28 visual features that provided support for the letters that were most consistent with those features. The Pandemonium model had the capacity to learn which features were especially discriminating among letters, and adjusted the weights for these features accordingly. As we shall see, the Pandemonium model predates by some 20 years important developments in letter and word recognition models. It is quite amazing that the Pandemonium model worked so well given the computational hardware limitations in the late 1950s and early 1960s.

Although most models of word recognition assume a first step of primitive feature identification, there are still many unanswered questions in this initial stage of processing: First, what is the glue that puts the features together? Specifically, once vertical lines, horizontal lines, and intersections have been detected, how does one put the features together to identify the letter T? We typically do not perceive free-floating features (for a review of the binding issue, see Treisman, 1996, 1999). Second, what happens in the feature analytic models when distortions occur that modify the feature, e.g., does a 15° rotated vertical line still activate the vertical line detector? Third, and along the same lines, what are the critical features when the letters are distorted via different fonts or a novel style of handwriting?

Reading still proceeds in an acceptable fashion even though there are considerable changes in the critical set of features (see Manso de Zuniga, Humphreys, & Evett, 1991). Interestingly, there is some evidence that there may be differences in the way people process printed words and cursive handwriting. For example, case mixing disrupts reading performance with printed words (Mayall, Humphreys, & Olson, 1997) but can actually facilitate performance with handwriting (Schomaker & Segers, 1999). This suggests that distinctive word contours are more critical in handwriting recognition (for a description of a computational model of handwriting, see Schomaker & Van Galen, 1996). Fourth, are features across letters coded serially in reading, e.g., from left to right in English orthography, or is there a parallel coding of features? Based on the work by Treisman (1986), one might expect that there is an early parallel coding of features that is followed by a more capacity demanding binding process (see, however, Shulman, 1990). As we will see, the distinction between parallel and serial processing in word recognition has been a central area of debate in the literature (for a recent discussion, see Rastle & Coltheart, 2006). Finally, are features within letters the critical level of analysis in word recognition or are there supraletter and/or even word-level features (e.g., Purcell, Stanovich, & Spector, 1978) that are more important? Although there has been considerable progress in understanding how features contribute to pattern recognition (for a review, see Quinlan, 2003), there are still many questions that need to be resolved in mapping features onto letters. In lieu of getting bogged down in some of the more important fundamental aspects of visual perception, let us take the leap of faith and assume we have made it to the letter. Surely, things must get a bit more tractable there.

## 2.2. Letters

Assuming that features play a role in letter recognition, and letters are crucial in word recognition, one might ask what variables are important in letter recognition. For example, does the frequency of a given letter in print influence its perceptability? Fortunately, there seems to be a relatively straightforward answer to this question. Appelman and Mayzner (1981) reviewed 58 studies that entailed 800,000 observations from a variety of paradigms that spanned 100 years of research. The conclusion from their review is very straightforward: letter frequency does appear to influence speeded tasks such as letter matching, naming, and classification tasks (e.g., is the letter a vowel or a consonant?). However, letter frequency does not appear to influence accuracy in perceptual identification tasks. The results from the Appelman and Mayzner study are intriguing for three reasons: First, *a priori*, one would clearly expect that frequency of any operation (perceptual, cognitive, or motoric) should influence performance, and hence, it is unclear why there is not a letter frequency effect in identification tasks. Second, as we shall see below, there is a consistent word level frequency effect in both response latency tasks and perceptual identification tasks, and hence, there at least appears to be a difference between frequency effects at different levels within the processing system, i.e., letters vs. words. Third, this is our first exposure of a general theme that runs across the word recognition literature, i.e., different tasks or analyses yield different patterns of data, and so it is incumbent upon the researcher to build a task of not only the targeted dimensions in word processing, but also the tasks that are used to tap these dimensions.

### 2.3. Features, Letters, and Word Interactions: Some Initial Models

Important theoretical issues regarding letter recognition date back to questions that were originally posed by Cattell (1885). The interest here is to define the perceptual unit in word recognition. *A priori*, it would seem obvious that the letter should be the primary unit of analysis in visual word recognition, i.e., words are made up of letters. However, Cattell (1885, 1886) reported evidence that was initially viewed as inconsistent with this notion. Cattell found that some words can be named more quickly than single letters. The problem this finding posed was very simple: how could the letter be the critical unit of analysis in word recognition, if words could be named more quickly than the letters that presumably make up the words? Along with the Cattell results, it was also reported that the exposure duration necessary to identify a word was in some cases less than the exposure duration necessary to identify a single letter. In fact, Erdmann and Dodge (1898) reported that the exposure duration necessary to identify four to five letters in a display was sufficient to read single words that could contain as many as 22 letters. Again, the conundrum is that if words can be better perceived than letters then how can letters be the basic unit of perception, since words are made up of letters?

Of course, an alternative account of this pattern of data is simply that subjects can use any available information regarding orthographic redundancy and lexical-level information to facilitate word processing, and such information is unavailable when single letters are presented. For example, if you thought you saw the letters T and H at the beginning of a short briefly presented word and the letter T at the end then you are likely to guess that there was the letter A between the TH and T, producing the word THAT. This was labeled the sophisticated guessing account of some of the initial findings. However, because of a seminal study by Reicher (1969), it appeared that there was more to this phenomena than simply sophisticated guessing. In Reicher's study, on each trial, one of three stimuli was briefly flashed (e.g., a single letter, K, a word, WORK, or a nonword, OWRK), after which a patterned mask was presented. After the mask was presented, subjects were presented with two letters (e.g., D and K) adjacent to the position of the previous target letter for a forced-choice decision. The remarkable finding here is that subjects produced reliably higher accuracy when the first stimulus was a word than when it was a single letter or a nonword. Because both the letters D and K produce acceptable words within the WOR context, subjects could not rely on pre-existing lexical knowledge to bias their response one way or the other (for an alternative view, see Krueger & Shapiro, 1979; Massaro, 1979). Hence, it appeared that subjects actually *see* letters better when embedded in words than when embedded in nonwords. This finding was termed the word-superiority effect and was also reported in a study by Wheeler (1970), so it sometimes also is referred to as the Reicher-Wheeler effect.

There were two important subsequent findings that constrained the interpretation of the word superiority effect. First, the effect primarily appears under conditions of patterned masking (masks that involve letter-like features) and does not occur under energy masking (masks that involve high-luminance contrasts, e.g., Johnston & McClelland, 1973; Juola, Leavitt, & Choe, 1974). In fact, it appears that the interfering effect of the mask is primarily on performance in the letter alone condition and does not produce much of a breakdown

in the word condition (Bjork & Estes, 1973). Second, letters are also better recognized when presented in pronounceable nonwords (e.g., *MAVE*), compared to unpronounceable nonwords or alone (e.g., Carr, Davidson, & Hawkins, 1978; McClelland & Johnston, 1977). Thus, the word-superiority effect does not simply reflect a word-level effect.

The importance of the word-superiority effect derives not only from the information that it provides about letter and word recognition, but also from its historical impact on the level of modeling that researchers began to use to influence their theory development. Specifically, this effect led to the development of a quantitative model of word and letter recognition developed by McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982; also see Paap, Newsome, McDonald, & Schvaneveldt, 1982). As noted earlier, this type of modeling endeavor set the stage for the explosion of interest in connectionist models of cognitive processes (e.g., McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989).

Figure 1 provides an overview of the architecture of the McClelland and Rumelhart (1981) model. Here, one can see the three basic processing levels; feature detectors, letter detectors, and word detectors. These levels are attached by facilitatory (arrowed lines) and/or inhibitory (knobbed lines) pathways. As shown in Figure 1, there are inhibitory connections within the word level and within the letter level. Very simply, when a stimulus is presented, the flow of activation is from the feature level to the letter level and eventually onto the word level. As time passes, the letter-level representations can be reinforced, via the facilitatory pathways, by the word-level representations and vice versa. Also, as time passes, within both the letter and word level representations, inhibition from highly activated representations will decrease the activation at less activated representations, via the within-level inhibitory pathways.

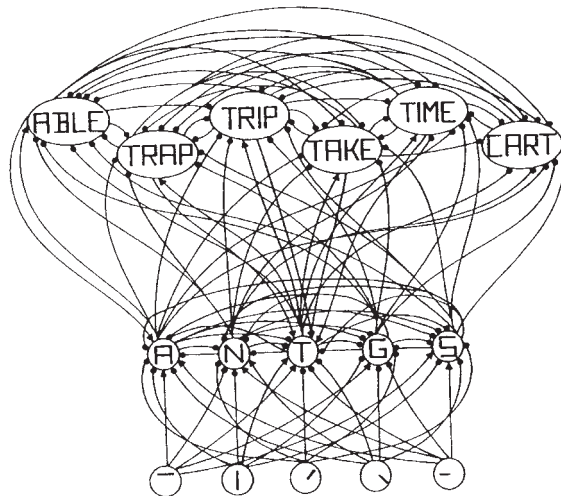


Figure 1. McClelland and Rumelhart's (1981) interactive activation model of letter recognition.



How does the model account for the word-superiority effect? The account rests heavily on the notion of cascadic processes in the information processing system (see Abrams & Balota, 1991; Ashby, 1982; McClelland, 1979). Specifically, a given representation does not necessarily need to reach some response threshold before activation patterns can influence other representations, but rather, there is a relatively continuous transfer of activation and inhibition across and within levels as the stimulus is processed. Consider the letter alone condition in the Reicher paradigm, described earlier. When a letter is presented, it activates the set of features that are consistent with that letter. These featural detectors produce activation for the letter detectors that are consistent with those features, and inhibition for the letter detectors that are inconsistent with those features. Although there is some activation for words that are consistent with the letter and some inhibition for words that are inconsistent with the letter, this effect is relatively small because there is little influence of a single letter producing activation at the word level. Now, consider the condition wherein the letter is embedded in a word context. In a word context, there is now sufficient partial information from a set of letters to influence word-level activation patterns and this will produce a significant top-down influence onto letter-level representations, i.e., increase activation for consistent letters and decrease activation for the inconsistent letters. It is this higher-level activation and inhibition that overrides the deleterious influence of the patterned mask.

In passing, it is worth noting here that there is also evidence by Schendel and Shaw (1976) that suggests that features (e.g., lines) are better detected when the features are part of a letter than when presented alone. Hence, it is possible that there is also a letter superiority effect. Such an effect would appear to be easily accommodated within the McClelland and Rumelhart-type architecture by assuming that there are also top-down influences from the letter level to the feature level.

Interestingly, there is another phenomenon called the pseudoword superiority effect that would at first glance appear to be problematic for the McClelland and Rumelhart model. Specifically, letters are also better detected when embedded in pronounceable nonwords than when embedded in unpronounceable nonwords (Baron & Thurston, 1973; Carr et al., 1978), or presented in isolation (e.g., Carr et al., 1978; McClelland & Johnston, 1977). However, the interactive activation model can also accommodate this effect. Specifically, when letters are embedded in pronounceable nonwords, it is likely that there will be some overlap of spelling patterns between the pseudoword and acceptable lexical entries. For example, the pronounceable nonword *MAVE* activates 16 different four-letter words that share at least two letters within the McClelland and Rumelhart network. Thus, the influence of orthographic regularity appears to naturally fall out of the interaction across multiple lexical entries that share similar spelling patterns within the language. As we shall see below, the influence of orthographic regularity on word recognition performance has been central to many of the recent developments in word recognition research.

Although some orthographic regularity effects appear to naturally fall from this model, there are some additional intriguing insights from the model regarding other orthographic regularity effects. Consider, for example, the impact of bigram frequency. For example, the vowel pair *EE* occurs in many more words than the cluster *OE*. The available

evidence indicates that there is relatively little impact of bigram frequency on letter recognition within a Reicher-type paradigm (Manelis, 1974; McClelland & Johnston, 1977; Spoehr & Smith, 1975). McClelland and Rumelhart have successfully simulated this finding within their interactive activation framework. Although high-frequency letter clusters are more likely than low-frequency letter clusters to activate many word-level representations, this activation will be compensated by the fact that there will also be more word-level inhibition across those activated representations. Because, as noted above, there are influences of the number of lexical representations that share more than two letters, the lack of an influence of bigram frequency would appear to indicate that there may be a critical limit in the amount of overlap across lexical representations that is necessary to overcome the deleterious effects of within-level inhibition. (Bigram frequency also has very little influence on other lexical-processing tasks, such as naming or lexical decision; for example, see Andrews, 1992; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995.) One question that arises from this apparent lack of an influence of bigram frequency is why there are influences of neighbors only when the neighbors share more than two letters.

In addition to bigram frequency, one might ask whether positional frequency influences letter recognition. Positional frequency refers to the probability that a given letter(s) will occur in a given position within a word. Mayzner and Tresselt (1965) tabulated the summed positional frequency for single letters, bigrams, trigrams, tetragrams, and pentagrams (Mayzner, Tresselt, & Wolin, 1965a, 1965b, 1965c) across a set of 20,000 words. This metric should reflect the orthographic structure across words within a given language. In fact, one might expect influences of such a metric to fall quite nicely out of the McClelland and Rumelhart-type model. In fact, Massaro, Venezky, and Taylor (1979) reported evidence of a large impact of summed positional frequency within a Reicher-type paradigm. Their results indicated that both summed positional frequency and a rule-based metric of orthographic regularity (see discussion below) were found to influence letter recognition performance. Thus, at least at the level of letter recognition, there does appear to be an influence of positional letter frequency in a Reicher-type paradigm. Because letter position must be coded in the McClelland and Rumelhart model, one might expect this effect to naturally fall from the combined facilitatory and inhibitory influences across lexical-level representations. However, there are some limitations to such harsh coding. As discussed below in the section on orthographic neighborhood effects, the coding of position of letters within words has become a very active area of research recently.

In sum, the interactive activation model provides a cogent quantitative account of what appears to be evidence of multiple levels within the processing system working in concert to influence letter recognition (for an alternative view, see Massaro & Cohen, 1994). A particularly important aspect of this model is that “other” similar lexical-level representations appear to have an influence on the ease of recognizing a given letter within a word. It appears that letter- or word-level representations do not passively accumulate information, as in a logogen-type model (see Morton, 1969), but letters and words appear to be recognized in the context of similar representations that either reinforce or diminish the activation at a given representation. We shall now turn to some discussion of the dimensions that define “similarity” in such networks.

### 3. GETTING FROM LETTERS TO WORDS: INFLUENCES OF SUBLEXICAL LEVELS OF ORGANIZATION

The journey from letters to words has been a central concern in word recognition models. Although there are many distinct issues that arise in this area, one of the major theoretical issues has been the specification of the “rules” that are used in translating an orthographic pattern into an acceptable lexical/phonological representation. Unfortunately, as we shall see, such a translation process is far from easy in the English orthography.

#### 3.1. Specifying the “Rules” of Translation

One of the most evasive goals encountered in the analysis of English orthography is the specification of the functional unit(s) of sublexical organization. An obvious spelling-to-sound mapping might involve a simple one-to-one correspondence between graphemic units (single letters or letter clusters) and phonemes. Obviously, such an analysis fails relatively quickly in English because some graphemes, like PH, can serve as one phoneme in words like PHILOSOPHY, and two phonemes in a word like UPHILL. Likewise, even single letters are quite ambiguous such as the C in the word CAT and CIDER. English orthography simply does not allow a one-to-one mapping of spelling to sound.

Although a simple mapping of spelling to sound may not work for all words, it is still possible that one may gain considerable insight into the vast majority of words via an analysis of the regularities in the orthography. Such an enterprise was undertaken in a number of large-scale studies of English orthography in the late 1960s and early 1970s (e.g., Haas, 1970; Hanna, Hanna, Hodges, & Rudorf, 1966; Venezky, 1970; Wijk, 1966). For example, Venezky coded the grapheme-to-phoneme correspondences across a set of 20,000 medium- to high-frequency words. Through an in-depth analysis of the consistency of grapheme-to-phoneme patterns, Venezky distinguished between two large classes of grapheme-to-phoneme correspondences. Predictable patterns are those which can be based upon the regular graphemic, morphemic (minimal meaningful units, e.g., REDISTRIBUTION = RE + DISTRIBUTE + TION), or phonemic features of the words in which they occur, whereas, unpredictable patterns do not appear to fit within any predictable class (e.g., CHAMOIS). The important question is to what degree are patterns predictable when one considers similarities across words within the language. For example, some correspondences appear to be relatively invariant (predictable invariant patterns), e.g., the grapheme F always corresponds to the sound /f/ with the only exception being in the word OF. On the other hand, other graphemes have many variations, each of which appear to be relatively predictable (predictable variant patterns). For example, the letter C most typically corresponds to the phoneme /K/, but corresponds to the phoneme /S/ in many words when it is succeeded by the letter I, Y, or E.

As Henderson (1982) points out, there are a number of sublexical constraints within the grapheme-to-phoneme system in English, which are called phonotactic constraints. For example, because certain stop consonant sequences are not permissible in English

(e.g., /b/p/ and /p/b/), whenever one is confronted with such a sequence of letters (e.g., PB or BP) the correspondence is such that the first phoneme is silent (e.g., SUBPOENA). Thus, in this case, the phonological constraints of the language drive the grapheme-to-phoneme conversion of the spelling patterns. There also appear to be predictable constraints on the grapheme-to-phoneme mapping that are derived at the morphemic and syllabic levels. For example, the graphemic sequence MB corresponds to two separate phonemes when it segments syllables such as in *ambulance* and *amber*, but only one phoneme at word ending positions, such as in *tomb* and *bomb*. Unfortunately, as Henderson points out, the situation becomes somewhat more complex when one considers that MB also only corresponds to one phoneme when it precedes inflectional affixes (e.g., *bombing*), but not when it precedes other morphemes (*bombard*). Moreover, there appear to be other rule-type constraints that are simply based upon allowable grapheme-to-phoneme correspondences in particular positions within words. For example, the CK spelling pattern corresponds to the phoneme /K/, but the CK pattern does not occur at the beginning of words; in these later cases, the C to /K/ correspondence or the K to /K/ correspondence occurs. Using sophisticated permutation analyses, Kessler and Treiman (2001) have also shown that the spelling-to-sound consistency of a syllabic segment (i.e., onset, vowel, and coda) increases substantially when the other two segments are taken into account. These results may support the contention that English spelling is not as chaotic or irregular as popularly thought (Kessler & Treiman, 2003).

For demonstrative purposes, we have only touched upon some of the problems that one encounters in attempting to understand the regularity of spelling-to-sound correspondences in English orthography. Although ultimately it may be possible to specify such grapheme-to-phoneme rules in English, it is noteworthy that even with the relatively complex rule system developed by Venezky, and others, Coltheart (1978) estimated that 10–15% of the words would still be unpredictable, i.e., irregular. Likewise, Wijk (1969) notes that about 10% of the words will not fit his Regularized English. This may be an underestimate, because as Henderson points out, of the most common 3000 words, as many as 21% violate Wijk's regularization rules. Interestingly, Coltheart, Curtis, Atkins, and Haller (1993), using a learning algorithm to generate grapheme-to-phoneme rules, found that these rules mispronounced 22% of monosyllabic words, a figure which is consistent with Henderson's estimate. Also, because of computational limits inherent in two-layer networks (Hinton & Shallice, 1991), the two-layer network of Zorzi, Houghton, and Butterworth (1998) dual-process model was found to be incapable of learning exception words, and it failed to learn correct phonological codes for about 19% of its 2774 monosyllabic word corpus.

Of course, even if one could develop a rule-based system of spelling-to-sound translation that would accommodate all words in English, this would not necessarily indicate that such a rule-based system is represented in readers of English. In fact, even if such a rule-based system were represented, this would not be sufficient evidence to indicate that such rules are critical in fluent word recognition. Hence, instead of providing a detailed discussion of the enormously complex rule systems that have been developed to capture the mapping of orthography onto phonology in English, the present discussion will focus

on the empirical evidence regarding how readers use sublexical information in word recognition tasks. The interested reader is referred to Henderson (1982), Wijk (1966, 1969), and Venezky (1970) for excellent treatments of the search for rule-based translations of spelling-to-sound in English (for a description of the algorithms used to identify single-letter, multiletter, and context-sensitive rules, see Coltheart et al., 1993).

### 3.2. If Not Rules, Then What? The Controversy Regarding Dual-Route and Single-Route Models of Pronunciation

#### 3.2.1. Dual Route Perspective

If it is unlikely that there will be a limited number of rules that specify the translation from spelling-to-sound in English (i.e., an assembled route), it is possible that there is a second route (the lexical or direct route) that also plays a role in recognizing words. In the second, lexical, route the reader may map the orthographic string onto a lexical

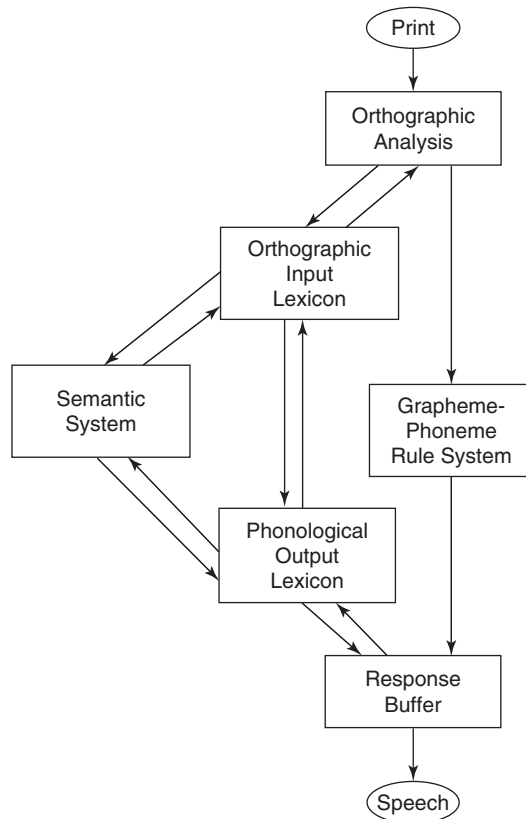


Figure 2. Coltheart et al.'s (2001) DRC model of word recognition.

representation and then access the programs necessary for pronouncing a given word aloud either directly from that representation or via access to a semantic representation. Figure 2 displays the dual-route cascaded (DRC) model of word reading developed by Coltheart, Rastle, Perry, Langdon, and Ziegler (2001). In their model, the lexical route is a straightforward extension to the interactive-activation model discussed above. One notable difference is that in the Coltheart et al. model, separate lexicons exist for orthography and phonology.

It is important to note here that because the world's orthographies differ with respect to the regularity of spelling-to-sound correspondences, orthographies also appear to differ with respect to the weight placed on the assembled and lexical routes. For example, if the alphabetic system in a given language is unequivocal in mapping orthography to phonology, as in a language such as Serbo-Croatian, then one might find little or no impact of the lexical route in speeded pronunciation performance (Frost, Katz, & Bentin, 1987). The reader can rely totally on the assembled route, because it always produces the correct response. However, in English, and even to a greater extent in other languages such as in Hebrew (e.g., Frost et al., 1987), the mapping between orthography and phonology is far less transparent. Hence, one should find increasing lexical effects in speeded pronunciation performance as one decreases the transparency of the spelling-to-sound correspondences (also referred to as the orthographic depth hypothesis). In support of this prediction, Frost et al. have reported larger frequency and lexicality effects in Hebrew compared to English which in turn produced larger effects compared to Serbo-Croatian. Similarly, there is evidence that readers of a shallow orthography like Serbo-Croatian make lexical decisions based on a prelexically computed phonological code; in contrast, phonological effects are relatively difficult to obtain in English lexical decision (Frost, 1998). Thus, comparisons across orthographies that differ with respect to the regularity of the spelling-to-sound correspondence support the notion that two routes are more likely in languages that have relatively deep orthographies.

If the inadequacy of a rule-based system demands a lexical route in English orthography, then one might ask what evidence there is for a role of an assembled route. Why would subjects ever use an assembled route to name a word aloud, if, by necessity, there must be a lexical route? One piece of evidence that researchers originally identified is the relative ease with which individuals can name nonwords (e.g., *blark*) aloud. Because nonwords do not have a direct lexical representation, it would appear that a nonlexical route is necessary for naming nonwords. However, this piece of evidence was soon disabled by evidence from activation-synthesis-type approaches (e.g., Glushko, 1979; Kay & Marcel, 1981; Marcel, 1980), in which the pronunciation of a nonword could be generated by the activation of similarly spelled words. Activation-synthesis theorists argued that pronunciation performance is always generated via analogies to words represented in the lexicon, thus minimizing an important role for the assembled route.

However, there is a second, and more powerful, line of support for the role of an assembled route in English that involves the performance of acquired dyslexics, who appeared to produce a double dissociation between the two routes. Specifically, one

class of dyslexics, surface dyslexics, appears to have a selective breakdown in the lexical route, but have an intact assembled route. These individuals are likely to regularize irregular words and exception words, e.g., they might pronounce *PINT* such that it rhymes with *HINT* (e.g., Marshall & Newcombe, 1980; McCarthy & Warrington, 1986; Shallice, Warrington, & McCarthy, 1983). A second class of acquired dyslexics, phonological (deep) dyslexics, appears to have an intact lexical route but an impaired phonological route. These individuals can pronounce irregular words and other familiar words that have lexical representations, however, when presented a nonword that does not have a lexical representation there is considerable breakdown in performance (Patterson, 1982; Shallice & Warrington, 1980). The argument here is that phonological dyslexics have a selective breakdown in the assembled route. Recently, Coltheart et al. (2001) simulated these two acquired dyslexias in the DRC model by selectively lesioning different components of the model. Specifically, surface dyslexia was simulated by lesioning the orthographic lexicon, while phonological dyslexia was simulated by dramatically slowing the sublexical process. These simulations nicely mimicked the neuropsychological data. For example, the degree of impairment of the orthographic lexicon produced regularization error rates that correlated highly with actual regularization error rates exhibited by surface dyslexics of varying severity. Furthermore, the model also correctly simulated the pseudohomophone advantage shown by phonological dyslexics. Specifically, these individuals pronounce pseudohomophones (e.g., BRANE) more accurately than non-pseudohomophonic nonwords (e.g., BRONE), reflecting the larger impact of the lexical route as the influence of the sublexical route is decreased (for a review of pseudohomophone effects in naming performance, see Reynolds & Besner, 2005a).

### 3.2.2. *Parallel distributed processing*

Although it would appear that there is compelling evidence for a dual-route architecture, there are important alternative models that have been developed by Seidenberg and McClelland (1989) and Plaut, McClelland, Seidenberg, and Patterson (1996) that also do an excellent job of handling some of the major findings that were originally viewed as strong support for the dual-route model. These parallel-distributed-processing (PDP) models could be viewed as a second generation of the original McClelland and Rumelhart (1981) model of letter recognition described above. One of the major differences between the two classes of models is that the later models were specifically developed to account for lexical tasks such as word pronunciation and the lexical decision task, whereas, the McClelland and Rumelhart model was developed in large part to account for letter recognition performance. A second major difference between the models is that the McClelland and Rumelhart model involves localized representations for the major processing codes (i.e., features, letters, and words), whereas, the later models involve distributed representations, e.g., there is not a single representation that reflects the word DOG. A third difference is that the McClelland and Rumelhart model assumes the existence of a specific architecture (i.e., sets of features, letters, and words along with the necessary connections), whereas, the latter models attempts to capture the development of the lexical processing system via the influence of a training regime. However, given

these differences, both models account for performance by assuming a flow of activation across a set of relatively simple processing units and have been detailed sufficiently to allow for mathematical tractability. We shall now turn to a brief introduction to the Seidenberg and McClelland model, which was the first in a series of parallel distributed processing models of word recognition.

As shown in Figure 3, the Seidenberg and McClelland model involves a set of input units that code the orthography of the stimulus and a set of output units that represent the phonology entailed in pronunciation. All of the input units are connected to a set of hidden units (units whose only inputs and outputs are within the system being modeled, i.e., no direct contact to external systems, see McClelland & Rumelhart, 1986, p. 48), and all of the hidden units are connected to a set of output units. The weights in the connections between the input and hidden units and the weights in the connections between the hidden units and phonological units do not involve any organized mapping before training begins. During training, the model is presented an orthographic string which produces some phonological output. The weights connecting the input and output units are adjusted according to the back-propagation rule, such that the weights are adjusted to reduce the difference between the correct pronunciation and the model's output. During training, Seidenberg and McClelland presented the model with 2897 English monosyllabic words (including 13 homographs, resulting in 2884 unique letter strings) at a rate that is proportional to their natural frequency of occurrence in English. The exciting result of this endeavor is that the model does a rather good job of producing the phonology that corresponds to regular words, high-frequency exception words, and even some nonwords that were never presented. Although there is clearly some controversy regarding the degree to which the model actually captures aspects of the data (e.g., see Besner, 1990; Besner, Twilley, McCann, & Seergobin, 1990), the fact that it provides a quantitative account of aspects of simple pronunciation performance (without either explicit Venezky-type rules or even a lexicon) is quite intriguing and it presented a powerful challenge to the available word-recognition models.

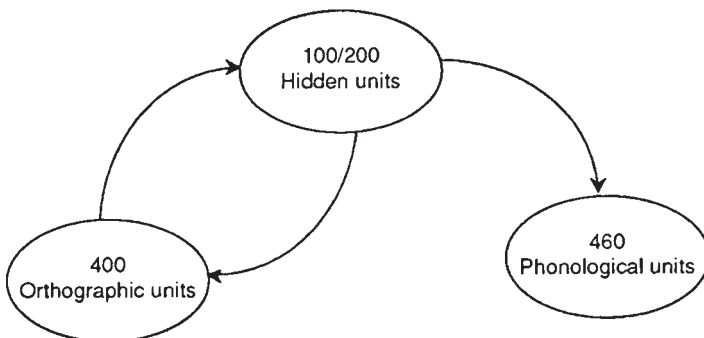


Figure 3. Seidenberg and McClelland's (1989) implemented connectionist architecture.



One of the more important results of the Seidenberg and McClelland model is its ability to capture the frequency by regularity interaction. This finding was initially viewed as rather strong support for a dual-route model (cf., Andrews, 1982; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Paap & Noel, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984a). The interaction is as follows: for high-frequency words, there is very little impact of the correspondence between orthography and phonology (but see Jared, 1997), whereas, for low-frequency words there is a relatively large impact of such a correspondence. The dual-route framework accommodated this finding by assuming that for high-frequency words the frequency modulated lexical route is faster than the frequency independent assembled route, and hence, any inconsistent information from the assembled route does not arrive in time to compete with the pronunciation that is derived from the lexical route. For example, the incorrect assembled pronunciation for the high-frequency word HAVE (such that it rhymes with GAVE) should not arrive in time to compete with the fast and correct lexical pronunciation. However, if one slows up the lexical route by presenting a low-frequency word (e.g., PINT), then one finds that the assembled output has time to interfere with the lexically mediated route and hence response latency is slowed down. The important point for the dual-route model is that the output of a low-frequency lexically mediated response can be inhibited by the availability of phonological information that is produced via the assembled route.

Although the dual-route model provides a natural account for this interaction, this pattern also nicely falls from the Seidenberg and McClelland single route model. That is, the error scores produced by the model (a metric that is assumed to map onto response latencies) for high-frequency regular words and exception words are quite comparable, however, for low-frequency words, the error scores are larger for exception words than for regular words. Thus, one does not have to assume separate routes (or even a lexicon) to handle the frequency by regularity interaction, because this pattern naturally falls from the correspondences between the frequency of a particular spelling-to-sound correspondence even in a relatively opaque alphabetic system such as English. The interaction between frequency and regularity for a specific set of words, and the predictions from Seidenberg and McClelland's model for this same set of words are displayed in Figure 4.

Interestingly, the spelling-sound consistency of a word's neighborhood also influences naming performance, and this neighborhood effect appears to produce an additional influence above and beyond the grapheme-to-phoneme regularity (Glushko, 1979; Jared, McRae, & Seidenberg, 1990). Consistency refers to the degree to which similarly spelled words are pronounced similarly. In particular, studies of consistency have focused on the rime (i.e., the vowel and subsequent consonants in a monosyllabic word). A word that shares both the orthographic rime and phonological rime with most or all of its neighbors is relatively consistent, whereas a word that shares the orthographic rime with its neighbors but has a different pronunciation than most of its neighbors is relatively inconsistent. Regular words that have many "friends" (e.g., *spoon* is consistent because of *moon*, *noon*, etc.) are named faster than regular words that have many "enemies" (e.g., *spook* is inconsistent because of *book*, *took*, etc.). Jared et al. (1990) provided evidence that there are consistency effects in pronunciation primarily under conditions when the neighbors

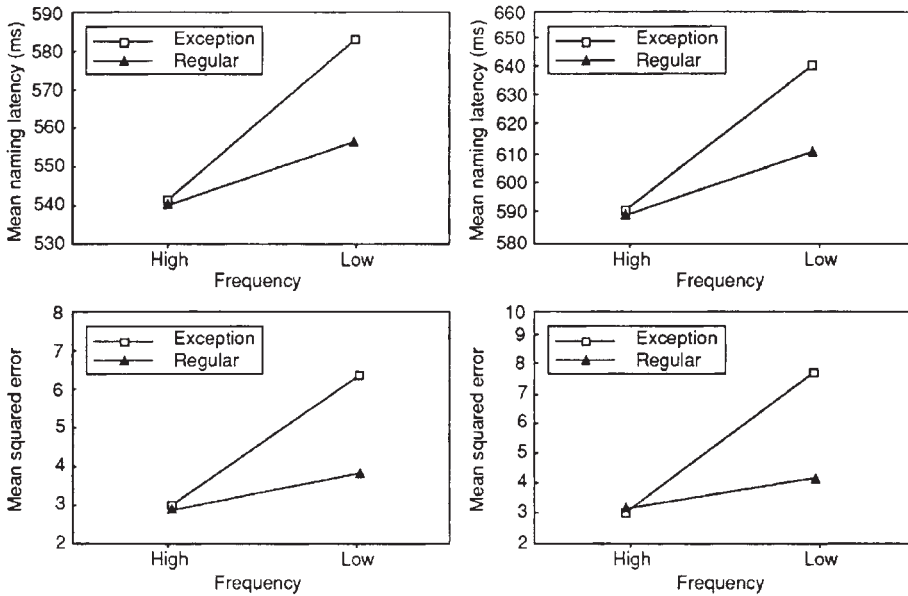


Figure 4. Results and simulations of the Seidenberg (1985, left graph) and Seidenberg et al. (1984a, Experiment 3, right graph) studies: Experimental results (upper graphs) and simulations from the Seidenberg and McClelland (1989) model (lower graphs).

that have consistent spelling patterns (i.e., friends) are higher in frequency than the neighbors that have inconsistent spelling patterns (i.e., enemies). Such neighborhood frequency effects would appear to fall quite nicely from the Seidenberg and McClelland (1989) model. Alternatively, a rule-based model might suggest that the consistency of the neighbors defines the rules of translation from orthography to phonology. However, because of the difficulties noted above in specifying such rules, it is appealing that the Seidenberg and McClelland model can capture such neighborhood effects, without the appeal to rules.

### 3.2.3. Regularity vs. consistency revisited

Because many irregular words (i.e., words whose pronunciation violates grapheme-phoneme correspondence (GPC) rules) are also inconsistent at the rime level, regularity and consistency have typically been confounded. However, these two dimensions are indeed separable (e.g., Andrews, 1982; Kay & Bishop, 1987). Obviously, distinguishing regularity and consistency is important in testing contrasting predictions of models of word recognition. Specifically, the DRC model predicts large effects of regularity and small effects of consistency, and PDP models predict small effects of regularity and large effects of consistency. In general, the results of the studies that have distinguished between consistency and regularity have shown that rime consistency has a larger influence

than regularity on latencies and errors than regularity (for discussion, see Cortese & Simpson, 2000; Jared, 2002). In fact, Cortese and Simpson found that the PDP model of Plaut et al. (1996) simulated the naming data on a selected set of words that crossed regularity and consistency better than the Coltheart et al. (2001) DRC model.

### 3.2.4. *Regularity vs. consistency in words and nonwords*

Of course, consistency is a continuous variable that can be measured at various levels (e.g., rimes, graphemes). In large-scale studies, Treiman and colleagues (Treiman, Kessler, & Bick, 2002; Treiman et al., 1995) have found that rime-level consistency is a better predictor of word naming performance than grapheme-to-phoneme level consistency. However, it appears that for nonword naming performance, the pattern is a bit more complicated. For example, in contrast to the results by Treiman and colleagues regarding word naming performance, Andrews and Scarratt (1998) reported that nonword reading is affected more by consistency at the grapheme-to-phoneme level than by rime-level consistency. Moreover, in their analysis of 20 nonwords (taken from Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994) in which regularity and consistency pull in opposite directions, Cortese and Simpson (2000) found that grapheme-to-phoneme rules predicted the preferred pronunciation in 14 nonwords, whereas rime consistency predicted the preferred pronunciation in only 5 nonwords. Consider *jind*. The GPC rule for *i* is /I/, but consistency favors the /aInd/ pronunciation found in *bind*, *blind*, *hind*, *mind*, etc. Seidenberg et al. found that 23 of 24 participants pronounced *jind* in a fashion that is consistent with GPC rules. Therefore, it is quite possible that subjects may rely on different types of information when pronouncing a set of nonwords than when processing words.

Zevin and Seidenberg (2006) have recently claimed that consistency effects in nonword naming tasks are more consistent with the PDP perspective than the DRC perspective. By varying the training experience with each new run of a PDP model (also see Harm & Seidenberg, 1999), the model could be tested in terms of pronunciation variability (i.e., the degree to which pronunciations vary across subjects) that is exhibited by college readers (Andrews & Scarratt, 1998; Treiman et al., 2002). Both the PDP model and college readers exhibited considerable variability in their pronunciation of nonwords derived from inconsistent words (e.g., *thead*, *moup*), but not in their pronunciation of nonwords derived from consistent words (e.g., *nust*). This characteristic is difficult to assess in the DRC model because it is not clear how rules are acquired in the most recent model and how different versions of the model could be implemented. In addition to the insights provided regarding consistency effects in nonword naming, the extension of the models to individual variability, as opposed to overall mean performance, is an important next step in model development.

### 3.2.5. *Feedback consistency*

Heretofore, we have been primarily discussing the directional *feedforward* mapping of orthography onto phonology in our consideration of regularity and consistency effects.

For example, PINT is feedforward inconsistent because it does not rhyme with its orthographic neighbors (e.g., *mint*, *hint*, *tint*, etc.). However, there is another form of mapping which reflects a *feedback* influence. Specifically, feedback consistency reflects the manner in which a specific phonological pattern is spelled in different ways. Figure 5 illustrates the syllabic structure derived by linguistic distinctions between onsets and rimes (see further discussion below) and also shows how consistency can be computed along four dimensions: (a) feedforward onset, (b) feedforward rime, (c) feedback onset, and (d) feedback rime. For example, the rime in *tone* is feedback inconsistent because /on/ is spelled OWN as in GROWN, and OAN, as in MOAN. As one might guess, many words are inconsistent in both directions. Stone, Vanhoy, and Van Orden (1997) first decoupled feedforward consistency from feedback consistency and the effects of both variables were obtained in lexical decision performance. In addition, reliable and equivalent feedback consistency effects were reported by Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) for lexical decision and naming performance, whereas in French, Ziegler, Montant, and Jacobs (1997) found larger feedback consistency effects in lexical decision than in naming performance. The influence of feedback consistency in visual word recognition is theoretically important because it suggests that phonological activation provides feedback onto the orthographic representation (also see Pexman, Lupker, & Jared, 2001) during isolated visual word processing. However, it should also be noted that there is currently some debate regarding the unique effect of feedback consistency. For example, Peereman, Content, and Bonin (1998) have argued that feedback consistency effects in French are eliminated when familiarity is controlled (also see Kessler, Treiman, & Mullennix, 2005).

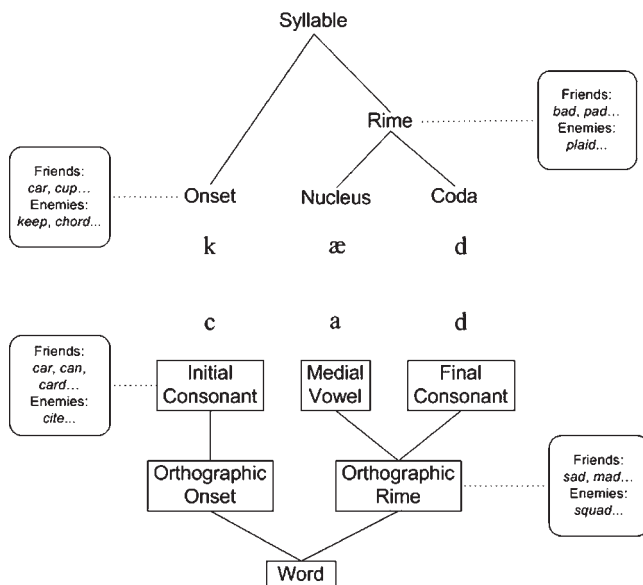


Figure 5. Feedforward and feedback onset and rime organization for single syllabic structure.

### 3.2.6. *Potential problems with Seidenberg and McClelland model*

Consistency effects would appear to arise naturally from the PDP architecture developed by Seidenberg and McClelland (1989). Although this model provided an interesting alternative to the dual-route model, it also generated a number of important problems that needed to be resolved (for a discussion of these issues, see Coltheart et al., 1993). First, it is unclear how such a model might handle the fact that some acquired dyslexics appear to only have an intact assembled route, while others appear to only have an intact lexical route (for some discussion of this issue, see Patterson, Seidenberg, & McClelland, 1989). Second, as described below, it appears that meaning-level representations can influence pronunciation and lexical decision performance. Thus, without some level of semantic input, it is unclear how an unembellished Seidenberg and McClelland model could account for such effects. Third, Besner (1990) and Besner et al. (1990b) have documented that the phonological error scores and the orthographic error scores do a rather poor job of simulating some characteristics of nonword performance. Fourth, the Seidenberg and McClelland model mapped error scores indirectly onto response latency instead of providing a direct metric for response latency.

### 3.2.7. *Further developments in the PDP architecture*

In response to the challenges to the Seidenberg and McClelland (1989) model, Plaut et al. (1996) substantively updated the representations and architecture of the PDP model (henceforth PMSP96) to address these problems. First, the model incorporated improved orthographic and phonological representations that allow it to not only correctly pronounce all the monosyllabic words in the training corpus, but also to name nonwords with a much greater facility. Second, in Seidenberg and McClelland's (1989) "triangle" model framework (see Figure 6), skilled reading is supported by the joint contributions of a phonological and a semantic pathway. Although only the phonological pathway was implemented in the Seidenberg and McClelland model, a prototype semantic pathway was implemented in PMSP96 (see Simulation 4), which may be useful in accommodating meaning-level influences in word recognition (see Strain, Patterson, & Seidenberg, 1995). Third, the authors extensively discussed how the network could handle the acquired dyslexia data that, as described above, was central to the development of the DRC model. For example, it is possible to simulate phonological dyslexia, i.e., better word reading than nonword reading, by a selective impairment of the phonological pathway. Similarly, surface dyslexia, i.e., normal nonword but impaired exception word reading, was satisfactorily simulated after training a new network that incorporated an isolated, semantically supported phonological pathway. In normal readers, the semantic and phonological pathway work together to support the pronunciation of exception words. Should the semantic pathway be damaged, the semi-competent isolated phonological pathway manifests symptoms similar to that of surface dyslexia (Plaut, 1997). Of course, at this point one might ask whether the inclusion of a semantic "route" makes the PDP model functionally equivalent to a dual-route model. For example, does the network, over the course of training, partition itself into two sub-networks, one that handles regular words, and one that handles exception words? Plaut et al. (1996) tested this intriguing hypothesis and found little support for this contention. Generally, the system did not fractionate

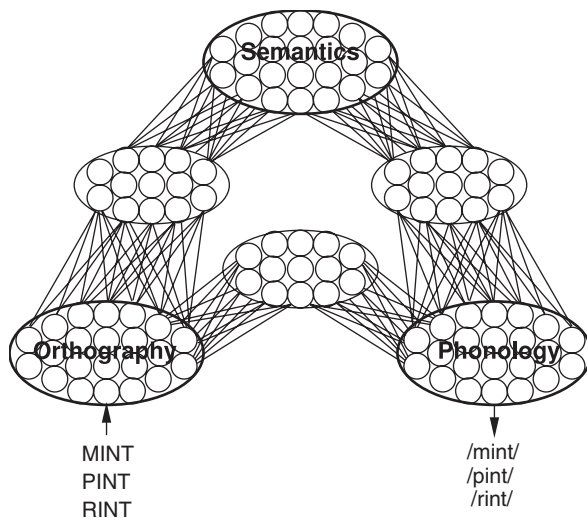


Figure 6. Seidenberg and McClelland's (1989) triangle connectionist framework for lexical processing.

itself such that one part learned spelling-sound rules and another part encoded the exceptions to these rules, but both components contributed to performance. Finally, it is important to note that the Plaut et al. model was a recurrent network that eventually settled into a steady state and hence response latencies could be evaluated in the model. That is, activation of phonological units changes over time as information in the system accumulates and is shared among network units. This property contrasts with the error measure that was used to evaluate the original Seidenberg and McClelland model. In the Plaut et al. model, when a word is recognized, its corresponding grapheme units become activated, and, in turn, this activation is propagated throughout the network.

Considerable debate continues between advocates of PDP and DRC approaches to word recognition. Although the PDP models seem ideally suited for handling consistency effects, the DRC model is particularly adept at handling data consistent with serial processing. Consider, for example, the *position of irregularity effect*. An irregular/inconsistent word can be irregular/inconsistent at the first phoneme position (e.g., *chef*), the second phoneme position (e.g., *pint*), the third phoneme position (e.g., *plaid*), or beyond (e.g., *debris*). Because the sublexical process in the DRC operates in a serial fashion, it is more susceptible to earlier than later irregular/inconsistent sublexical interference. In contrast, the PDP model processes words in parallel, and so does not predict a position of irregularity effect. Although there have been some methodological concerns noted (see Cortese, 1998; Zorzi, 2000), the evidence indicates that latencies are longer for words that have early irregular/inconsistent patterns than late irregular/inconsistent patterns (e.g., Coltheart & Rastle, 1994; Cortese, 1998; Rastle & Coltheart, 1999). These results, along with others (see Rastle & Coltheart, 2006), appear to support the serial component of the DRC-type framework over the parallel nature of the PDP framework. Of course, it is possible that such effects may ultimately reflect input or output processes beyond the

scope of the currently implemented PDP framework. Indeed, Seidenberg (2005) acknowledges that important challenges exist, but the PDP approach accounts for many behavioral phenomena as well as providing a more natural interface between reading and underlying principles of the nervous system. In this light, Seidenberg argues that the PDP architecture is important because it generalizes well to other cognitive domains. However, a proponent of the DRC approach would argue that such generality should not outweigh the fact that the devil is in the details of the fit of a particular model of word recognition with the available evidence (see, for example, Rastle & Coltheart, 2006). Clearly, the debate continues.

More recent connectionist models of reading have shifted their emphasis from understanding how people *pronounce* letter strings aloud to understanding how *meaning* is computed (Seidenberg, 2005). For example, Harm and Seidenberg (2004) proposed a model that considers how the meaning of a word is computed by orthographic and phonological processes working cooperatively. It is also apparent that one glaring limitation of both dual-route and connectionist models is their inability to process multisyllabic words. One model that has made some progress in this respect is the connectionist multi trace memory model of Ans, Carbonnel, and Valdois (1998). While a full description of this interesting model is beyond the scope of this chapter, the Ans et al. model proposes two sequential procedures for reading: first, a holistic procedure that draws on knowledge about entire words, and if that fails, an analytic procedure that is dependent on the activation of subsyllabic segments. Including two reading procedures allows the model to name monosyllabic words, multisyllabic words, and nonwords, and also allows it to account for dissociations between skilled and pathological reading. Although the Ans et al.'s model may *prima facie* resemble the dual-route model, it does not compute phonology from orthography using different computational principles. Instead, pronunciation is always supported by the memory traces laid down by previously encountered exemplars.

### 3.3. Superletter Sublexical Codes : What's the Evidence for their Functional Role?

At this point, it should be noted that we have yet to discuss specific types of sublexical but supraletter influences on word recognition. We have generally grouped together a set of effects under the regularity/consistency umbrella, focusing on the theoretical implications of such effects for current models. We shall now turn to a brief discussion of three distinct levels of sublexical representation that have been at the center of this area of research: onsets and rimes, morphology, and syllables. The goal here is to simply acquaint the reader with the attempts that have been used to decompose the sublexical units.

#### 3.3.1. Onsets and rimes

As noted earlier, researchers have made a distinction between the onset and rime unit within syllables. For example, Treiman and her colleagues (e.g., Treiman, 1989; Treiman & Chafetz, 1987; Treiman & Danis, 1988; Treiman & Zukowski, 1988) have argued that there is an intermediate level of representation in lexical processing between graphemes and syllables (also see Kay & Bishop, 1987; Patterson & Morton, 1985). They argue that

syllables are not simply strings of phonemes but there is a level of subsyllabic organization that is used both in speech production and recognition of visual strings. This subsyllabic distinction is between the onset and rime of a syllable. The onset of a syllable can be identified as the initial consonant or consonant cluster in a word. For example, /s/ is the onset for sip, /sl/ is the onset for slip, and /str/ is the onset for strip. The rime of a word involves the following vowel and any subsequent consonants. For example, in SIP, SLIP, and STRIP, /ɪp/ would be the rime. Thus, syllables have a subsyllabic organization in that each syllable is composed of an onset and a rime.

Although our primary interest is in visual word processing, it is interesting to note that there has been evidence from a number of quite varied research domains that supports the distinction between onsets and rimes in English. For example, there is evidence for this distribution from the types of speech errors that speakers produce (Dell, 1986; MacKay, 1972), the distributional characteristics of phonemes within syllables (Selkirk, 1982), along with the types of errors that subjects produce in short-term memory tasks (Treiman & Danis, 1988). Thus, the support for the onset and rime distinction clearly extends beyond the work in visual word recognition, and is driven more by phonological principles that have been developed in linguistics.

In one of the first studies addressing onset and rime organization in visual word recognition, Treiman and Chafetz (1987) presented strings like FL OST ANK TR to subjects with the task being to determine whether two of the strings in these four strings of letters could be combined to form a real word. In this case, one can see that FL and ANK can be combined to produce FLANK, with FL corresponding to the onset of the word FLANK and ANK corresponding to the rime. Now, consider performance in conditions where the strings again correspond to words but they are not broken at onsets and rimes. For example, a subject might be presented FLA ST NK TRO. For these items, the correct answer is again FLANK, but now the FLA and NK do not correspond to onsets and rimes. The results of the Treiman and Chafetz experiments indicated that anagram solutions were better when the breaks corresponded to onset-rime divisions compared to when the breaks did not. A similar pattern was found in a lexical decision task. In this study, the items were again presented such that there was either a break that matched the onset-rime division (e.g., CR//ISP, TH//ING) or a break that did not match the onset-rime division (e.g., CRI//SP and THI//NG). The results indicated that lexical decisions were reliably faster when the break matched the onset-rime division. Thus, Treiman and Chafetz argued that onset and rime units play a role in visual word recognition.

### 3.3.2. *Syllables*

If the distinction between onsets and rimes plays a functional role en route to word recognition then one would also expect a functional role for the syllable. At this level, it is quite surprising that there has been considerable disagreement regarding the role of the syllable in visual word recognition. For example, Spoehr and Smith (1973) argued for a central role of the syllable, whereas, Jared and Seidenberg (1990) have questioned the role of the syllable as a sublexical unit. In fact, as Seidenberg (1987) points out there is



even some disagreement regarding where syllabic boundaries exist. For example, according to Howard's (1972) rules that emphasize intrasyllabic consonant strings surrounding a stressed vowel, CAMEL would be parsed as (CAM)+(EL), whereas, according to Selkirk's (1980) more linguistically based view that emphasizes the maximal syllable onset principle CAMEL would be parsed (CA)+(MEL). Obviously, before one can address the functional role of the syllable in visual word recognition, one must have some agreement on how to parse words into syllables. Fortunately, for the majority of words, there is agreement on how words are parsed into syllables.

The question here of course is whether a word like ANVIL is parsed into (AN)+(VIL) en route to word recognition. It should again be emphasized here that the concern is not whether subjects have access to syllabic information, surely they must, i.e., most subjects can accurately decompose most words into syllables. The more important issue is whether this information is used in accessing the lexicon for visually presented words.

Prinzmetal, Treiman, and Rho (1986) reported an intriguing set of experiments that investigated the impact of syllabic structure on early level perceptual operations in word recognition. These researchers used a paradigm developed by Treisman and Schmidt (1982) in which feature integration errors are used to examine perceptual groupings. The notion is that if a set of strings (e.g., letters or digits) forms a perceptual group then one should find migration of features (e.g., colors) toward that group. In the Prinzmetal et al. study, subjects were presented with words such as ANVIL and VODKA. At the beginning of each trial, subjects were given a target letter with the task being to report the color of the target letter that would appear in the upcoming display. After the target letter was designated, subjects were presented a letter string with each of the letters in different colors. The data of interest in such studies are the types of errors that subjects make as a function of syllabic structure. Consider the third letter position in the words ANVIL and VODKA. In the word ANVIL the third letter is part of the second syllable, whereas, in the case of VODKA the third letter is part of the first syllable. Now, if the syllable produces a perceptual grouping, then one might expect errors in reporting the colors such that the D in VODKA might be more likely to be reported in the same color of the O, compared to the K, whereas, the V in ANVIL might be more likely to be reported in the color of the I, compared to the N. This is precisely the pattern obtained in the Prinzmetal et al. study.

It is interesting to note here that Adams (1981) provided evidence that the letters that border adjacent syllables often have relatively low bigram frequencies. In fact, the NV and DK are the lowest bigram frequencies in the words ANVIL and VODKA. In general, if one considers relatively high-frequency bisyllabic words, there appears to be a decrease in frequency of the bigrams that occur at syllabic boundaries. This bigram trough may actually increase the likelihood of feature errors, due to the frequency of the orthographic neighbors of the target instead of an actual subsyllabic parsing en route to word recognition. Although Seidenberg (1987, Experiment 3) provided some initial evidence that the effects observed in the original Prinzmetal et al. paradigm were due to such bigram troughs, as opposed to actual syllabic boundaries, more recent work by Rapp (1992) found that one can obtain syllabic effects even when one controls for such bigram troughs.

The role of the syllable has not been implemented in most models of word recognition that have been primarily built to process monosyllabic words. One exception to this is the connectionist model proposed by Ans et al. (1998), discussed earlier. Based on the evidence discussed above and the findings from the literature on spoken word processing (e.g., Stevens & Blumstein, 1978), this model parses words into syllabic units in the phonological output. Presumably, this phonological output could serve as an access to a semantic system; however, this was not implemented in the current model.

More recently, Rastle and Coltheart (2000) have proposed a complex set of rules for syllable segmentation, stress assignment, and vowel reduction for disyllabic words in their DRC model. In their study, the assignment of stress to a set of nonwords by the model was similar to that provided by human subjects. Also, words that violated the rules resulted in longer naming latencies, an effect that is consistent with predictions of the DRC model. However, it is important to note that, like previous studies on regular and irregular monosyllabic words, regularity in the Rastle and Coltheart study was confounded with spelling-sound consistency (Chateau & Jared, 2003). In their naming study of disyllabic words, Chateau and Jared found that the feedforward consistency of the segment containing the first vowel grapheme and subsequent consonants and the second vowel grapheme predicted naming latencies and errors. Moreover, the consistency measures derived by Chateau and Jared nicely predicted the outcome reported by Rastle and Coltheart. Of course, if readers use sublexical rules when processing multisyllabic words, the DRC model would be better equipped to explain such a result, but consistency effects are better handled by PDP models. Clearly, more research on multisyllabic words is necessary to determine both the behavioral influence of syllables and stress patterns en route to word recognition and also the best way to model such effects.

### 3.3.3. *Morphemes*

Another sublexical unit that has received considerable attention in the literature is the morpheme. One of the most compelling reasons that morphemes might play a functional role in word recognition is the generative nature of language. Rapp (1992) provides CHUMMILY as an interesting example. Although we may have never encountered the nonword CHUMMILY, we may assume that it means something like in a chummy way or friendly because it appears to have the morphological form CHUMMY + LY. Linguistic models of lexical representation assume that there is some base form of representation and a set of rules that are used to construct other forms of that item. The present question is whether a given form of a word such as JUMPED is parsed as (JUMP)+(ED) en route to word recognition. As in the case of syllables, we are not questioning whether morphemes are represented in the processing system, the question is whether morphemic analyses play a role in processes tied to visual word recognition.

Much of the early theoretical and empirical work regarding the role of the morpheme in visual word recognition was originally developed by Taft and Forster (1975, 1976; also see Taft, 1979a, 1979b, 1985, 1987). They argued that readers first decompose polymorphemic words into constituent morphemes. Readers then access lexical files

that are listed under the root morpheme. For example, if the word CHARACTERISTIC was presented, the reader would first access the root word CHARACTER and once this root word was accessed the subject would search through a list of polymorphemic words with the same root morpheme, e.g., CHARACTERISTIC, UNCHARACTERISTIC, CHARACTERIZED, CHARACTERISTICALLY, UNCHARACTERISTICALLY, etc. There have been a number of studies reported in the literature that support the notion that there is a morphemic level of analysis in visual word recognition. For example, Taft (1979a, 1979b) found an effect of printed word frequency of the root morpheme (the sum of frequencies of all words with a given root) in lexical decision performance for items that were equated in surface frequencies (see, however, caveats by Bradley, 1979). This would appear to support the contention that root morphemes do play a special role in word recognition and it is not simply the raw frequency of the actual lexical string that is crucial.

Another approach to morphological analyses in word recognition involves long-term morphemic priming (e.g., Stanners, Neiser, & Painton, 1979a). In these studies, subjects are most often presented a sequence of lexical decision (word/nonword) trials. At varying lags within the sequence, subjects might be presented two forms of a given word with the same root. The interesting comparison is the influence of an earlier presentation of a given root form on later lexical decisions to the actual root. For example, if either JUMP or JUMPED is presented earlier in a lexical decision task, what impact does this presentation have on later lexical decision performance on the root form JUMP? Stanners, Neiser, Herson, and Hall (1979b) found that both JUMP and JUMPED equally primed later lexical decisions to JUMP. Presumably, subjects had to access JUMP to recognize JUMPED and hence there was as much long-term priming from JUMPED as for the actual stem itself. Interestingly, Lima (1987) has found that mere letter overlap does not produce such an effect. For example, she reported that ARSON does not prime SON, but DISHONEST does prime HONEST. Thus, it does not appear that mere letter overlap is producing this long-term priming effect (for a summary of evidence favoring non-orthographic accounts of morphemic priming effects, see review by Feldman & Andjelkovic, 1992).

Because the PDP perspective has achieved prominence as a general theory of language processing, research on morphological decomposition has taken on new theoretical significance. One main reason that this topic has received such attention is that distinct morphemic representations do not exist in PDP models (e.g., Plaut & Gonnerman, 2000; Rueckl, Mikolinski, Raveh, Miner, & Mars, 1997). Rather, morphemic effects are thought to emerge from interactions among orthography, phonology, and semantics (Gonnerman, Seidenberg, & Andersen, 2005). A recent cross-modal lexical decision study by Gonnerman et al. (2005) found support for this view. They reported that facilitation for visually presented targets was related to the semantic and phonological overlap found in prime–target pairs. In contrast, morphemic overlap did not produce additional facilitation above and beyond semantically and phonologically related items. For example, *sneer* facilitated *snarl* to the same degree as *teacher* facilitated *teach*. Also, weakly related pairs (e.g., *lately-late*) produced less facilitation than strongly related pairs.

Interestingly, Rastle, Davis, and New (2004) reported a morphological effect that was independent of semantics. In their lexical decision study, masked primes (presented for 42 ms) that maintained a morphological relationship only (e.g., *corner-corn*) facilitated targets as much as primes that maintained both a semantic and morphological relationship with the target (e.g., *cleaner-clean*), whereas a control condition (e.g., *brothel, broth*) did not produce priming. Thus, it appears from the Rastle et al. study that decomposition is somewhat independent of the semantic information available from the stem. This outcome seems more consistent with localist models (e.g., the DRC model) than distributed models (e.g., PDP models). However, given the Gonnerman et al. results discussed above, it is clear that further work is needed on this important topic.

We have only touched upon some of the very interesting issues that have arisen in morphological analyses in visual word recognition. We suspect that this will be an area of very active research in the future, and refer the reader to Baayen and Schreuder (2003), Feldman and Basnight-Brown (2005), and Sandra and Taft (1994) for more comprehensive treatments of this important area.

#### 4. LEXICAL-LEVEL VARIABLES

By lexical-level variables, we refer to the impact of variables that have been quantified at the whole word level. For example, word frequency is a lexical variable. Specifically, a researcher can investigate the influence of the printed frequency of a given word (e.g., DOG vs. SILO) on word recognition task performance.

##### 4.1. Length

One might ask whether there is a word length effect in visual word recognition tasks, as measured by the total number of letters in a given word. Obviously, if the letter is a crucial player in word recognition then one should find consistent effects of letter length. Interestingly, there has been some disagreement on this simple topic. There is clear evidence that longer words take more time in perceptual identification (McGinnies, Comer, & Lacey, 1952), and produce longer fixation durations in reading (see Just & Carpenter, 1980), but the effect of length in lexical decision and naming performance has been a bit more inconsistent (for a review, see New, Ferrand, Pallier, & Brysbaert, 2006).

The role of letter length in naming performance has been the focus of a number of recent studies. For example, Gold et al. (2005) found that individuals with a loss of semantic/lexical input, produced exaggerated length effects, compared to individuals with dementia of the Alzheimer's type. Gold et al. suggested that these results may be supportive of greater reliance on the serial sublexical route in individuals with semantic dementia. Consistent with this possibility, Weekes (1997) found length effects for nonwords and no length effects for words. Coltheart et al. (2001) interpreted the Weekes results as being critical to the DRC, i.e., the small or non-existent length effects for words is due to the parallel pathway used in the lexical route, whereas, the large length effects

for nonwords reflects the serial analysis demanded by the sublexical route. In a study of speeded naming performance of over 2400 single syllable words, Balota et al. (2004) obtained clear effects of length that were modulated by word frequency. Moreover, low-frequency words produced larger length effects than high-frequency words.

There is some controversy regarding length effects in the lexical decision task. Because the lexical decision task has been taken as a premier task to develop word recognition models, this is a troublesome finding (for a review, see Henderson, 1982). Chumbley and Balota (1984) reported relatively large length effects in the lexical decision task when the word and nonwords were equated on length and regularity. It is possible that inconsistent results with respect to past word-length studies using the lexical decision task may have been due to using a relatively small range of lengths of words. In this light, the recent study by New et al. (2006) is noteworthy. Specifically, they analyzed length effects in a dataset of lexical decision latencies to 33,006 words taken from Balota et al. (2002). They found an interesting quadratic relationship between length and lexical decision performance, such that there was a facilitatory effect from 3 to 5 letter in length null effect for 5–8 letters in length and a clear inhibitory effect for 8–13 letter words. The long words appear to demand some serial processing. Interestingly, the short words indicate that there may be an ideal length, based on the average length of words, and that very short words actually may produce a decrement in performance. Finally, it should also be noted that frequency does appear to modulate the length effect, since Balota et al. (2004) reported that length effects were larger in lexical decisions for low- than high-frequency words, similar to the pattern obtained in speeded naming performance mentioned above. Thus, the effects of word length in lexical decision performance appear to depend on both the frequency and the particular lengths of the words.

#### 4.2. Word Frequency

The frequency with which a word appears in print has an influence on virtually all word recognition tasks. For example, word frequency effects have been found in lexical decision performance (e.g., Forster & Chambers, 1973), naming performance (e.g., Balota & Chumbley, 1984), perceptual identification performance (e.g., Broadbent, 1967), and online reading measures such as fixation duration and gaze duration measures (e.g., Rayner & Duffy, 1986; Schilling, Rayner, & Chumbley, 1998). This, of course, should not be surprising because printed word-frequency should be related to the number of times one experiences a given word; experience with an operation should influence the ease of performing that operation.

Although it would appear to be obvious why word-frequency modulates performance in word recognition tasks, the theoretical interpretations of such effects have been quite varied. For example, the activation class of models based in large part on Morton's (1969, 1970) classic Logogen model, assume that frequency is coded via the activation thresholds in word recognition devices (logogens). High-frequency words, because of the increased likelihood of experience, will have lower activation thresholds than low-frequency words. Therefore, in order to surpass a word recognition threshold, the

activation within such a logogen will need to be boosted by less stimulus information for high-frequency words than for low-frequency words. Coltheart et al.'s (2001) DRC model nicely captures frequency effects via the activation patterns in the lexical route. The PDP models of Seidenberg and McClelland (1989) and Plaut et al. (1996) assume that frequency is coded in the weights associated with the connections between the units. Interestingly, there are hybrid models (e.g., Zorzi et al., 1998), which implement lexical and sublexical processing using connectionist principles, and so frequency effects could arise in both pathways.

A third class of word recognition models that we have yet to describe are referred to as ordered search models (e.g., Forster, 1976, 1979; Rubenstein, Garfield, & Millikan, 1970). According to these models, the lexicon is serially searched with high-frequency words being searched before low-frequency words. For example, as shown in Figure 7, Forster (1976) has argued that the lexicon may be searched via several indexing systems: orthographic, phonological, and syntactic/semantic access bins. Each of these bins involves a frequency ordered search, i.e., high-frequency words are searched before low-frequency words, and once the target is located the subject has immediate access to the word's master lexicon representation. Although such a model may seem cumbersome, Murray and Forster (2004) have recently provided intriguing evidence supporting this position, since rank frequency (as in rank in the search bin) appears to be a better predictor of word-frequency effects than actual log frequency values. It is noteworthy that there are additional models that are hybrids of the activation and search models such as in the Becker (1980), Paap et al. (1982), and the Taft and Hambly (1986) models. For example, Becker suggests that activation processes define both sensorily and semantically defined search sets. These search sets are then compared to the target stimulus via a frequency-ordered search process.

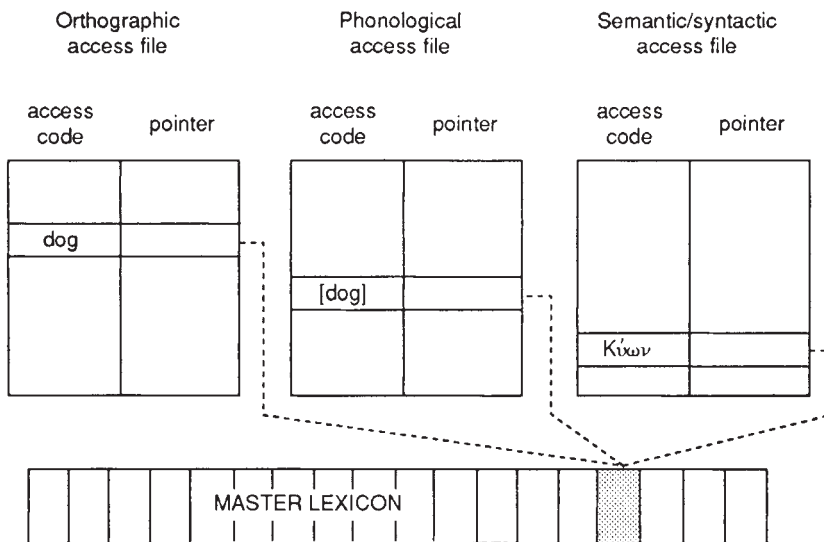


Figure 7. Architecture of Forster's (1976) serial search model of word recognition.

An important question that has arisen regarding word frequency effects is the locus of the effect in the tasks used to build models of word recognition. The models mentioned above all suggest that frequency is central to the interworkings of the models, as it should be. However, there is also evidence that suggests there are (a) decision components of the lexical decision task (Balota & Chumbley, 1984; Besner, Davelaar, Alcott, & Parry, 1984; Besner & McCann, 1987), (b) post-access components related to the generation and output of the phonological code in the pronunciation task (Balota & Chumbley, 1985; Connine, Mullennix, Shernoff, & Yelens, 1990), and (c) sophisticated guessing aspects of the threshold identification task (Catlin, 1969, 1973) that are likely to exaggerate the influence of word frequency. Because of the importance of task analyses, we will use this as an opportunity to review some of these issues regarding the lexical decision task.

Consider, for example, the Balota and Chumbley (1984) model of the lexical decision task displayed in Figure 8. Balota and Chumbley have suggested that because of the demands of the task, subjects place particular emphasis on two pieces of information that are obvious discriminators between words and nonwords, i.e., the familiarity and meaningfulness (FM dimension) of the stimuli. Nonwords are less familiar and also less meaningful than words. However, both words and nonwords vary on these dimensions; in fact the distributions may overlap (e.g., the nonword CHUMMINGLY is probably more familiar and meaningful than the low-frequency word TARADIDDLE). Frequency effects in the lexical decision task may be exaggerated because low-frequency words are more similar to the nonwords on the FM dimension than are high-frequency words. Hence, when there is insufficient information to make a fast "word" response the subject is required to engage in an extra checking process (possibly checking the spelling of the word). This time-consuming extra checking process is more likely to occur for low-frequency words than for high-frequency words, thereby exaggerating any obtained influence of word frequency. Hence, one should expect a larger influence of word-frequency in the lexical decision task than in the naming task, and in general this is what is found

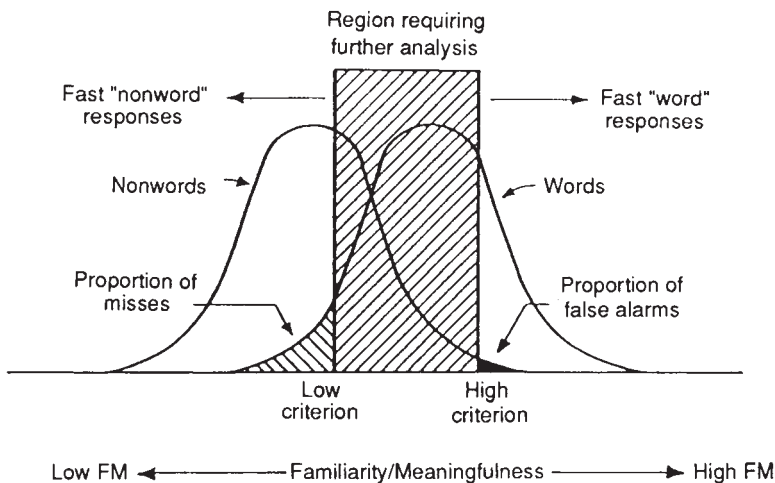


Figure 8. Balota and Chumbley's (1984) two-stage model of the lexical decision task.

(see Balota et al., 2004). Balota and Spieler (1999) have implemented a hybrid model of the lexical decision task that not only accommodates word-frequency effects and other effects, but also accounts for the reaction time distributional aspects of performance, i.e., the shape of the reaction time distribution. It is also important to note that Ratcliff, Gomez, and McKoon (2004) have also argued that decision processes tied to the lexical decision task are critical in understanding word frequency effects, along with other variables, and have nicely modeled such effects with a single-process diffusion model.

There has been considerable controversy in the literature regarding the locus of word-frequency effects in the tasks used to build word recognition models (e.g., see Andrews & Heathcote, 2001; Balota & Chumbley, 1990; Monsell, Doyle, & Haggard, 1989). Of course, the primary intent of the task analysis work is to caution researchers that not all word-frequency effects can be unequivocally attributed to access processes in the tasks that are used to measure word recognition. Although a full discussion of this work is beyond the scope of the present review, it is sufficient to note here that there is little disagreement that word-frequency influences processes involved in word recognition, and hence will need to be incorporated into all models of word recognition. However, as exemplified throughout this review, understanding the operations in the tasks used to build models of word recognition is a paramount first step in building adequate models.

### 4.3. Familiarity

A variable that is highly correlated with frequency is word familiarity. Familiarity is typically based on untimed ratings. For example, subjects may be asked to rate each word on a 7 point scale ranging from extremely unfamiliar to extremely familiar. The importance of familiarity norms was motivated by Gernsbacher (1984) who persuasively argued that the available printed word frequency norms by Kučera and Francis (1967) and Thorndike and Lorge (1944) may not be the most sensitive estimates of the impact of frequency of occurrence on lexical representations. For example, frequency norms typically do not take into account spoken word frequency, and are based on dated and relatively limited samples of word use. Gernsbacher (1984) pointed out that *boxer*, *icing*, and *joker* have the same objective frequency value (according to Kučera & Francis, 1967) as *loire*, *gnome*, and *assay*. Recently, there are a number of more extensive norms that have been developed based on a multifold increase in the sample size compared to the original norms (e.g., Baayen, Piepenbrock, & van Rijn, 1993; Burgess & Livesay, 1998; Zeno, Ivens, Millard, & Duvvuri, 1995). As one might expect, when comparing different frequency norms, the more recent norms are better predictors of both naming and lexical decision performance than the still commonly used Kučera and Francis (1967) norms (see Balota et al., 2004; Zevin & Seidenberg, 2002). Hopefully, cognitive science researchers who are investigating or controlling word frequency will begin to use these more recent norms.

Although the norms are becoming better, it is still the case that they are only a proxy for frequency of exposure. Hence, some researchers still argue that subjective familiarity ratings are a better measure of sheer exposure to a word. However, one might ask what sorts of information do subjects use when making an untimed familiarity rating? Standard



instructions for familiarity ratings tend to be vague and may encourage the use of other types of information. For example, more meaningful stimuli tend to be rated more familiar. In fact, Balota, Pilotti, and Cortese (2001) found that the familiarity ratings of Toglia and Battig (1978) were related to meaningfulness, a semantic variable. As an alternative to standard familiarity ratings, Balota et al. (2001) had participants rate monosyllabic words in terms of subjective frequency. Participants estimated how often they read, heard, wrote, said, or encountered each word based on the following scale: 1 = never, 2 = once a year, 3 = once a month, 4 = once a week, 5 = every two days, 6 = once a day, 7 = several times a day. Balota et al. found that these ratings were less influenced by meaningfulness than the Toglia and Battig (1978) familiarity ratings. Hence, subjective frequency ratings may be more appropriate than traditional familiarity ratings because they are less influenced by semantic factors. Indeed, Balota et al. (2004) found that the subjective frequency ratings were highly predictive of both lexical decision and naming performance above and beyond a host of other correlated variables, such as objective word frequency, length, neighborhood size, spelling-to-sound consistency, etc.

#### 4.4. Age of Acquisition

Within the past decade there has been considerable interest in the influence of the age at which words are acquired on various measures of lexical processing (for a recent review, see Juhasz, 2005). There have been a number of reports suggesting that age of acquisition (AoA) produces a unique influence on word recognition performance (e.g., Brown & Watson, 1987; Morrison & Ellis, 1995) above and beyond correlated variables such as word frequency. The intriguing argument here is that early acquired words could play a special role in laying down the initial orthographic, phonological, and/or semantic representations that the rest of the lexicon is built upon. Moreover, early acquired words will also have a much larger cumulative frequency of exposure across the lifetime.

There are at least two important methodological issues regarding AoA effects. The first concerns the extent to which AoA produces a unique effect in word recognition tasks like naming and lexical decision. One of the problems with assessing this issue is that AoA is correlated with many other variables, including length, frequency, and imageability. Moreover, one might expect an AoA effect not because early acquired words have a special influence on the lexicon, but rather because early acquired words have a greater cumulative frequency, even when objective frequency is held constant (for example, see Lewis, Gerhand, & Ellis, 2001). Although most studies have not teased these possibilities apart, Juhasz and Rayner (2003) found a unique effect of AoA in eye fixation data in reading and Bonin, Barry, Méot, and Chalard (2004) have demonstrated significant effects of objective AoA in word naming and lexical decision performance.

The second issue is concerned with whether or not AoA should be considered an outcome variable (Zevin & Seidenberg, 2002, 2004) or a standard independent (or predictor) variable. Zevin and Seidenberg have argued that AoA predicts word recognition performance because the age at which a word is learned is affected by many factors, and

hence, this is related to the correlated variables issue noted above. They focus on frequency trajectory, which reflects the distribution of exposures that one has with words over time. Some words such as *potty* occur fairly frequently during early childhood but not adulthood, whereas other words such as *fax* occur frequently during adulthood, but not childhood. Therefore, frequency trajectory should influence AoA, and indeed the two variables are correlated. In addition, Zevin and Seidenberg (2004) examined the influence of frequency trajectory and cumulative frequency in naming. They found little evidence for frequency trajectory, whereas cumulative frequency produced a unique effect on naming performance (for an alternative interpretation of the Zevin & Seidenberg findings, however, see Juhasz, 2005). Given the potential theoretical importance of AoA, it appears that this variable will continue to be at the center of considerable empirical and theoretical work in the next several years.

#### 4.5. Orthographic Neighborhood Effects

Although estimates vary, the average adult reader is likely to have about 50,000 words in their lexicon. Because these words are based on a limited number of 26 letters, there must be considerable overlap in spelling patterns across different words. One of the major tasks of an acceptable model of word recognition is to describe how the system selects the correct lexical representation among neighborhoods of highly related orthographic representations. Of course, it is possible that the number of similar spelling patterns may not influence lexical processing and that only a single representation must pass threshold for recognition to occur. However, as already mentioned, it appears that words are not recognized in isolation from other orthographically related representations.

Coltheart, Davelaar, Jonasson, and Besner (1977) introduced the orthographic neighborhood or N metric. N refers to the number of words that could be generated by changing only a single letter in each of the positions within a word. For example, the orthographic neighbors of the word FALL include MALL, FELL, FAIL, BALL, FULL, CALL, among others. There are two major ways that researchers have investigated the influence of N. First, consider the influence of the sheer number of orthographic neighbors. In naming performance, the results are rather straightforward: as the number of orthographic neighbors increases, response latency decreases, and this effect is larger for low-frequency words than high-frequency words (see Andrews, 1989, 1992; Balota et al., 2004). In contrast, in lexical decision performance, increases in N increase response latencies to nonwords, and for word targets the results range from facilitatory Andrews (1989, 1992; Forster & Shen, 1996) to no effect (Coltheart et al., 1977) to some conditions producing inhibitory effects (see, for example, Johnson & Pugh, 1994). In an excellent review of this literature, Andrews (1997) has argued that the variance across the studies of orthographic neighborhood size in lexical decision appears to be in part due to variability in list contexts (e.g., nonword type). It should also be noted that there is evidence of facilitatory effects of large Ns in semantic classification studies (Forster & Shen, 1996; Sears, Lupker, & Hino, 1999a). Finally, it should be noted that the evidence from eye-fixation patterns while people are reading indicate that there is an inhibitory effect of words with large Ns. Importantly, Pollatsek, Perea, and Binder (1999) have shown that

with the same set of words that produces facilitatory effects in lexical decision performance, these words produce inhibitory effects in eye-fixation durations. Clearly, the effects of orthographic N are highly dependent upon the task constraints, and most likely a host of other variables such as individual processing speed (see, e.g., Balota et al., 2004).

A second way to investigate the influence of orthographic neighborhoods is to consider the frequency of the neighbors, i.e., does the stimulus have higher-frequency neighbors or lower-frequency neighbors? In lexical decision performance, there is evidence that targets with higher-frequency neighbors indeed produce inhibition in lexical decision performance, compared to words with lower-frequency neighbors (e.g., Grainger, 1990, 1992; Grainger & Jacobs, 1996; Carreiras, Perea, & Grainger, 1997, but see Pollatsek et al., 1999). However, there is even some conflict here, because in a series of experiments involving both naming and lexical decision performance, Sears, Hino, & Lupker (1995) found facilitation for low-frequency targets with large neighborhoods and higher-frequency neighbors.

Given that word recognition unfolds across time, it is not surprising that both frequency of the neighbors and the size of the neighborhoods should play a role in word recognition tasks. In this light, it is useful to mention the Luce and Pisoni (1989) neighborhood activation model, which they applied to auditory word recognition performance. This model takes into consideration target frequency, neighbor frequency, and neighborhood size via R. D. Luce's (1959) choice rule. Specifically, the probability of identifying a stimulus word is equal to the probability of the stimulus word divided by the probability of the word plus the combined probabilities of the neighbors. Of course, it is possible that the neighborhoods of the neighbors may play a role along with the degree of overlap of the neighbors. At this level, it is noteworthy that recent simulations by Sears, Hino, & Lupker (1999b) have shown that both the Plaut et al. (1996) and the Seidenberg and McClelland (1989) models appear to predict facilitatory effects of neighborhood size that are greater for low-frequency words than for high-frequency words, which is overall most consistent with the data in this area.

Facilitatory neighborhood effects for low-frequency words would appear to be difficult to accommodate within models that have a competitive interactive activation component (e.g., the DRC model of Coltheart et al., 2001, or the multiple read-out (MROM) model of Grainger & Jacobs, 1996). Specifically, the larger the neighborhood, the more competition one should find. Moreover, the facilitatory effects of N produce particular difficulties for serial search models, such as Forster's classic bin model. Specifically, the more items that need to be searched, the slower response latency should be. This is opposite to the most common pattern reported in this literature.

An interesting variation on the influence of orthographic N is the transposed letter effect. Specifically, Chambers (1979) and Andrews (1996) found that words like SLAT produce slower response latencies in lexical decision performance, because these items have a highly similar competitor SALT. Andrews (1996) also found this pattern in naming performance. Note that SLAT is not an orthographic neighbor of SALT, but is very similar because two letters in adjacent positions are switched. As Perea and Lupker (2003) have

recently argued, the influence of transposed letter stimuli is inconsistent with most available models of word recognition, because these models typically code letters by positions within the words. These results are more consistent with recent input coding models such as SOLAR (Davis, 1999), and SERIOL (Whitney, 2001) that use spatial coding schemes for input of letters, that are not simply position specific (also see Davis & Bowers, 2004). Clearly, this is an important new area of research that extends the original work on orthographic N effects and has important ramifications of how the visual system codes the spatial position of the letters within words.

#### **4.6. Phonological Neighborhood Effects**

Although the influence of orthographic neighbors has dominated work in visual word recognition, it is quite possible that phonological neighbors may also play a role. Indeed, work by Yates, Locker, and Simpson (2004) has recently shown that lexical decision performance is facilitated by words with large phonological neighborhoods (also see Yates, 2005). Here, a phonological neighbor reflects a change in one phoneme, e.g., GATE has the neighbors HATE and GET, and BAIT. Yates et al. have also noted that previous studies of orthographic neighborhood size have typically confounded phonological neighborhood size. Although this is a relatively new area of exploration, it indeed is quite intriguing regarding the role of phonology in early access processes (see earlier discussion of feedback consistency effects), and has potentially important implications for how phonology is coded in the extant models (also see Ziegler & Perry, 1998).

### **5. SEMANTIC VARIABLES FOR ISOLATED WORDS**

There have been a number of reports in the literature that indicate that semantic variables associated with lexical representations can modulate the ease of word recognition (see review by Balota, Ferraro, & Connor, 1991, of the early work in this area). This is an intriguing possibility because many models of word recognition would appear to indicate that the word must be recognized before the meaning of the word is determined. For example, within a logogen model, the lexical representation will need to reach threshold before the meaning of the word becomes available. How could it be otherwise? How could the system have access to the meaning without knowing what the stimulus is? Of course, this has some similarity to the word superiority effect described earlier wherein it was argued that the word level information is activated before the letters that make up the word have been recognized, via cascaded top-down activation. In fact, recent computational models by Coltheart et al. (2001) and Plaut et al. (1996) would appear to be able to handle such cascaded influences of meaning en route to making a speeded naming response.

Although there has been a considerable amount of work attempting to specify which semantic variables play a role in word recognition, much of this work has been open to alternative interpretations. Here, we shall briefly review this work emphasizing the primary findings with respect to each of the major variables.

### 5.1. Concreteness/Imageability Effects

Because concreteness is highly correlated with imageability, we will lump these variables together here. Concreteness refers to whether a word can be the object of a sense verb (e.g., touch, see, hear, etc.), whereas imageability typically involves subjects rating words on a low to high imageability scale. One might expect that high-imageable words (e.g., CARROT) may be better recognized than low-imageable words (e.g., FAITHFUL), because of the influence of the more salient referent being activated. Although the early evidence suggested that there were indeed effects of the concreteness/imageability variables (e.g., Boles, 1983; Day, 1977; Paivio & O'Neill, 1970; Rubenstein et al., 1970; Winnick & Kressel, 1965), some of this work was questioned because of the potential for confounding variables (see, for example, Schwanenflugel, Harnishfeger, & Stowe, 1988). However, there are indeed studies that are less susceptible to such criticism and have confirmed that there are concreteness/imageability effects in lexical decision, which are larger for low-frequency words than high-frequency words (e.g., de Groot, 1989; James, 1975; Kroll & Merves, 1986). Of course, this finding in and of itself is not terribly compelling evidence for an influence of meaning en route to word recognition performance, because one could argue that subjects place a premium on semantics in discriminating words from nonwords in the lexical decision task. Hence, the results from the naming task are indeed more noteworthy. Although the effects are clearly smaller, there is also evidence of an effect of concreteness/imageability in naming (e.g., Bleasdale, 1987). In the Balota et al. (2004) study of over 2000 monosyllabic words, they found that there was a reliable unique effect of imageability in naming (based on norms developed by Cortese & Fugett, 2004) after other variables were controlled, but this effect was quite small compared to lexical decision.

Imageability has played a special role in recent work exploring naming performance. Specifically, Strain et al. (1995) found an intriguing interaction between word frequency, spelling-to-sound consistency, and imageability. They found that low-frequency words with inconsistent spelling to sound mappings produced the largest imageability effects. This was viewed as reflecting greater input from preexisting semantic representations for items with relatively low spelling to sound mapping, i.e., low-frequency inconsistent words, which they viewed as consistent with the tripartite connectionist framework, as exemplified by the Plaut et al. (1996) model. It should also be noted, however, that there is some controversy regarding potential correlated variables that may have contributed to this pattern (see Monaghan & Ellis, 2002; Strain, Patterson, & Seidenberg, 2002).

### 5.2. Meaningfulness

A second semantic variable that could play a role in word recognition is the meaningfulness of the stimulus. One way of measuring meaningfulness is simply to count the number of dictionary meanings for each word (for further discussion of different metrics of meaningfulness, see Millis & Button, 1989). Again, the early work in this area was controversial. For example, Jastrzembski (1981) found initial evidence for a facilitatory effect of number of dictionary meanings, while, Gernsbacher (1984) argued that this was

likely due to familiarity being confounded with meaningfulness. Azuma and Van Orden (1997) found an effect of number of meanings in lexical decision performance, but this seemed to depend on the relatedness of the meanings for a word. In fact, Azuma and Van Orden argued that the relatedness of the meanings is more important than the sheer number of meanings. As described below, this may be related to more recent notions of semantic connectivity. Balota et al. (2004) found a small and unique effect of subject-rated meaningfulness that was larger in lexical decision than in naming performance. Finally, it is noteworthy that Rodd (2004) has recently provided evidence that the effect of number of meanings in speeded naming is larger for inconsistent spelling to sound mappings. This, of course, is consistent with the theoretically important observation of an increased influence of a semantic variable (imageability) for low-frequency inconsistent items, reported by Strain et al. (1995) described above.

As noted, meaningfulness is typically defined by the number of dictionary meanings, which can vary in subtle but related ways. For example, the word DOG can mean the four-legged animal, but it can serve as an adjective such as in “My car is a real dog,” wherein the meaning of the word DOG is extended to another form. These might be considered different shades of the same meaning as opposed to distinct meanings of the word. In this light, there has also been some intriguing work investigating word recognition performance on homographs (e.g., the word ORGAN has two very different meanings referring to musical meaning and bodily system). It appears that such items can produce a facilitatory effect in both naming and lexical decision performance (see Hino & Lupker, 1996; Hino, Lupker, & Pexman, 2002). Interestingly, although one finds facilitation in naming and lexical decision, Hino et al. (2002) found inhibition in semantic categorization. These authors argued that only when attention is directed to retrieve semantic information, as in the semantic categorization task, will one find interference effects. A similar pattern was observed by Balota and Paul (1996) in a semantic relatedness judgment task. Finally, in neutral contexts, on-line measures of reading performance, as reflected by eye-fixation durations, suggest that there is interference when ambiguous words have relatively equally dominant interpretations (e.g., CLUB means to hit and organization, with similar frequencies), as if the meanings are competing for interpretation (for a review, see Duffy, Morris, & Rayner, 1988; Morris, this volume). Again, we find that task constraints strongly modulate the influence of a variable.

### 5.3. Grounding Semantics in Large-Scale Databases

There have been a number of recent attempts to ground semantics via analyses of large databases of natural language. This approach avoids some of the pitfalls in trying to quantify meaning as feature lists (e.g., the word DOG may include the features furry, barks, four-legged, pet) or some abstracted prototype (e.g., the modal DOG that is based on your experience with all DOGs). These more recent approaches include Burgess and Livesay’s (1998) hyperspace analogue of language (HAL) and Landauer and Dumais’ (1997) latent semantic analysis (LSA). HAL and LSA capture the meaning of words from the context in which a given word appears. Hence, the meaning of DOG is an evolving concept dependent upon an individual’s experience with DOG in various linguistic contexts.

Buchanan, Westbury, and Burgess (2001) have shown that estimates from HAL indeed predict lexical decision performance (for a detailed discussion of this work, see Burgess, this volume). It is noteworthy that an early study by Schwanenflugel, Harnishfeger, and Stowe (1988) provided evidence that a variable referred to as contextual availability can have an influence on isolated word recognition in lexical decision performance above and beyond influences of correlated variables such as concreteness, familiarity, length, etc. Contextual availability refers to how easily a subject is able to think of contexts in which a given word might occur.

An intriguing alternative approach has recently been developed by Steyvers and Tenenbaum (2005). They have utilized recently developed graph theoretic techniques to look at metrics of connectivity (along with other metrics) of meanings of words in a set of large-scale databases including Roget's (1911) Thesaurus, Miller's (1990) WordNet, and Nelson, McEvoy, and Schreiber's (1998) word association norms. Based on analyses of these databases, Steyvers and Tenenbaum have shown that semantic memory has a small-scale network structure in which a relatively small number of concepts serve as communication hubs for the rest of the semantic network. If semantic networks are represented in terms of the structure hypothesized by Steyvers and Tenenbaum, then words characterized by a high degree of connectivity with other words may be processed more quickly than words characterized by sparse connections. Indeed, Steyvers and Tenenbaum found evidence for such an effect in naming and lexical decision performance, above and beyond more standard lexical variables (also see Balota et al., 2004).

#### **5.4. Additional Semantic Variables that Produce Effects in Isolated Word Recognition Paradigms**

Because of space limitations, we shall only briefly mention a few other findings that would appear to indicate that meaning can have an early influence in word recognition performance. First, there is evidence that concreteness of a word can influence the time taken to generate an associate from that word (e.g., de Groot, 1989). Because subjects must recognize a word en route to generating an associate, this effect might be due to word recognition processes. Second, and along these same lines, Chumbley and Balota (1984) have found that the time taken to generate associates from one group of subjects can be used as a predictor of lexical decision performance for the same set of words when presented in isolation to a second group of subjects, above and beyond other related variables such as frequency, length, etc. Third, Whittlesea and Cantwell (1987) found that providing meaning for a nonword can produce a word-superiority effect, and also a study by Forster (1985) indicated that providing meaning for a nonword can produce a masked form priming effect in the lexical decision task. Both the word-superiority effect and the masked form priming effect would appear to tap relatively early lexical processes. Finally, there is evidence from masked semantic priming studies (reviewed below) suggesting that highly masked primes (that subjects apparently cannot consciously recognize) produce semantic priming effects, i.e., facilitate the processing of related targets compared to unrelated targets (see Holender, 1986, and the accompanying commentary for a discussion of the degree of conscious processing of the primes in these studies).

At the very least, such threshold priming effects suggest that under presentation conditions that minimize conscious processing of the prime, meaning access can still occur.

## 5.5. Summary

The possibility that meaning-level representations play a role in isolated word recognition has relatively far reaching implications for current models of word recognition. Most of the available models emphasize the stages that subjects use in accessing the mental lexicon, with relatively little direct influence of meaning-level variables. However, when reminded that the role orthographic patterns play in reading is to convey meaning and not simply to convey lexicality then one might easily envisage an architecture that incorporates a relatively early influence of meaning. At this level, it should be no surprise that meaning-level representations may contribute to relatively early perceptual analyses and aid in constraining the percept, i.e., recognition of the word. Although recent connectionist and dual-route models of word processing acknowledge such effects, the devil is in the details of implementing such meaning-level influences.

## 6. CONTEXT/PRIMING EFFECTS

Heretofore, we have primarily discussed the literature that deals with variables that influence isolated visual word recognition. Of course, readers typically encounter words in the context of other words. We now turn to a summary of the influences of contexts (hereafter referred to as primes) on word recognition processes. In these studies, two letter strings are typically presented and the researcher manipulates the relation between the two strings. For example, the strings may be orthographically related (COUCH-TOUCH), phonologically related (MUCH-TOUCH), semantically related (FEEL-TOUCH), or unrelated (NAIL-TOUCH). By manipulating the types of relationships between the primes and targets one can obtain evidence regarding the architecture of the word recognition system. For a more detailed discussion of this rich literature, see Neely (1991), Hutchison (2004), McNamara (2005) for important reviews of the semantic priming literature and Kinoshita and Lupker (2003b) for a volume dedicated to masked priming effects.

### 6.1. Orthographic Priming Effects

An interesting approach to identifying the access code in word recognition is the masked orthographic priming paradigm developed by Evett and Humphreys (1981, also see Humphreys, Besner, & Quinlan, 1988; Humphreys, Evett, Quinlan, & Besner, 1987). In this paradigm, subjects are briefly presented two letter strings that are both preceded and followed by pattern masks. The two letter strings vary in terms of orthographic, phonological, or semantic relatedness. Here, we focus on the orthographic priming conditions. There are a number of interesting findings in these masked priming studies: first, on most trials, subjects are unable to consciously identify the prime items and hence any influence of the prime items presumably reflects early access processes. Second, subjects are better at identifying the second letter string when it shares letters with the first letter



string even though these shared letters are presented in different case. For example, relative to a baseline (e.g., harmless-ATTITUDE), there are priming effects for both *identity priming* (e.g., attitude-ATTITUDE) and *form priming* (e.g., aptitude-ATTITUDE). Third, in lexical decision, evidence for nonword repetition priming (e.g., flirp-FLIRP) is clearly less powerful than word repetition priming (Forster, 1998). Although earlier studies actually failed to find nonword repetition priming effects in lexical decision (see, for example, Forster and Davis, 1984), more recent studies have observed reliable effects (Bodner & Masson, 1997; Sereno, 1991). Fourth, in masked repetition priming studies, the effects of target word frequency and prime-target repetition are additive (Forster & Davis, 1984), a finding which is more consistent with search-class than with activation-class models of lexical access. Fifth, eye-tracking studies by Rayner, McConkie, and Zola (1980) using orthographic priming techniques have provided compelling evidence for a case independent orthographic code being used to access words in the parafovea while reading (for reviews, see Balota & Rayner, 1991; Rayner, 1998).

A particularly intriguing aspect of the masked priming literature is that within a range of short-duration primes, there is a relatively linear relationship between the *duration* of the masked prime and the *magnitude* of the priming effect (Forster & Davis, 1984). Specifically, a prime with a duration of 30ms produces a priming effect of about 30ms, whereas a prime with a duration of 20ms produces a priming effect of about 20ms. Forster (1998) has argued that this is most consistent with an *Entry Opening* process where the prime has the influence of opening the target's lexical representation, allowing the target to be processed more rapidly. This Entry Opening account of masked priming nicely accommodates the equivalent masked repetition effects for high-frequency and low-frequency words, i.e., the masked prime has the effect of opening the lexical representation (Forster & Davis, 1984). However, it is unclear how the Entry Opening model accounts for nonword repetition priming effects, since nonwords, by definition, have no pre-existing lexical representations. To address such nonword effects, Bodner and Masson (1997) have proposed that masked priming effects are driven by a nonlexical locus, specifically, the retrieval of episodic memory traces established during previous encounters with the stimulus (for an episodic trace view of lexical processing, see Goldinger, 1998). This account implies that masked nonword primes operate nonlexically to facilitate orthographic processing (for an alternative explanation, see Forster, 1998).

Finally, task-specific effects have also been observed in masked priming. For example, there is evidence for a phenomenon called the *masked onset priming effect*. This effect was first reported by Forster and Davis (1991), who found that naming latencies to a target were facilitated when the prime and target shared the initial letter (e.g., save-SINK) compared to when they did not (e.g., farm-SINK). Further work by Kinoshita (2000) has revealed that this effect is position-dependent and is observed *only* when the initial onset (not the letter) is shared. For pairs bingo-BLISS, which has a common initial letter but different onsets (i.e., /B/ vs. /BL/), the effect was eliminated. Kinoshita argued that this supported a serial left-to-right procedure in naming performance, and may reflect articulatory planning rather than orthography-to-phonology computations (see also Schiller, 2004). The onset effect is only observed with tasks that require articulation, such as

speeded naming, and not with lexical decision (Forster & Davis, 1991). Positing an articulatory nonlexical priming component for speeded naming may also explain why nonword repetition priming effects, which are equivocal in lexical decision, are more consistent in speeded naming (Masson & Isaak, 1999).

## 6.2. Phonological Priming Studies

There has been considerable debate concerning the role of phonological codes in word recognition (for an excellent review of this literature, see Frost, 1998). The extremes range from all words must be recognized via a phonological (assembled) code to the notion that many words (e.g., high-frequency words for skilled readers) are only accessed via an orthographic (addressed) code. Although there is controversy regarding the role of a phonological code in visual word recognition, there is considerably less debate regarding the importance of phonological codes in reading text, wherein, phonological codes produce representations that appear better suited for aspects of comprehension that place considerable demands on the working memory system (e.g., Baddeley, Eldridge, & Lewis, 1981; Besner, 1987; Slowiaczek & Clifton, 1980). It is possible that such phonological codes become active after lexical access has taken place in such reading studies. The more narrow issue here is whether phonological codes are necessary en route to word recognition. With this in mind, we now turn to the phonological priming literature.

Evett and Humphreys (1981) used the masked priming paradigm, described above, also to investigate the viability of a phonological access code, under conditions wherein conscious processing was limited. The results of this study indicated that there was priming for pairs that were orthographically and phonologically related (e.g., bribe-TRIBE) compared to pairs that were orthographically related but phonologically unrelated (break-FREAK). Moreover, the effect occurred across case changes. In addition, in a similar masked priming paradigm, Humphreys, Evett, and Taylor (1982) found that identification accuracy was higher for targets (e.g., SHOOT) that followed homophonic primes (e.g., chute) compared to targets that followed graphemically related (e.g., short) or unrelated primes (trail). However, there was no facilitation from a nonword phonologically related prime (e.g., smorl-SMALL), suggesting a lexical locus for the priming effect.

Evidence for phonological mediation has also been obtained with an associative priming paradigm, which permits conscious, albeit brief, processing of primes. For example, Lukatela and Turvey (1994) compared priming effects across four conditions at different stimulus onset asynchronies (SOAs): *standard semantic* priming (e.g., TOAD-FROG), *word homophonic* priming (e.g., TOWED-FROG), *nonword homophonic* priming (e.g., TODE-FROG), and an orthographic control condition (e.g., TOLD-FROG). At short (i.e., 50ms) SOAs, the three related conditions produced comparable facilitation priming effects, compared to the control condition. However, at longer SOAs (i.e., 250ms), TODE became a stronger prime than TOWED. These findings reinforce the role of phonology in early visual lexical access, and also suggest that although word homophone primes (i.e., TOWED) are initially effective, they are quickly suppressed when the system detects the mismatch between their orthography and the addressed spelling of TOAD.

It is important to point out that the validity of the findings described above rests on the assumption that the orthographic control (e.g., TOLD) is as orthographically similar to the critical associate (e.g., TOAD) as the homophone (e.g., TOWED) (Pollatsek, Perea, & Carreiras, 2005). Some have failed to replicate the homophone/pseudohomophone advantage described above (see, for example, Davis, Castles, & Iakovidis, 1998) and Pollatsek et al. argued that this inconsistency may be due to imperfect matching of controls to homophones. After controlling for this potential confound, Pollatsek et al. still observed early phonological effects in a Spanish lexical decision task, strengthening the assertion that phonological coding of the primes takes place relatively early in the word recognition process.

Interestingly, the importance of phonological codes in word identification has been demonstrated in both orthographically shallow languages, where there is a direct mapping between orthography and pronunciation (e.g., Serbo-Croatian, for a review, see Carello, Turvey, & Lukatela, 1992) and orthographically deep languages, where the mapping appears to be more arbitrary (e.g., Chinese, for a review, see Tan & Perfetti, 1998). Clearly, phonological information can constrain visual word recognition even in logographic scripts where one would expect meaning to be derived directly from ideograms (Hoosain, 1991). For example, Tan and Perfetti (1999) sequentially presented pairs of Chinese words in a meaning-judgment task, in which subjects were asked to judge whether the two words had the same meaning or not. On trials where participants were supposed to make a “no” judgment (i.e., the two words had different meanings), the “no” response had longer latencies when the foil was homophonous with the base word compared to when it was not.

There have been additional tasks used to investigate the early influence of phonological processes. For example, Van Orden (1987; Van Orden, Johnston, & Hale, 1988) used a semantic categorization task, in which subjects had to decide whether a given word was a member of a semantic category. The intriguing finding here is that subjects produced considerably higher error rates for words that were homophones of an exemplar (e.g., MEET for the category FOOD), compared to an orthographically related control (e.g., MELT). This finding suggests a clear role of phonological information in accessing the semantics necessary for category verifications, and nicely converges with the results from the Tan and Perfetti (1999) study with Chinese characters. Jared and Seidenberg (1991) replicated this pattern showing that this effect is more likely to occur for low-frequency words. This pattern also appears to be consistent with the earlier observation of an interaction between frequency and spelling-to-sound regularity that was observed in word pronunciation performance (also, see Rodd, 2004). In another paradigm, Ziegler, Ferrand, Jacobs, Rey, and Grainger (2000) used an *incremental priming technique*, by manipulating the duration of the prime, which provides a window into the time-course of masked priming effects. They found clear orthographic and phonological priming effects in both naming and lexical decision performance, with the naming task being more dependent upon phonological priming. This study is particularly noteworthy because it provides a method to help understand the temporal locus of such priming effects. Finally, it is also worth noting that just as in the case of orthographic priming, there is also evidence of phonological priming in the parafoveal priming paradigm in more natural reading contexts. Specifically, Pollatsek, Lesch, Morris, and Rayner (1992) found that previews that were homophones of targets

(e.g., site-cite) facilitated performance (both in pronunciation latencies and fixation durations during reading), compared to nonhomophonic previews that were controlled for orthographic similarity (e.g., cake-sake). Lee, Binder, Kim, Pollatsek, and Rayner (1999) have extended this work with a *fast-priming* paradigm (for a description, see Sereno & Rayner, 1992), a task which taps early stages of word processing. They observed an interesting prime by word frequency interaction; specifically, homophonic priming was primarily obtained with high-frequency word primes. Taken together, these findings not only support the role of phonology as an access code, but also suggest that lexical information may be guiding phonological coding early in fixations during reading (Lee et al., 1999).

### 6.3. “Semantic” Priming Effects

The semantic (associative) priming paradigm is clearly the most studied area of priming. (Because of space limitations, the present section will be limited to single word priming studies, see Morris, this volume, for a review of sentential semantic priming effects.) This enterprise began with a seminal study by Meyer and Schvaneveldt (1971). They found that subjects were faster to make lexical decisions to word pairs when the words were related (e.g., CAT-DOG) compared to when the words were unrelated (e.g., CAT-PEN). The prevailing *zeitgeist* was ready to welcome such a finding for a number of reasons: first, the dependent measure was response latency and response latency measures were becoming the mainstay of cognitive experiments. Second, the study nicely demonstrated top-down contextual influences (e.g., semantic relations) on what appeared to be a bottom up, stimulus driven word recognition processes. This was a major emphasis in Neisser’s (1967) *Cognitive Psychology* that was published a few years earlier. Third, the effect was quite robust and easily replicated. Fourth, the semantic priming task appeared to be ideally suited to map out the architecture of meaning-level representations and the retrieval operations that act upon such representations; both of these issues would at least appear to be critical to higher-level linguistic performance.

#### 6.3.1. *Semantic or associative effects?*

There is little controversy that across the major tasks used to build word recognition models (threshold identification, lexical decision, pronunciation, and on-line measures of eye-movements during reading), words are better recognized when embedded in semantically related contexts compared to unrelated contexts. However, there are many questions that have arisen regarding this effect. For example, one might ask if the effect is truly “semantic” (i.e., reflects similarity in semantic features, Smith, Shoben, & Rips, 1974 or category membership, Collins & Quillian, 1969), or if it primarily reflects associative relationships among items. For example, DOG and CAT share a semantic and associative co-occurrence relationship, whereas RAT and CHEESE appear to primarily share an associative relationship. Two recent reviews of this topic appear to come to somewhat different conclusions. Lucas (2000) argued that there was indeed evidence that semantic priming effects truly reflected “semantic” information, whereas, Hutchison (2003) concluded that, with a few exceptions, a simple associative account could handle most of this literature. Of course, teasing apart semantic influences from associative influences has been rather difficult because these relationships typically co-occur. In an attempt to

address this issue, researchers have attempted to identify items that are of the same category (e.g., *glove-hat*) but do not entail a strong associative relation, e.g., are not produced in associative production norm studies in which subjects are asked to generate associates to a given word (see, for example, Palermo & Jenkins, 1964). The results from three such studies (e.g., Lupker, 1984; Schreuder, Flores d'Arcais, & Glazenberg, 1984; Seidenberg, Waters, Sanders, & Langer, 1984b) indicate that there is still some priming with such stimuli in both lexical decision and in pronunciation, although the pure semantic effects are somewhat smaller in pronunciation.

One must be cautious in accepting the conclusion that there are pure nonassociative semantic priming effects. This caution is warranted for the following reasons: first, and foremost, it is unclear whether the relatively small, but "pure," semantic priming effects might be due to some lingering associative-level relationship for words that researchers believe only have a semantic relationship (e.g., GLOVE-HAT are probably more likely to co-occur compared to the pair GLOVE-PEN). Second, as noted below, there is evidence that priming can occur across mediated pairs within the memory network. Thus, it is at least possible that some of the priming from GLOVE to HAT is due to GLOVE priming CLOTHES and CLOTHES priming HAT. Third, when one considers low-category dominance pairs, words that are categorically related but may have little associative relationship, one finds that there is relatively little priming in pronunciation performance (Keefe & Neely, 1990; Lorch, Balota, & Stamm, 1986); however, in lexical decision performance, there appears to be equivalent priming for high- and low-category dominance pairs (e.g., Lorch et al., 1986; Neely, Keefe, & Ross, 1989). The difference between pronunciation and lexical decision performance is particularly noteworthy here. As noted below, a number of researchers have suggested that at least part of the priming effect observed in the lexical decision task may be due to a type of post-lexical checking process. Subjects can use the relatedness between the prime and target to bias their "word" response because nonwords by definition are never semantically related to the primes. In fact, Neely et al. (1989) have found that the priming effect for low-dominance exemplars in the lexical decision task depends upon the ratio of nonwords to words. Neely et al. argued that the nonword/word ratio should modulate the likelihood of the checking process being engaged in the lexical decision task. Hence, because of the task-specific list context effect in this study (i.e., the effect of the nonword/word ratio), one may question the argument for a pure semantic priming effect in access processes (also see Balota & Paul, 1996). In the following discussion, we will use the term "semantic" priming effects, however, the reader by now should understand that many of these effects could be primarily "associative" in nature.

### 6.3.2. *Mediated priming effects*

At an intuitive level, the finding that subjects are better at recognizing words that are embedded in related contexts compared to unrelated contexts is no great surprise. (Although, as described below, it is not so intuitive what mechanisms are responsible for such effects.) However, the priming literature has also provided some very counterintuitive findings. Consider the two words LION and STRIPES. These two

words do not have any obvious direct relation, but do have an indirect relation through the word TIGER. Such items have been referred to as mediated pairs and the research addressing mediated priming effects has provided some interesting results. First, in a standard lexical decision task in which subjects only respond to the target string, there is little evidence for mediated priming (cf. Balota & Lorch, 1986; de Groot, 1983; den Heyer, Sullivan, & McPherson, 1987). However, if one changes the lexical decision task so that subjects either (a) make lexical decisions about the prime and target (McNamara & Altarriba, 1988) or (b) only make a response to word targets and not respond to non-word targets (den Heyer, Sullivan, & McPherson, 1987), mediated priming does occur in the lexical decision task. Moreover, when one now turns to the pronunciation task, one does find mediated priming effects (Balota & Lorch, 1986). Researchers have again argued that checking processes tied to the lexical decision task can strongly control when mediated priming effects will be found in this task (e.g., Balota & Lorch, 1986; McNamara & Altarriba, 1988; Neely, 1991). The notion is that checking for a relationship between the prime and target will not yield a successful outcome for mediated prime–target pairs, because such pairs do not share any obvious relationship. Thus, a negative outcome from the checking process may override the mediated influence from the prime to the target.

### 6.3.3. *Threshold priming effects*

A second important finding in this literature deals with threshold semantic priming effects, mentioned earlier. In the initial studies in this area, researchers first determined each subject's threshold wherein he or she can no longer discriminate between the presence or absence of a stimulus. These thresholds are then used in a later semantic priming task, in which the prime is presented at a subject's threshold and the target is presented in a lexical decision task. The intriguing finding here is that there still is evidence for semantic priming effects, under conditions in which subjects apparently can no longer make presence/absence decisions about the prime item (Balota, 1983; Carr & Dagenbach, 1990; Dagenbach, Carr, & Wilhelmson, 1989; Fowler, Wolford, Slade, & Tassinari, 1981; Marcel, 1983; Marcel & Patterson, 1978). There have also been similar findings reported in the pronunciation task (Carr, McCauley, Sperber, & Parmelee, 1982; Hines, Czerwinski, Sawyer, & Dwyer, 1986). Although there is some concern regarding whether subjects are truly at an objective presence/absence threshold (see Cheesman & Merikle, 1984; Holender, 1986; Merikle, 1982), it is clear that primes presented under very degraded conditions still produce semantic priming effects. It is noteworthy that the threshold priming literature has also been extended to functional neuroimaging techniques. For example, in an event-related potential/functional magnetic resonance neuroimaging study, Dehaene et al. (1998) used number primes that were so briefly presented that participants were unable to discriminate them from foils. Nevertheless, these primes influenced performance on a semantic comparison task (press one key if the target is less than 5 and another key if the target is greater than 5), and modulated hemodynamic measures of brain activity. As in the mediated priming studies, these studies indicate that conscious access to a prime–target relationship does not appear to be a necessary condition for obtaining semantic priming effects.

In some studies of threshold priming, stimuli and/or targets are repeated across trials, with thresholds being carefully monitored. There has been a recent debate about whether such effects in these paradigms reflect unconscious access to meaning at the whole-word level (Abrams & Greenwald, 2000; Damian, 2001; Naccache & Dehaene, 2001). For example, Abrams and Greenwald (2000) argued that threshold priming effects in these studies may reflect automatized stimulus–response mappings that develop as participants make responses to visible targets across trials (for an alternative view, Damian, 2001, but also see Kunde, Kiesel, & Hoffmann, 2003). Specifically, after participants repeatedly (and consciously) classify *smut* and *bile* as negative words, *smile* (*smut-bile* hybrid) subsequently functions as a negative valence masked prime (Abrams & Greenwald, 2000). No significant priming is found for masked primes that had *not* earlier appeared as a target to be classified. These findings question the traditional premise that threshold primes are analyzed at the whole-word level, and suggest that subconscious processing may instead involve sublexical analyses. There is, however, some recent evidence that these findings are specific to words. With numbers, masked primes are apparently able to provide access to long-term semantic memory (Greenwald, Abrams, Naccache, & Dehaene, 2003). The important point here is that one needs to be cautious in interpreting “threshold” priming effects when stimuli are repeated across trials.

It is also noteworthy that the masked priming paradigms have been extended to the domain of social psychology. The overarching question of interest is whether affective states can be automatically triggered by threshold-level primes. For example, Fazio, Sanbonmatsu, Powell, and Kardes (1986) found evidence for automatic attitude activation using an adjective connotation task (i.e., rate a target as “good” or “bad”). They observed that participants rated a negative valenced target (e.g., DISGUSTING) more quickly when it was preceded by a negative prime (e.g., COCKROACH), compared to a control item. Wittenbrink, Judd, and Park (1977) found similar priming effects under highly masked prime conditions. In a highly cited paper by Devine (1989) using the Neely (1977) automatic and controlled distinction in semantic priming, there was clear evidence of automatic activation of racial prejudice at short SOAs that was ultimately controlled at longer SOAs (see also Payne, 2001; Lambert et al., 2003). In reviewing this literature, Fazio (2001) has argued that such attitude priming is automatic and unconscious (also see De Houwer, Hermans, & Eelen, 1998).

Automatic influences from masked primes have also been detected using more ecologically valid paradigms. Bargh and Chartrand (1999) provide a comprehensive review of this literature. For example, Bargh, Chen, and Burrows (1996) found that participants presented with highly masked primes that presumably activated “rudeness” traits (e.g., *rude*, *impolite*, *obnoxious*) were more likely to interrupt a subsequent conversation than if they were primed with “politeness” traits (e.g., *respect*, *considerate*, *polite*). Collectively, the evidence from attitude and affect priming in social psychology is in-line with the evidence from semantic masked priming in visual word recognition. Given the cascading nature of the models that we discussed earlier, such a pattern might be expected. However, this literature also clearly demonstrates that one needs to be cautious and use converging evidence to evaluate whether such effects are in the purest sense unconscious.

#### 6.3.4. Backward priming effects

The third area that is counterintuitive is backward priming. There are two types of backward priming effects. First, there is evidence (Balota, Boland, & Shields, 1989; Kiger & Glass, 1983) that indicates one can still find semantic priming (DOG-CAT vs. PEN-CAT) even when the prime (DOG or PEN) is presented temporally after the target (CAT). These results suggest that early on in target processing, subsequent related prime information/activation can actually “catch-up” to influence response latencies to the target. Such an effect would appear to most naturally fall from a cascadic framework in which partial activation is released from representations before such representations have reached threshold.

A second type of backward priming effect is backward semantic priming. In backward semantic priming, prime–target pairs are presented that entail directional relations, e.g., BELL is related to BOY in the BELL-BOY direction, but not in the BOY-BELL direction. Koriat (1981) and Seidenberg et al. (1984b) have reported evidence of backward priming in the lexical decision task. However, when one turns to the pronunciation task, there is relatively little evidence of backward priming (Seidenberg et al., 1984b), except under short stimulus onset asynchronies (see Kahan, Neely, & Forsythe, 1999; Peterson & Simpson, 1989). It is possible that at short SOAs, there is sufficient temporal overlap between the target and the context to produce the first type of backward priming, noted above, even in naming.

### 6.4. Syntactic Priming

If associative/semantic context does indeed influence lexical processing, then it is quite possible that syntactically appropriate vs. inappropriate contexts might also influence lexical processing. In fact, effects of syntactic context on word recognition might be quite informative. At one level, one might argue that associative pathways between syntactically appropriate words might be represented within the lexicon, simply due to associative co-occurrence of such pairs (c.f., Ratcliff & McKoon, 1988). Likewise, one might argue that syntactic tags within lexical representations might produce priming to consistent syntactic representations. Alternatively, one might argue that syntactic representations are only engaged after word recognition and hence one might not expect syntactic priming effects in word recognition tasks.

One of the first syntactic priming studies was reported by Goodman, McClelland, and Gibbs (1981). Goodman et al. found that subjects were faster to make lexical decisions to targets (e.g., *oven*) that followed syntactically appropriate primes (e.g., *my*) compared to syntactically inappropriate primes (e.g., *he*). Seidenberg et al. (1984b) replicated this pattern in a lexical decision task, but only obtained marginal effects in the pronunciation task. As in the priming studies mentioned above, Seidenberg et al. argued that the syntactic priming effect in the lexical decision task was probably due to some post-lexical processing of the relation between the prime and target. At first, it appeared that Seidenberg et al.’s arguments are not totally correct, because West and Stanovich (1986) obtained relatively large syntactic



priming effects in both the pronunciation task and the lexical decision task. However, Sereno (1991) argued that the past syntactic priming studies have used relatively long prime–target SOAs, and hence may be due to attentional expectancies. In a series of studies, with highly masked primes, Sereno found clear syntactic priming effects in lexical decision that were eliminated in naming, consistent with the Seidenberg et al.'s original arguments about task-specific post-lexical checking processes producing the syntactic priming effects.

### 6.5. Prime Type by Factor Interactions

Of course, the importance of the semantic priming literature is not simply the demonstration that certain factors produce facilitation in the lexical decision and naming tasks, but its importance extends to the intriguing interactions that have been uncovered. As an example, consider the following intriguing pattern of interactive effects: (a) semantic priming effects are larger for low-frequency words than for high-frequency words (Becker, 1979); (b) semantic priming effects are larger for degraded words compared to non-degraded words (Becker & Killion, 1977; Borowsky & Besner, 1991); (c) there are additive effects of stimulus degradation and word frequency (see Balota & Abrams, 1995; Becker & Killion, 1977; Borowsky & Besner, 1991). Traditionally, this constellation of findings has been used to support independent, sequentially organized stages in lexical processing (Borowsky & Besner, 1993; Plourde & Besner, 1997; Sternberg, 1969). In contrast, Plaut and Booth (2000) have argued that a single-mechanism PDP model, implemented with a sigmoid activation function, can more parsimoniously simulate these effects, along with additional findings in the literature. This debate has recently resurfaced, with Borowsky and Besner (2005) contending that there is insufficient evidence that the PDP model implemented by Plaut and Booth (2000) can simultaneously achieve high lexical decision accuracy and correctly simulate the joint effects of stimulus quality, word frequency, and priming in speeded lexical decision. Instead, they argue that the available evidence is more consistent with serially organized processing stages that are differentially sensitive to degradation, semantic relatedness, and word frequency. Evidence for independent stages of processing is especially intriguing when considering the human word recognition architecture.

### 6.6. Theoretical Accounts of Semantic Priming Effects

The importance of the semantic priming paradigm has not simply been restricted to models of word recognition, but also has extended to more general issues concerning representation and retrieval processes. We shall now briefly discuss some of the theoretical issues that have been nurtured by this literature, but the interested reader should see Neely (1991), Hutchison (2003), and McNamara (2005) for a full discussion of these theoretical mechanisms.

#### 6.6.1. *Automatic spreading activation*

The notion that semantic/lexical memory may be represented by nodes that reflect concepts and that such conceptual nodes are interconnected via associative/semantic

pathways has been central to a number of developments in cognitive psychology (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975; Posner & Snyder, 1975). As Anderson (1983) points out, the spreading activation metaphor has probably been most strongly supported by the semantic priming paradigm. When a node in memory becomes activated via stimulus presentation or via internal direction of attention, the notion is that activation spreads from that node along associative pathways to nearby nodes. Thus, the reason that subjects are faster to recognize DOG when it follows CAT, compared to when it follows PEN is because the underlying representation for these two words are connected via an associative/semantic pathway and when CAT is presented activation spreads from its underlying node to the node underlying DOG. Thus, the representation for DOG needs less stimulus information to surpass threshold.

Although there is a limited capacity version of spreading activation theory (e.g., Anderson & Bower, 1973), by far, most of the work in the priming literature has addressed the automatic nature of the spreading activation mechanism. In one of the clearest expositions of this mechanism, Posner and Snyder (1975) argued that the automatic spreading activation mechanism was (a) fast-acting, (b) independent of subjects' conscious control, and (c) primarily produces facilitation for related targets and little inhibition for unrelated targets, compared to an appropriate neutral baseline condition (see Neely, 1977). Because of controversies regarding the adequacy of a given neutral prime condition (see, for example, Balota & Duchek, 1989; de Groot, Thomassen, & Hudson, 1982; Jonides & Mack, 1984; Neely, 1991), we will primarily focus on Posner and Snyder's first two characteristics.

There are a number of important semantic priming results that would appear to support Posner and Snyder's automatic spreading activation mechanism. First, the evidence for semantic priming under highly masked priming conditions, reviewed above, is consistent with the notion that priming effects are independent of consciously controlled processing (e.g., Balota, 1983; Dehaene et al., 1998; Fowler et al., 1981; Marcel, 1983). Second, the evidence that there are mediated priming effects at relatively short prime--target SOAs (e.g., from LION to STRIPES), when it is unlikely that subjects have sufficient time to generate an attentional expectancy for the mediated target also supports the notion of an automatic spread of activation within a memory network. Finally, the findings that prime-expectancy instructions (Neely, 1977) and relatedness proportion manipulations have relatively little impact at short SOAs (den Heyer, Briand, & Dannenbring, 1983), but strong influences at long SOAs, support the notion that the automatic spreading activation mechanism is relatively fast acting (i.e., occurs at short SOAs), decays quickly, and is independent of subjects' conscious expectations.

Although there appears to be support for something akin to an automatic spreading activation mechanism, there are some caveats. For example, initially, there was little evidence of priming effects occurring across unrelated words (e.g., facilitation from LION to TIGER in LION-CHALK-TIGER compared to FROG-CHALK-TIGER, e.g., Gough, Alford, & Holley-Wilcox, 1981; Masson, 1991; Ratcliff & McKoon, 1988). Clearly, if the effect is automatic, one would expect such effects. In this light, it is noteworthy that more recent studies by Joordens and Besner (1992), McNamara (1992), and

Balota and Paul (1996) have obtained such priming effects. Of course, one might expect such priming effects to be relatively small because the unrelated word may have the effect of shifting attention away from the related prime and this shift may override any pure spreading activation effect. A second potential problem with the automatic nature of spreading activation is that semantic priming effects can be eliminated when subjects process the primes in a very shallow fashion, e.g., responding to whether a given letter is in the prime or an asterisk is beside the prime (e.g., Henik, Friedrich, & Kellogg, 1983; Smith, 1979; Smith, Theodor, & Franklin, 1983). Unless the shallow processing task eliminates processing of the prime at the lexical level, one should expect automatic spreading activation and semantic priming effects under shallow processing conditions (for further discussion of this issue, see Besner, Smith, & MacLeod, 1990). Finally, Balota, Black, and Cheney (1992) have shown that prime-expectancy instructions (e.g., subjects are instructed to expect exemplars from the TREE category when presented the prime METALS) can influence naming performance even at very short prime–target SOAs. Thus, although there is support of an automatic spreading activation mechanism involved in semantic priming tasks, it appears that we still do not fully understand the constraints under which this mechanism operates (for a recent discussion of the automatic nature of spreading activation, see Neely & Kahan, 2001).

#### 6.6.2. *Attentional/expectancy effects*

A second mechanism that presumably underlies semantic priming effects is a more attention-based expectancy factor (Balota, 1983; Becker, 1980; Favreau & Segalowitz, 1983; Neely, 1976, 1977). Here, when the prime is presented subjects generate expectancies about potential candidate targets. When the expectancy is correct, facilitation occurs, however, when the expectancy is incorrect, inhibition occurs. This expectancy-based model of priming falls naturally from the work of Posner and Snyder (1975) and Neely (1977), wherein, instructional manipulations and list context effects have larger influences at long SOAs (when expectancies have had time to be generated) than at short SOAs. Of course, at one level, the impact of an attentional-based expectancy mechanism should not be surprising because it simply reflects the probability of correctly predicting the target word when given the prime. The more intriguing work here is the specification of the parameters that modulate the expectancy effects, i.e., the rate at which expectancies are generated across time, the duration at which the expectancy is maintained, and the characteristics of such an expectancy set (for a detailed discussion of a semantic expectancy model, see Becker, 1980, 1985).

#### 6.6.3. *Backward-checking accounts*

As noted above, a number of researchers have argued that priming effects in the lexical decision task may reflect influences at a post-lexical decision level (e.g., Balota & Lorch, 1986; de Groot, 1984; Forster, 1979, 1981; Neely, 1976, 1977; Neely & Keefe, 1989; Seidenberg et al., 1984b; Stanovich & West, 1983). Subjects can rely on finding a relationship between the prime and target to bias the “word” response in the lexical decision task, because nonwords are never related to the primes. This would have the effect of

facilitating “word” decisions to related prime–target trials and possibly inhibiting “word” decisions to unrelated prime–target trials. As described above, there is considerable support for such a mechanism in the lexical decision task. For example, the finding that there is backward priming in the lexical decision task (e.g., priming from BOY to BELL) suggests that subjects can use the target to check in a backwards direction (BELL to BOY) for a potential relationship to the prime item. Although the backward checking mechanism would appear to be primarily a nuisance variable tied to the lexical decision task, one might argue that this checking process may reflect a tendency in natural language processing to integrate the meaning of the current word with the ongoing comprehension of the previous words (for a full discussion of the backward checking mechanism, see Neely & Keefe, 1989). As noted above, in support of this possibility, Kahan et al. (1999) found some evidence of backward checking at short SOAs even in naming performance.

#### 6.6.4. *Compound-cue model*

Ratcliff and McKoon (1988) developed a model that takes a quite different approach to priming effects in the lexical decision task. The model is based on a formal model of episodic recognition memory developed by Gillund and Shiffrin (1984). In Ratcliff and McKoon’s model, items in short-term memory serve as a compound cue with the more recently presented items having a larger influence on the output of the retrieval process. If the prime and target are associated then this will provide a higher familiarity value than if the prime and target are not associated. Familiarity is then used to predict response latency via a random-walk decision process (Ratcliff, 1978), wherein, high-familiar compound cues produce relatively fast “yes” decisions and low-familiar compound cues produce relatively slow “no” decisions. Intermediate values of familiarity produce relatively slower and less accurate decisions. Hence, if familiarity is modulated by the degree to which primes and targets are either directly associated or share associates in memory, then one should find that related prime–target pairs will produce higher familiarity values and faster response latencies than unrelated prime–target pairs.

Although the compound cue model does provide an interesting alternative to prime-induced mechanisms, there are some limitations to this approach. For example, the model is primarily a model of the lexical decision task, and hence, does not account for the wealth of interesting priming data from the pronunciation task, along with other tasks. Neely’s (1991) tripartite (spreading activation, attentional expectancies, and backward checking) framework accounts for both lexical decision and pronunciation results by assuming logogen-type word recognition devices that are also connected to a phonological output system used for pronunciation. Second, and more importantly, the distinction between the compound cue model and the spreading activation framework may be more apparent than real. In both frameworks, it is necessary to map the influence of relationships between words onto priming effects. Within the spreading activation framework, this mapping involves the preactivation of related concepts in memory, whereas, within the compound cue model, this mapping is based on a rule that computes familiarity based on associations within long-term memory. At this level, the major distinction between the spreading activation framework and the compound cue model involves this mapping process.

#### 6.6.5. *Plaut and Booth's (2000) Single-mechanism connectionist model*

In contrast to Neely's tripartite framework described above, Plaut and Booth have claimed that a distributed network model can account for semantic priming lexical decision phenomena using a single mechanism. Implementing a distributed attractor network with distributed orthographic and semantic representations (Plaut, 1995), Plaut and Booth were able to account for a number of theoretically interesting findings, including the surprising observation that only participants with high perceptual ability exhibited the priming by frequency interaction (i.e., greater priming for low-frequency words); participants with low perceptual ability showed equal priming for both high- and low-frequency targets. Like the Seidenberg and McClelland (1989) model, however, the connectionist view of priming faces challenges. For example, as mentioned earlier, there is an ongoing debate about whether semantic priming is better accommodated by a single-mechanism account or by separate mechanisms that invoke distinct sets of computational principles (see Borowsky & Besner, 2006). Nevertheless, this work represents an interesting advance in that it includes a computationally implemented architecture that has been applied across a number of cognitive domains and accommodates some intriguing data in the priming literature and takes a step toward tackling the important topic of individual differences.

#### 6.6.6. *Masson's (1995) distributed memory model of semantic priming*

Masson's model, based also on distributed connectionist principles, provides a framework for accommodating semantic priming in speeded naming that neither appeals to spreading activation nor compound cues. In this model, conceptual knowledge is represented via distributed orthographic, phonological, and semantic units that are connected by weighted pathways. Importantly, Masson's network, a Hopfield (Hopfield, 1982) net variant, does not distinguish between input, hidden, and output units. The basic principle in the model is that semantically related words have very similar patterns of activation in the semantic units. When a semantically related prime is presented, activation in the semantic units starts moving toward a pattern that is similar to the pattern of activation of the to-be-presented target. When the target appears, the overlap between its pattern and the pattern of activation in the semantic units helps the phonological units converge more rapidly on the target's pattern, and hence, speeds naming responses. This model is able to account for the *intervening stimulus effect*, which, as mentioned above, is the observation that interpolating an unrelated word between the prime and the target reduces the priming effect in naming performance, a finding that the spreading activation framework does not readily predict. However, it is also the case that this model has not been extended to the wealth of data that Neely's tripartite framework appears to be able to handle.

### 6.7. Summary of Context/Priming Effects

The priming literature has provided an extremely rich data base to develop models of context effects, memory retrieval, and word recognition. Because of space limitations, we

were unable to provide a review of other important models of semantic priming effects such as Becker's (1980) verification model, Norris' (1986) plausibility-checking model, and Forster's (1976) bin model. Each of these models provides intriguing alternative perspectives on semantic priming effects. At this point in theory development, it appears that no single model of priming readily accounts for the richness and diversity of this literature, and it would appear that multiple mechanisms will need to be postulated to account for the breadth of semantic priming effects.

## 7. ATTENTIONAL CONTROL, MODULARITY, AND TIME CRITERION MODELS

The models reviewed earlier appear to have a relatively passive architecture wherein different systems accumulate information across time. However, in some instances, it may be advantageous for the reader to modulate the contribution of a given pathway depending upon the task demands or particular reading context. For example, one might expect different emphases on distinct systems when proofreading, comprehending, or checking for grammaticality. Virtually, every theory of word recognition posits multiple ways of accessing or computing the phonological code from print. In the DRC model, one can compute a phonological code via the lexical route, which maps the whole word onto a lexical representation to access phonology, or via the sublexical route, which computes the phonology via the spelling-to-sound correspondences in the language; in PDP models, the phonology can be computed by differential emphasis on the direct orthographic to phonological connections or the indirect connections via semantics. The question naturally arises whether there is any control of which processing pathway influences performance in a given task. This is important because it brings into question the modularity of the lexical processing system (see Fodor, 1983).

One way to examine the control issue is to present words that place different demands on the lexical and sublexical information. For example, within a DRC model, nonwords should bias the sublexical pathway. However, low-frequency exception words should bias the lexical pathway, since the sublexical pathway would lead to regularization errors for low-frequency exception words, i.e., pronouncing *pint* such that it rhymes with *hint*. Monsell et al. (1992) found that naming latencies to high-frequency irregular words were faster and more accurate when embedded in lists with other irregular words, than when mixed with nonwords. Monsell et al. suggested that exception word context directed attention to the lexical pathway, which is more appropriate for naming exception words, than the sublexical pathway. Additional studies have found similar influences of pathway priming (e.g., Rastle & Coltheart, 1999; Reynolds & Besner, 2005b; Simpson & Kang, 1994; Zevin & Balota, 2000).

Although intuitively appealing, the evidence for route priming has been quite controversial. Specifically, work by Kinoshita and Lupker (2002, 2003a) suggests that much of the earlier findings can be accounted for by a time criterion model. The time criterion perspective is important in a number of domains so we will briefly review it here.

Specifically, there is evidence that participants adopt a time criterion whereby they are likely to produce a response at a latency that is biased toward the average of the latencies in a block of trials. Consider the word-frequency effect (presumably a reflection of the lexical route). In two pure independent blocks, assume that a set of low-frequency words produces response latencies on the average of 700ms and a set of high-frequency words produces response latencies on the average of 600ms. If one now embeds these same words in the context of nonwords that produce an average response latency of 700ms, the word-frequency effect will likely diminish. That is, latencies to the low-frequency words will remain relatively the same (because the latencies for both low-frequency words and nonwords are 700ms), whereas latencies to the high-frequency words will increase considerably, i.e., migrate toward the time criterion invoked by mean latency of the nonwords. Hence, the word-frequency effect will decrease in the context of nonwords not because of a decreased reliance on the lexical pathway, but rather because of a change in the temporal criterion to produce a response.

The evidence clearly suggests that participants do adopt a time criterion based on the difficulty of items within a block. However, there is also evidence that appears to be consistent with a pathway control perspective above and beyond the time criterion effects. For example, all of the effects reported by Zevin and Balota (2000) hold even after the response latencies to the context items are partialled out via analyses of co-variance. Of course, if the time criterion model were the only responsible variable in this study, one should not find this pattern. Moreover, Kinoshita, Lupker, and Rastle (2004) have recently provided evidence that one can indeed modulate the lexicality effect (words faster than nonwords) via list context manipulations. However, they were unable to modulate the regularity effect (regular words faster and/or more accurate than exception words) by list context manipulations. In addition, Reynolds and Besner (2005b) have recently demonstrated that one can find lexical and sublexical pathway switching above and beyond any response latency criterion effects. Although there is accumulating evidence for some level of pathway control, further work is clearly necessary in this area. Indeed, the extent to which attentional systems modulate the information in distinct pathways has important implications for future modeling endeavors, and quite naturally would accommodate task-specific influences that have been emphasized in the present chapter. Moreover, time criterion perspectives are important in understanding how the word recognition system adjusts to the local constraints of an experiment and may have important implications for other cognitive paradigms that rely on response latency measures. At this level, time criterion effects may be viewed as an example of attentional control.

## **8. DEVELOPMENTS OF NEW APPROACHES AND ANALYTIC TOOLS TO GUIDE THE JOURNEY FROM FEATURES TO MEANING**

In the following sections, we will describe some recent developments in approaches to studying word recognition. Again, this is not a comprehensive review, but simply a brief summary to expose the reader to some of the interesting techniques that are helping researchers constrain how humans process visual words.

## 8.1. Neuroimaging Techniques

In the past decade, tremendous advances in neuroimaging methodology have provided another window into the dynamics of lexical processing (also see Just and Mason, this volume). Specifically, neuroimaging techniques like positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and event-related potentials (ERPs) now allow researchers to localize and measure the time course of activity of brain regions that are recruited by a particular cognitive task (Fiez & Petersen, 1998). One particularly exciting development has been the advent of *event-related* fMRI designs (Dale & Buckner, 1997). In PET studies and early fMRI studies, *blocked designs* (i.e., experimental conditions are blocked) were mandatory, making paradigms like the lexical decision task impractical. Event-related fMRI allows researchers to extract the fMRI BOLD (blood oxygen level dependent) response for specific trials and to conduct standard word recognition experiments in the scanner.

As Fiez and Petersen point out, neuroimaging allows one to make both coarse as well as refined fractionations of brain regions that are involved in reading. For example, Petersen, Fox, Posner, Mintun, and Raichle's (1988) seminal study elegantly demonstrated that remarkably distinct brain regions were activated by different levels of single-word processing. Specifically, at a relatively coarse level, PET scans revealed that compared to appropriate baseline conditions, occipital areas were active for passive viewing of words (orthography), temporal areas were active for reading words aloud (phonology), and frontal regions were active when participants generated verbs to nouns (semantics). By varying tasks demands and contrasting neural activation in reading aloud versus a control condition, these researchers were able to identify broadly the functions of different regions.

More recently, research designs have been employed to make finer-grained differentiations of regions that support different reading operations. For example, in a PET study of speeded naming, Fiez, Balota, Raichle, and Petersen (1999) manipulated the following three variables: lexicality (word vs. pronounceable nonword), frequency (high vs. low), and spelling-to-sound consistency (consistent vs. inconsistent). As discussed earlier, these variables have been central in the developments of models of word recognition, and so it is useful to explore the underlying circuitry. Fiez et al. (1999) found a number of noteworthy effects. First, a left frontal region showed effects of consistency and lexicality, indicating that this area may be involved in orthographic-to-phonological transformation. Second, there was greater activation for low-frequency words in a left temporal region and the supplementary motor area, which implicate these regions in the access and storage of lexical-level information. Third, effects of consistency were found bilaterally in the primary motor cortex, suggesting that consistency may influence both recognition *and* motor production systems; this surprising constraint has yet to be considered by extant theories of word recognition (but see, Kawamoto, Kello, Jones, & Bame, 1998). Fourth, the left inferior frontal gyrus showed a pattern analogous to the behavioral frequency by regularity interaction discussed earlier. Specifically, just as naming latencies are particularly slow for low-frequency inconsistent words compared to



low-frequency consistent words, high-frequency inconsistent words, and high-frequency consistent words (Seidenberg et al., 1984a), the left inferior frontal gyrus showed strong activation only for low-frequency inconsistent words. This study demonstrates how manipulating stimulus properties in a neuroimaging paradigm can be used to complement *and* extend theoretical accounts that have hitherto been based on behavioral data.

In the remainder of this section, we will review some recent neuroimaging studies and discuss how these studies contribute to our understand of word recognition. Obviously, due to space constraints, this review is selective. Also, rather than enumerating in minute detail which brain regions are activated by which task, we will be using more sweeping brushstrokes to describe the functional neuroanatomy of reading.

### 8.1.1. *Is there convergence across studies?*

A reasonable concern one may have regarding neuroimaging research is the extent to which findings generalize across laboratories and studies. Variability across studies may arise as a result of intersubject variability and slight differences in methodology, making it difficult to establish consistent regions of activation (Turkeltaub, Eden, Jones, & Zeffiro, 2002). A few articles have attempted to review results from multiple studies in order to answer this question. For example, Fiez and Petersen (1998) reviewed nine studies where participants read aloud single words, and found encouraging convergence between studies. Basically, they combined the data across the studies by merging foci from different experiments into a single figure (Figure 9).

Fiez and Petersen (1998) established that a set of areas are consistently active during word reading, including the supplementary motor area, the cerebellum, the anterior cingulate gyrus (BA 32), left-lateralized fusiform and lingual gyri (BA 18 and BA 37), the left inferior frontal gyrus (BA 44/45), bilateral activation in the anterior and posterior regions of the superior temporal gyrus (near BA 22), and dorsal and ventral portions of the post-central gyrus (near BA 4). Interestingly, a more sophisticated meta-analysis of 11 PET studies generated similar findings (also, see Price, 2004). A statistical map of

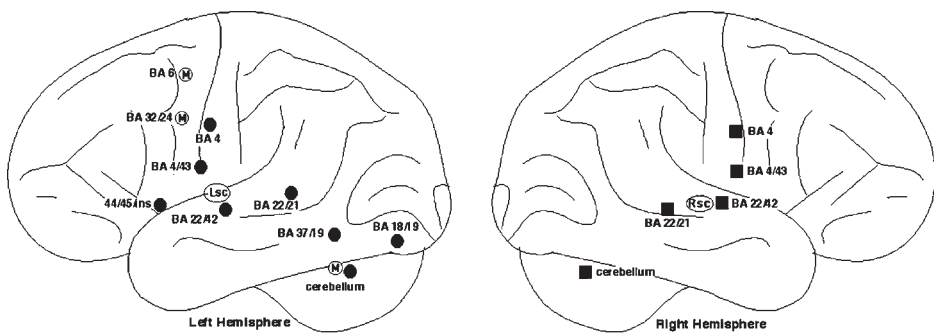


Figure 9. Brain regions that are consistently activated during word reading (Fiez & Petersen, 1998).

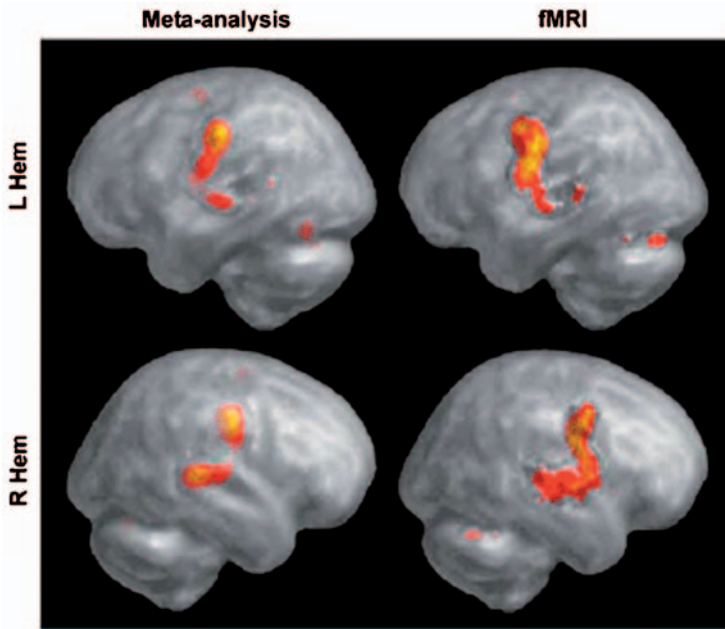


Figure 10. Statistical map generated by meta-analyses of 11 PET studies of word reading (Turkeltaub et al., 2002).

convergent foci by Turkeltaub et al. (2002) included the bilateral motor and superior temporal cortices, presupplementary motor area, left fusiform gyrus, and the cerebellum (see Figure 10). This map was successfully validated against new fMRI data of word reading, supporting the reliability of these findings.

Taken collectively, the regions identified by neuroimaging are broadly compatible with the neuropsychological literature (Fiez & Petersen, 1998; but see Price, 2004, for exceptions). More importantly, these analyses generate candidate regions of interest that can be used by researchers to test new hypotheses.

### 8.1.2. Controversies regarding targeted areas

We have already discussed how the left inferior frontal gyrus is sensitive to spelling-to-sound consistency manipulations (Fiez et al., 1999), implicating this region in processes that transform orthographic to phonological representations. This pattern has been nicely replicated in a number of other studies (see Herbster, Mintun, Nebes, & Becker, 1997; Rumsey et al., 1997). Note that some researchers have proposed another locus for the sublexical procedure, the left posterior superior temporal region (Simos et al., 2002), although this region may be associated more with phonological decomposition (Palmer, Brown, Petersen, & Schlaggar, 2004) than with orthographic-to-phonological transformation *per se*.

Interestingly, there is far less agreement about the neural markers of lexical processing. Typically, frequency and lexicality effects have been used as indicants of lexical processing. Word-frequency effects may mark brain regions involved in the access and representation of either localist or distributed lexical-level information (Fiez et al., 1999). Although word frequency is easily the most studied variable in word recognition, it has received surprisingly little attention in the functional neuroimaging literature. The literature suggests greater activation for low-frequency words in a left temporal region (near BA 22) and the supplementary motor area (BA 6) in speeded naming (Fiez et al., 1999), and greater activation in the left inferior frontal gyrus (BA 44/45) and a number of subcortical structures in lexical decision<sup>1</sup> (Fiebach, Friederici, Muller, & von Cramon, 2002).

A number of studies have also used lexicality effects (greater activation for words than nonwords) as a marker for the lexical processing (Binder et al., 2003; Fiebach et al., 2002; Ischebaeck et al., 2004; Rissman, Eliassen, & Blumstein, 2003; Simos et al., 2002). It seems plausible that greater activations for words reflect access to linguistic information, which may either be orthographic, phonological, or semantic (Fiez et al., 1999). Unfortunately, there seems to be little consensus on regions that show greater activation for words. Lexical decision studies have identified diverse regions, including the left angular gyrus (BA 39), left dorsal prefrontal cortex (BA 6/8), superior frontal gyri (BA 6/8/9), left rostral–ventral cingulate gyrus (BA 32/24), left posterior cingulate gyrus and precuneus (BA 23/29–31/7), and the junction of the left posterior middle temporal and inferior temporal gyri (BA 21/37). Curiously enough, in speeded naming, it is relatively difficult to find regions that show greater activations for words than nonwords. For example, in the Fiez et al. (1999) PET study, no region showed greater activation in the word condition compared to the nonword condition. This may be attributable to the lower spatial resolution of PET (compared to fMRI, but see Palmer, 2003), or to strategic effects induced by blocking. Using a more sensitive event-related paradigm with Japanese Kana words, Ischebeck et al. (2004) found greater activation for words in the left and right temporo–parietal areas (BA 39/40), the middle part of the left middle/inferior temporal gyrus (BA 21/20) and the posterior cingulate (BA 31).

The marked discrepancy between the lexical decision and naming findings may be partly attributable to differential demands of the two tasks; a theme that has consistently arisen in the current chapter. If indeed there are large task differences in the behavior, there clearly should be consequences for the neural underpinnings. For example, the meta-analyses we discussed earlier (e.g., Fiez & Petersen, 1998; Turkeltaub et al., 2002) were based on neuroimaging studies of naming. It will be interesting to see the degree to which statistical maps based on meta-analyses of *lexical decision* data show a similar pattern. This is a theoretically important question that has yet to be answered.

---

<sup>1</sup> The latter finding needs to be qualified; the study was done with German stimuli, which have regular spelling-to-sound correspondences. The regularity of the orthography may have facilitated sublexical processing (reflected in left inferior frontal gyrus activation) and attenuated lexical activity.

Nevertheless, in spite of the diversity of findings, two regions seem to be consistently associated with lexicality effects (word > nonword), the left angular gyrus (BA 39) and the left middle temporal gyrus (BA 21), a pattern which is nicely consistent with the Kana-naming study (Ischebeck et al., 2004). The left middle temporal gyrus has long been associated with language processing (Fiebach et al., 2002); other studies also implicate the region in the representation and processing of lexico-semantic (Pugh et al., 1996) and phonological word forms (Cohen et al., 2000). The left angular gyrus also seems to play a role in non modality-specific semantic processing (Binder et al., 2003).

PET and fMRI are not the only windows into the functional neuroanatomy of reading. While these measures have excellent spatial resolution, the intrinsic characteristics of these signals limit their temporal resolution. Event-related potentials (ERPs) have far more exquisite temporal resolutions, and so these measures are better suited to study the time course of word recognition processes. Scalp-measured ERPs reflect the brain electrical activity that is triggered by experimental stimuli, and capture in real-time cognitive processes on a millisecond basis (Kutas & Federmeier, 2000; Kutas and van Patten, this volume). Frequency effects, for example, are apparent in ERPs between 200 and 400 ms, and are most obviously correlated with a left anterior negative component called the lexical processing negativity (LPN) (King & Kutas, 1998). In a recent representative ERP study, Hauk and Pulvermüller (2004) investigated how word length and word frequency influenced the amplitude and peak latencies of event-related potentials in lexical decision. Early effects of length and frequency were observed, and the researchers interpreted the results as consistent with lexical access occurring as early as 150ms after the onset of visually presented words. More intriguingly, large length effects were observed in ERPs but not in the behavioral data; this dissociation demonstrates that psychophysiological measures may in some cases be more sensitive than behavioral data.

To recapitulate, in the foregoing discussion, we have briefly considered the neuroanatomical correlates of selected psycholinguistic effects. Clearly, this nascent work is exciting and informative, and many issues remain unexplored. In the final portion of this section, we will consider how neuroimaging has advanced our understanding of word recognition processes above and beyond traditional behavioral work.

### *8.1.3. What constraints are afforded by neuroimaging techniques?*

It is incontrovertible that we know a great deal more about the neuroanatomy of language today than a mere 10 years ago. Nevertheless, it is also clear that neuroimaging of cognitive processes is still a relatively new area of investigation. Even though the conclusions we have presented are preliminary and may be revised not too far in the future, we would contend that neuroimaging data is an essential adjunct to response latency and accuracy data. Most obviously, neural correlates of behavior provide another level of explanation (Marr, 1982) that reveals how reading processes are physically instantiated in the brain. Moreover, Palmer et al. (2004) have cogently argued that collecting functional neuroimaging data affords two other important advantages. One, brain activation data can powerfully complement behavioral measures. For example, young and older adults may perform identically on a task (in terms of response latencies and error rates),

but show marked differences in brain activity (also see Hauk & Pulvermüller, 2004; Schlaggar et al., 2002). In addition, neuroimaging data may also be useful in informing theories and adjudicating between competing models. For example, the DRC model (Coltheart et al., 2001) and the connectionist model (Plaut et al., 1996) adopt very different architectures for naming words that have been difficult to discriminate based on behavioral data, but it may be possible that converging evidence of the role of different neural substrates dedicated to specific operations may provide important information on how the brain implements the processes involved in word recognition.

## **8.2. Large-Scale Studies vs. Factorial Studies of Word Recognition**

Word recognition researchers have traditionally employed factorial designs where item variables of interest (e.g., length, frequency, etc.) have been “manipulated,” and other factors known to affect performance have been controlled. This approach has been useful, but there are some limitations (see Balota et al., 2004). Recently, researchers have examined word recognition performance for large sets of words that are not constrained by selection factors, e.g., virtually all monosyllabic words (see Balota & Spieler, 1998; Besner & Bourassa, 1995; Kessler, Treiman, & Mullennix, 2002; Spieler & Balota, 1997; Treiman et al., 1995). Such datasets are useful in a number of ways. For example, using standard predictor variables, Balota et al. (2004) accounted for 49 and 50 percent of the variance in the lexical decision and speeded naming performance, respectively for a dataset of 2428 words. This is a multifold increase over current computational models (for a discussion of pros and cons for using accounted for variance as a critical variable in evaluating a model’s performance, see Balota & Spieler, 1998; Seidenberg & Plaut, 1998). This outcome was obtained despite the success these computational models have had in accounting for performance at the factor level. The large-scale item-level analyses provide another potentially important constraint in the evaluation of theoretical approaches to word processing. More recently, Balota and colleagues have collected naming and lexical decision latencies for over 40,000 words (Balota et al., 2002; Balota et al., in press). The English Lexicon Project website (<http://ellexicon.wustl.edu>) provides a comprehensive data set of behavioral measures that researchers can easily access, via a search engine, along with a rich set of descriptive characteristics. Hopefully, this dataset will be helpful in extending current models to multisyllabic words, which as noted above is a potentially serious limitation in current models. Finally, as mentioned earlier, recent attempts to ground semantics in large scale natural databases of language use (e.g., Burgess & Livesay, 1998; Landauer & Dumais, 1997; Steyvers & Tenenbaum, 2005) have also been quite informative. Clearly, the computational power available today that affords analyses of these large-scale databases appears to be providing an important additional constraint on theory development.

## **8.3. RT Distributional Analyses**

In standard word recognition experiments, one compares the mean response latency across several conditions to determine if the predictions generated by an experimental hypothesis are correct. Of course mean performance is not the only estimate of performance across a set of trials. Researchers have long noted that means of conditions are only

one estimate available from performance. For example, in the standard Stroop task (i.e., naming the color that a word appears in), Heathcote, Popiel, and Mewhort (1991) provided a useful demonstration of how the shape of a response time distribution can provide useful information beyond estimates of central tendency. They found that the incongruent condition (e.g., the word *blue* appearing in the color red), compared to the neutral condition (e.g., the word *block* appearing in the color red), increased both the skewing and the central tendency of the reaction time distribution, but amazingly, the congruent condition (e.g., the word *red* appearing in the color red) *increased* skewing and *decreased* the central tendency, which basically masked any effect in means (for a replication of this pattern, see Spieler, Balota, & Faust, 1996). These researchers have fit reaction time distributions to ex-Gaussian functions, but other functions such as the Weibull or ex-Wald could also capture useful characteristics of the reaction time distributions. As theories become more precise regarding item level performance, there should be an increased level of sophistication regarding the predictions concerning the underlying reaction time distributions. For example, Balota and Spieler (1999) found that frequency and repetition influenced these parameters differently depending on the dependent measures, i.e., naming vs. lexical decision (however, see Andrews & Heathcote, 2001). Ratcliff et al. (2004) have recently used reaction time distributions to more powerfully test a diffusion model of lexical decision performance (see also Yap, Balota, Cortese, & Watson, in press). As models become more sophisticated, the precision of reaction time distribution analyses will be critical in their evaluation.

#### 8.4. Individual Differences

Just as one may be losing information when averaging across items to estimate means, one is also losing information when averaging across individuals. Of course, there are standard comparisons of individual differences as a function of age, acquired or developmental dyslexia, or other neuropsychological impairment (see Perfetti, this volume), however, another possibility is that individuals may produce particular profiles of lexical processing. For example, if indeed the dual-route model is correct, one might find that some subjects rely more on lexical pathways, while other subjects rely more on sublexical pathways, and this could indeed be tied to the manner in which they were originally taught to read or inherent individual differences in capacities. The recent explosion of interest in differences in working memory capacity has been quite successful in identifying distinct cognitive processing profiles (see, for example, Engle, Kane, & Tuholski, 1999). With the advent of large datasets on individual subjects (see megastudies mentioned earlier) it is quite possible that such differences could be observed (for processing speed modulating the effects of orthographic neighborhood size, see Balota et al., 2004). Of course, this may also push researchers to more closely consider the reliability of effects, which at least within one domain, semantic priming, appear to be surprisingly low (see Stolz, Besner, & Carr, 2005).

### 9. CONCLUDING REMARKS

In the present chapter, we have attempted to provide the reader with an overview of the major issues addressed in the word recognition literature. To conclude, we would like to summarize some of the major themes that have spanned a number of the sections.

First, in each of the sections, there has been evidence initially supporting a rather straightforward theoretical analysis and then there have been reports by “trouble-makers” that constrain the strength of the theoretical inferences available from a given task. For example, even in the word superiority paradigm, there have been arguments that partial information from the target letter could, in conjunction with the word-envelope, allow subjects to use a sophisticated guessing strategy to bias the correct choice (e.g., Krueger & Shapiro, 1979; Massaro, 1979). If this is the case, then the word-superiority effect may not reflect top-down impacts in perception, but rather, biases that occur at post-perceptual levels, based on partial information. Similar concerns were raised about the threshold identification, lexical decision, and pronunciation tasks. Of course, task analyses can be frustrating for theoreticians, however, before inferences can be made regarding the underlying locus or loci of a given variable, one should be especially careful in developing (or understanding) tasks that faithfully reflect such processes. Clearly, the adequacy of any theory rests on the quality of the tasks used to build that theory.

A second consistent theme that has surfaced in this review is whether there are separable analyses performed en route to word recognition or the apparent influences of multiple pathways are in large part merely a consequence of the activation and inhibition patterns across many lexical representations. Although some effects appear to be modeled quite well by interactive activation and parallel distributed processing systems, there have also been results that appear inconsistent with such systems. There are at least two likely outcomes to this area of work. First, more of the troublesome effects may fall from these models when networks that are closer to the size of an adult’s vocabulary are implemented (see Seidenberg & McClelland, 1990). Second, it may be necessary to implement sublexical processing modules within such connectionist models to incorporate the strong evidence for multiple distinct access pathways. Clearly, this is still a central issue in the current state of model development (see Andrews, 2006).

A third theme in the present review is the type of statistical interaction that has been repeatedly observed. The vast majority of interactions in this literature are of the nature that Factor A has more of an effect at the level of Factor B that produces the slowest or least accurate performance. Consider for example word frequency. We have reviewed evidence indicating that compared to high-frequency words, low-frequency words produce larger effects of bigram frequency, spelling-to-sound consistency, word-body strength, concreteness, semantic priming, task (lexical decision task vs. category verification vs. pronunciation), repetition priming, neighborhood size, among others. There are at least two noteworthy aspects of these interactions. First, one may wish to argue that because of the development of automaticity, high-frequency words are recognized via routes that effectively bypass many sublexical stages of analyses. Hence, if one is interested in identifying many of the intriguing sublexical aspects of word recognition, one should primarily investigate the processing of low-frequency words. Alternatively, as Loftus (1978) has noted, on a simply statistical level, this particular type of interaction is one of the most difficult to interpret. In fact, it is possible, that if one considered percentage of overall response latency change as a function of the levels for Factor A and B, or a z-score transform of the data (taking into account variability), many of these interactions would

disappear (for a detailed discussion of these issues, see Faust, Balota, Spieler, & Ferraro, 1999). Clearly, the assumption of a linear relations between response latency and underlying cognitive operations is a simplifying assumption, which will ultimately need to be faced by those studying the time-course of processes involved in visual word recognition, along with other cognitive operations.

In sum, we are hopeful that the reader agrees that at some level the word is to cognitive psychologists and psycholinguist as the cell is to biologists. Both entail many substructures and interact with many higher-level systems. The present overview of the word recognition literature may seem rather imposing, and sometimes it would appear that little progress is being made. However, this clearly is *not* the case; considerable progress has been made, especially within the last decade. Of course, the seductive simplicity of understanding lexical-level analyses surely is more apparent than real. As is often the case in a discipline, the more we know about a system, the more we develop procedures for generating and constraining our questions in the future. Given the new analytic methods that have come on line recently this will indeed be a very exciting next decade of research.

## ACKNOWLEDGMENTS

This work has been supported by Grant BCS 0001801 from the National Science Foundation and Grants AGO3991 and RO1 AG1702 from the National Institute on Aging. Thanks are extended to Matthew Traxler for helpful comments on an earlier version of this chapter.

## REFERENCES

- Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2*, 153–157.
- Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11*, 118–124.
- Adams, M. (1981). What good is orthographic redundancy? In: H. Singer, & O. J. L. Tzeng (Eds), *Perception of print* (pp. 197–221). Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Hemisphere Press.
- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory and Cognition, 10*, 565–575.



- Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802–814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234–254.
- Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, *35*, 775–800.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, *4*, 439–461.
- Andrews, S. (Ed.). (2006). *From inkmarks to ideas: Current issues in lexical processing*. Hove, UK: Psychology Press.
- Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 514–544.
- Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnaw wirts? *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1052–1086.
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multi-trace memory model of polysyllabic word reading. *Psychological Review*, *105*, 678–723.
- Appelman, I. B., & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception and Psychophysics*, *30*, 436–446.
- Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review*, *89*, 599–607.
- Azuma, T., & Van Orden, G. C. (1997). Why safe is better than fast: The relatedness of a word's meaning affects lexical decision times. *Journal of Memory and Language*, *36*, 484–504.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Baayen, R. H., & Schreuder, R. (Eds.). (2003). *Morphological structure in language processing*. Berlin: Mouton de Gruyter.
- Baddeley, A. D., Eldridge, M., & Lewis, V. J. (1981). The role of subvocalization in reading. *Quarterly Journal of Experimental Psychology*, *33A*, 439–454.
- Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior*, *22*, 88–104.
- Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1289–1302.

Balota, D. A., Black, S. R., & Cheney, M. (1992). Automatic and attentional priming in young and older adults: Reevaluation of the two-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 485–502.

Balota, D. A., Boland, J. E., & Shields, L. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, *28*, 14–36.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*, 89–106.

Balota, D. A., & Chumbley, J. I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comment on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General*, *119*, 231–237.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). *The English lexicon project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Accessed January 30, 2004, on the Washington University Web site: <http://lexicon.wustl.edu>.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Balota, D. A., & Duchek, J. M. (1989). Spreading activation in episodic memory: Further evidence for age independence. *Quarterly Journal of Experimental Psychology*, *41A*, 84–93.

Balota, D. A., Ferraro, F. R., Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In: P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–218). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 336–345.

Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 827–845.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition*, *29*, 639–647.

Balota, D. A., & Rayner, K. (1991). Word recognition processes in foveal and parafoveal vision: The range of influence of lexical variables. In: D. Besner, & G. W. Humphreys (Eds), *Basic processes in reading* (pp. 198–232). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science*, *9*, 238–240.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*, 32–55.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (in press). The English Lexicon Project. *Behavior Research Methods*.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462–479.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244.

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*, 207–228.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 252–259.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition*, *8*, 493–512.

Becker, C. A. (1985). What do we really know about semantic context effects during reading? In: D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 125–166). Toronto: Academic Press.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 389–401.

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *Quarterly Journal of Experimental Psychology*, *39A*, 467–478.

Besner, D. (1990). Does the reading system need a lexicon? In: D. A. Balota, G. B. Flores d'Arcaise, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 73–99). Hillsdale, NJ: Erlbaum.

Besner, D., & Bourassa, D. C. (1995). Localist and parallel distributed processing models of visual word recognition: A few more words. Paper presented at the Annual Meeting of the Canadian Brain, Behaviour, and Cognitive Science Society, Halifax, NS, Canada, June.

Besner, D., Davelaar, E., Alcott, D., & Parry, P. (1984). Basic processes in reading: Evidence from the FDM and the FBI. In: L. Henderson (Ed.), *Orthographies and reading* (pp. 121–135). Hillsdale, NJ: Erlbaum.

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and production: An examination of four classes of models. In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 201–219). Hillsdale, NJ: Erlbaum.

- Besner, D., Smith, M. C., & MacLeod, C. M. (1990a). Visual word recognition: A dissociation of lexical and semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 862–869.
- Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990b). On the association between connectionism and data: Are a few words necessary? *Psychological Review*, *97*, 432–446.
- Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., & Buchanan, L. (2003). Neural correlates of lexical access during visual word recognition. *Journal of Cognitive Neuroscience*, *15*, 372–393.
- Bjork, E. L., & Estes, W. K. (1973). Letter identification in relation to linguistic context and masking conditions. *Memory and Cognition*, *1*, 217–223.
- Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 582–594.
- Bodner, G. E., & Masson, M. E. J. (1997). Masked repetition priming of words and nonwords: Evidence for a nonlexical basis for priming. *Journal of Memory and Language*, *37*, 268–293.
- Boles, D. B. (1983). Dissociated imageability, concreteness, and familiarity in lateralized word recognition. *Memory & Cognition*, *11*, 511–519.
- Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory & Language*, *50*, 456–476.
- Borowsky, R., & Besner, D. (1991). Visual word recognition across orthographies: On the interaction between context and degradation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 272–276.
- Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 813–840.
- Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical–semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review*, *113*, 181–195.
- Bradley, D. (1979). Lexical representations of derivational relations. In: M. Aranoff, & M. Kean (Eds), *Juncture* (pp. 37–55). Cambridge, MA: MIT Press.
- Broadbent, D. E. (1967). Word-frequency effects and response bias. *Psychological Review*, *74*, 1–15.
- Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and naming latency. *Memory & Cognition*, *15*, 208–216.
- Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*, 531–544.

- Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting RT in a basic word recognition task: Moving on from Kucera and Francis. *Behavior Research Methods, Instruments, & Computers*, 30, 272–277.
- Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In: R. Frost, & L. Katz (Eds), *Orthography, phonology, morphology, and meaning* (pp. 211–226). Amsterdam: North-Holland.
- Carr, T. H., & Dagenbach, D. C. (1990). Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 341–350.
- Carr, T. H., Davidson, B. J., & Hawkins, H. L. (1978). Perceptual flexibility in word recognition: Strategies affect orthographic computation but not lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 674–690.
- Carr, T. H., McCauley, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 757–777.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 857–871.
- Cattell, J. M. (1885). The inertia of the eye and brain. *Brain*, 8, 295–312.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11, 63–65.
- Catlin, J. (1969). On the word frequency effect. *Psychological Review*, 76, 504–506.
- Catlin, J. (1973). In defense of sophisticated guessing theory. *Psychological Review*, 80, 412–416.
- Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, 18, 225–241.
- Chateau, D., & Jared, D. (2003). Spelling-sound consistency effects in disyllabic word naming. *Journal of Memory and Language*, 48, 255–280.
- Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, 36, 387–395.
- Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory and Cognition*, 12, 590–606.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area. Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123, 291–307.
- Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407–428.

- Collins, A., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240–247.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In: G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London : Academic Press.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608.
- Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In: S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1197–1220.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- Connine, C., Mullennix, J., Shernoff, E., & Yelens, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 1084–1096.
- Cortese, M. J. (1998). Revisiting serial position effects in reading aloud. *Journal of Memory and Language*, 39, 652–666.
- Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments and Computers*, 36, 384–387.
- Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory and Cognition*, 28, 1269–1276.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- Dagenbach, D., Carr, T. H., & Wilhelmsen, A. (1989). Task-induced strategies and near-threshold priming: Conscious influences on unconscious perception. *Journal of Memory and Language*, 28, 412–443.
- Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5, 329–340.
- Damian, M. F. (2001). Congruity effects evoked by subliminally presented primes: Automaticity rather than semantic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 154–165.
- Davis, C., & Castles, A., & Iakovidis, E. (1998). Masked homophone and pseudohomophone priming in children and adults. *Language and Cognitive Processes*, 13, 625–651.
- Davis, C. J. (1999). *The self-organizing lexical acquisition and recognition (SOLAR) model of visual word recognition* Doctoral dissertation, University of New South Wales, Sydney, New South

Wales, Australia, 1999. *Dissertation Abstracts International*, 62, 594. Available on the Web at [www.maccs.mq.edu.au/~colin/thesis.zip](http://www.maccs.mq.edu.au/~colin/thesis.zip).

Davis, C. J., & Bowers, J. S. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception and Performance*, 30, 923–941.

Day, J. (1977). Right-hemisphere language processing in normal right-handers. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 518–528.

de Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior*, 22, 417–436.

de Groot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of related prime-target pairs and the stimulus onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology*, 36A, 253–280.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word associations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 824–845.

de Groot, A. M. B., Thomassen, A., & Hudson, P. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory and Cognition*, 10, 358–370.

De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion*, 12, 145–169.

Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, 395, 597–600.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

den Heyer, K., Briand, K., & Dannenbring, G. (1983). Strategic factors in a lexical decision task: Evidence for automatic and attention-drive processes. *Memory & Cognition*, 11, 374–381.

den Heyer, K., Sullivan, A., & McPherson, C. (1987). *Mediated priming in a single-response lexical decision task*. Unpublished manuscript.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory & Language*, 27, 429–446.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In: A. Miyake, & P. Shah, (Eds), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). London, UK: Cambridge Press.

- Erdmann, B., & Dodge, R. (1898). *Psychologische Untersuchungen über das Lesen*. Halle, Germany: M. Niemery.
- Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology*, *33A*, 325–350.
- Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*, 777–799.
- Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in first- and second-language reading of fluent bilinguals. *Memory & Cognition*, *11*, 565–574.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15*, 115–141.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.
- Feldman, L. B., & Andjelkovic, D. (1992). Morphological analysis in word recognition. In: R. Frost, & L. Katz (Eds), *Phonology, orthography, morphology, and meaning* (pp. 343–360). New York: North-Holland Elsevier Science Publishers.
- Feldman, L. B., & Basnight-Brown, D. (2005). The role of morphology in visual word recognition: Stem senses and semantic richness. In: E. Grigorenko, A. Naples (Eds), *Single word reading: Cognitive, behavioral and biological perspectives*. Mahwah, NJ: Lawrence Erlbaum, in press.
- Fiebach, C. J., Friederici, A. D., Muller, K., & von Cramon, D. Y. (2002). fMRI evidence of dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience*, *14*, 11–23.
- Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron*, *24*, 205–218.
- Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Science*, *95*, 914–921.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Forster, K. I. (1976). Accessing the mental lexicon. In: R. J. Wales, & E. C. T. Walker (Eds), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In: W. E. Cooper, & E. Walker (Eds), *Sentence processing: Psychological studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical context on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, *33A*, 465–495.



- Forster, K. I. (1985). Lexical acquisition and the modular lexicon. *Language and Cognitive Processes, 1*, 87–108.
- Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27*, 203–233.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680–689.
- Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30*, 1–25.
- Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 696–713.
- Fowler, C. A., Wolford, G., Slade, R., & Tassinary, L. (1981). Lexical access with and without awareness. *Journal of Experimental Psychology: General, 110*, 341–362.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*, 71–99.
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104–115.
- Gernsbacher, M. A. (Ed.). (1994). *Handbook of psycholinguistics*. San Diego: Academic Press.
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256–280.
- Gibson, E., Osser, H., Schiff, W., & Smith, J. (1963). An analysis of critical features of letters, tested by a confusion matrix. In: *A basic research program on reading* (Cooperative Research Project No. 639). Washington, DC: U.S. Office of Education.
- Gibson, J. J., & Gibson, E. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*, 32–41.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.
- Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691.
- Gold, B. T., Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Snyder, A. Z., Salat, D. H., Fischl, B., Dale, A. M., Morris, J. C., & Buckner, R. L. (2005). Differing neuropsychological and

neuroanatomical correlates of abnormal reading in early-stage semantic dementia and dementia of the Alzheimer type. *Neuropsychologia*, 43, 833–846.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2005). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology (submitted).

Goodman, G., McClelland, K., & Gibbs, R. (1981). The role of syntactic context in word recognition. *Memory & Cognition*, 9, 580–586.

Gough, P. B., Alford, J. A., & Holley-Wilcox, P. (1981). Words and contexts. In: O. L. Tzeng, & H. Singer (Eds), *Perception of print: Reading research in experimental psychology* (pp. 85–102). Hillsdale, NJ: Erlbaum.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29, 228–244.

Grainger, J. (1992). Orthographic neighborhoods and visual word recognition. In: R. Frost (Ed.), *Orthography, phonology, morphology, and meaning* (pp. 131–146). Oxford, England: North-Holland.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518–565.

Greenwald, A. G., Abrams, R. L., Naccache, L., & Dehaene, S. (2003). Long-term semantic memory versus contextual memory in unconscious number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 235–247.

Haas, W. (1970). *Phonographic translation*. Manchester: Manchester University Press.

Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H. (1966). *Phoneme-grapheme correspondences as cues to spelling improvement*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Harm, M. W., & Seidenberg, M. S. (2004) Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111, 662–720.

Harm, M., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106, 491–528.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115, 1090–1103.

Healy, A. F., & Drewnowski, A. (1983). Investigating the boundaries of reading units: Letter detection in misspelled words. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 413–426.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340–347.

Henderson, L. (1982). *Orthography and word recognition in reading*. London: Academic Press.

Henik, A., Freidrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effects upon prime processing. *Memory and Cognition*, *11*, 366–373.

Herbster, A. N., Mintun, M. A., Nebes, R. D., & Becker, J. T. (1997). Regional cerebral blood-flow during word and non-word reading. *Human Brain Mapping*, *5*, 84–92.

Hines, D., Czerwinski, M., Sawyer, P. K., & Dwyer, M. (1986). Automatic semantic priming: Effect of category exemplar level and word association level. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 370–379.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1331–1356.

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 686–713.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74–95.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences*, *9*, 1–66.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Erlbaum.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science*, *79*, 2554–2558.

Howard, I. (1972). *A directional theory of rule application in phonology*. Dissertation at Massachusetts Institute of Technology.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Humphreys, G. W., Besner, D., & Quinlan, P. T. (1988). Event perception and the word repetition effect. *Journal of Experimental Psychology: General*, *117*, 51–67.

Humphreys, G. W., Evett, L. J., & Taylor, D. E. (1982). Automatic phonological priming in visual word recognition. *Memory and Cognition*, *10*, 576–590.

- Humphreys, G. W., Evett, L. J., Quinlan, P. T., & Besner, D. (1987). Priming from identified and unidentified primes. In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 105–126). London, UK: Erlbaum.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A micro-analytic review. *Psychonomic Bulletin and Review*, *10*, 785–813.
- Ischebeck, A., Indefrey, P., Usui, N., Nose, I., Hellwig, F., & Taira, M. (2004). Reading in a regular orthography: An fMRI study investigating the role of visual familiarity. *Journal of Cognitive Neuroscience*, *16*, 727–741.
- James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 130–136.
- Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529.
- Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory & Language*, *46*, 723–750.
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715.
- Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 92–105.
- Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General*, *120*, 358–394.
- Jastrzemski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, *13*, 278–305.
- Johnston, J. C., & McClelland, J. L. (1973). Visual factors in word perception. *Perception and Psychophysics*, *14*, 365–370.
- Johnson, N. F. & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, *26*, 240–346.
- Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, *96*, 29–44.
- Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 483–491.
- Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture processing. *Psychological Bulletin*, *131*, 684–712.
- Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1312–1318.

Juola, J. F., Leavitt, D. D., & Choe, C. S. (1974). Letter identification in word, nonword, and single letter displays. *Bulletin of the Psychonomic Society*, 4, 278–280.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.

Kahan, T. A., Neely, J. H., & Forsythe, W. J. (1999). Dissociated backward priming effects in lexical decision and pronunciation tasks. *Psychonomic Bulletin & Review*, 6, 105–110.

Kawamoto, A., Kello, C., Jones, R., & Bame, K. (1998). Initial phoneme versus whole word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 862–885.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or the body in question. Further dissection of the processes of sub-lexical spelling-sound translation. In: M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 449–470). London: Erlbaum.

Kay, J., & Marcel, A. J. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *The Quarterly Journal of Experimental Psychology*, 37A, 39–81.

Keefe, D. E., & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies. *Memory and Cognition*, 18, 289–298.

Kessler, B., & Treiman, R. (2001). Relations between sounds and letters in English monosyllables. *Journal of Memory and Language*, 44, 592–617.

Kessler, B., & Treiman, R. (2003). Is English spelling chaotic? Misconceptions concerning its irregularity. *Reading Psychology*, 24, 267–289.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language*, 47, 145–171.

Kessler, B., Treiman, R., & Mullennix, J. (in press). Feedback consistency effects in single-word reading. In E. J. Grigorenko & A. Naples (Eds.), *Single-word reading: Biological and behavioral perspectives*. Mahwah, NJ: Erlbaum.

Kiger, J. I., & Glass, A. (1983). The facilitation of lexical decisions by a prime occurring after the target. *Memory and Cognition*, 11, 356–365.

King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, 224, 61–64.

Kinney, G. C., Marsetta, M., & Showman, D. J. (1966). *Studies in display symbol legibility. Part XII. The legibility of alphanumeric symbols for digitized television* (ESD-TR066–117). Bedford, MA: The Mitre Corporation.

Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin and Review*, 7, 133–141.

- Kinoshita, S., & Lupker, S. J. (2002). Effects of filler type in naming: Change in time criterion or attentional control of pathways? *Memory and Cognition*, 30, 1277–1287.
- Kinoshita, S., & Lupker, S. J. (2003a). Priming and attentional control of lexical and sublexical pathways in naming: A re-evaluation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 405–415.
- Kinoshita, S., & Lupker, S. J. (Eds.). (2003b). *Masked priming: State of the art*. New York: Psychology Press.
- Kinoshita, S., Lupker, S. J., & Rastle, K. (2004) Modulation of regularity and lexicality effects in reading aloud. *Memory and Cognition*, 32, 1255–1264.
- Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory and Cognition*, 9, 587–598.
- Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 92–107.
- Krueger, L. E., & Shapiro, R. G. (1979). Letter detection with rapid serial visual presentation. Evidence against word superiority of feature extraction. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 657–673.
- Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kunde, W., Kiesel, A., & Hoffman, J. (2003). Conscious control over the content of unconscious cognition. *Cognition*, 88, 223–242.
- Kutas, M., & Federmeier, K. D. (2000): Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4, 463–470.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Lambert, A. J., Payne, B. K., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L., & Khan, S. K. (2003). Stereotypes as dominant responses: On the “social facilitation” of prejudice in anticipated public context. *Journal of Personality and Social Psychology*, 84, 277–295.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lee, Y., Binder, K. S., Kim, J., Pollatsek, A., & Rayner, K. (1999). Activation of phonological codes during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 948–964.
- Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition*, 78, 189–205.

- Lima, S. D. (1987). Morphological analysis in sentence reading. *Journal of Memory & Language*, 26, 84–99.
- Loftus, G. R. (1978). On interpretation of interactions. *Memory and Cognition*, 6, 312–319.
- Lorch, R. F., Balota, D. A., & Stamm, E. G. (1986). Locus of inhibition effects in the priming of lexical decision: Pre- or post-lexical access? *Memory and Cognition*, 14, 95–103.
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*, 7, 618–630.
- Luce, P. A., & Pisoni, D. B. (1989). *Neighborhoods of words in the mental lexicon*. Unpublished manuscript.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, 123, 107–128.
- Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior*, 23, 709–733.
- MacKay, D. (1972). The structure of words and syllables: Evidence from words and speech. *Cognitive Psychology*, 3, 210–227.
- Manelis, L. (1974). The effect of meaningfulness in tachistoscopic word perception. *Perception and Psychophysics*, 16, 182–192.
- Manso de Zuniga, C. M., Humphreys, G. W., & Evett, L. J. (1991). Additive and interactive effects of repetition, degradation, and word frequency in the reading of handwriting. In: D. Besner, & G. Humphreys (Eds), *Basic processes in reading visual word recognition* (pp. 10–33). Hillsdale, NJ: Erlbaum.
- Marcel, A. J. (1980). Surface dyslexia and beginning reading: A revised hypothesis of pronunciation of print and its impairments. In: M. Coltheart, K. Patterson, & J. C. Marshall (Eds), *Deep dyslexia* (pp. 227–258). London: Routledge & Kegan Paul.
- Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, 15, 197–237.
- Marcel, A. J., & Patterson, K. E. (1978). Word recognition and production: Reciprocity in clinical and normal studies. In: J. Requin (Ed.), *Attention and performance VII* (pp. 209–226). Hillsdale, NJ: Erlbaum.
- Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.
- Marshall, J. C., & Newcombe, F. (1980). The conceptual status of deep dyslexia: An historical perspective. In: M. Coltheart, K. Patterson, & J. C. Marshall (Eds), *Deep dyslexia* (pp. 1–21). London: Routledge & Kegan Paul Ltd.

- Massaro, D. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 595–609.
- Massaro, D. M., Venezky, R. L., & Taylor, G. A. (1979). Orthographic regularity, positional frequency and visual processing of letter strings. *Journal of Experimental Psychology: General*, 108, 107–124.
- Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1107–1128.
- Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In: D. Besner, & G. W. Humphreys (Eds), *Basic processes in reading: Visual word recognition* (pp. 233–263). Hillsdale, NJ: Erlbaum.
- Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 3–23.
- Masson, M. E. J., & Isaak, M. I. (1999). Masked priming of words and nonwords in a naming task: Further evidence for a nonlexical basis for priming. *Memory and Cognition*, 27, 399–412.
- Mayall, K., & Humphreys, G. W. & Olson, A. C. (1997). Disruption to word or letter processing? The origins of case mixing effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1275–1286.
- Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1, 13–32.
- Mayzner, M. S., Tresselt, M. E., & Wolin, B. R. (1965a). Tables of pentagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1, 144–186.
- Mayzner, M. S., Tresselt, M. E., & Wolin, B. R. (1965b). Tables of pentagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1, 79–143.
- Mayzner, M. S., Tresselt, M. E., & Wolin, B. R. (1965c). Tables of pentagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1, 33–78.
- McCarthy, R., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, 22, 359–380.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception and Psychophysics*, 22, 249–261.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.



- McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.
- McGinnies, E., Comer, P. B., & Lacey, O. L. (1952). Visual-recognition thresholds as a function of word length and word frequency. *Journal of Experimental Psychology*, *44*, 65–69.
- McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1173–1190.
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Hove, UK: Psychology Press.
- McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language*, *27*, 545–559.
- Merikle, P. M. (1982). Unconscious perception revisited. *Perception and Psychophysics*, *31*, 298–301.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.
- Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–312.
- Millis, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition*, *17*, 141–147.
- Monaghan, J., & Ellis, A. W. (2002). What, exactly, interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 183–206.
- Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). The effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*, 43–71.
- Monsell, S., Patterson, K., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translations of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 452–467.
- Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 116–153.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
- Morton, J. (1970). A functional model for memory. In: D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*, 721–756.

- Naccache, L., & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*, 215–229.
- Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition*, *4*, 648–654.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: D. Besner, & G. Humphreys (Eds), *Basic processes in reading: Visual word recognition* (pp. 236–264). Hillsdale, NJ: Erlbaum.
- Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In: H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds), *The nature of remembering: Essays in honor of Robert G. Crowder*, (pp. 69–93). Washington, DC: American Psychological Association.
- Neely, J. H. & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective/retrospective processing theory. In: G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207–248). New York: Academic Press.
- Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1003–1019.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton, Century, Crofts.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved June 1, 2000, from the University of South Florida Web site: <http://www.usf.edu/FreeAssociation/>.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin and Review*, *13*, 45–52.
- Noble, C. E. (1952). An analysis of meaning. *Psychological Review*, *59*, 421–430.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, *22*, 93–136.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: The University of Illinois Press.
- Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review*, *89*, 573–594.
- Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, *53*, 13–24.

Paivio, A., & O'Neill, B. J. (1970). Visual recognition thresholds and dimensions of word meaning. *Perception and Psychophysics*, 8, 273–275.

Palermo, D. S., & Jenkins, J. J. (1964). *Word association norms: Grade school through college*. Minneapolis: University of Minnesota Press.

Palmer, E. D. (2003). *Neural mechanisms underlying single word reading: An fMRI investigation*. Unpublished doctoral dissertation, Washington University at St. Louis, St. Louis, MO.

Palmer, E. D., Brown, T. T., Petersen, S. E., & Schlaggar, B. L. (2004). Investigation of the functional neuroanatomy of single word reading and its development. *Scientific Studies of Reading*, 8, 203–223.

Patterson, K. E. (1982). The relation between reading and phonological coding: Further neuropsychological observations. In: A.W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 77–111). San Diego, CA: Academic Press.

Patterson, K. E., & Morton, J. (1985) From orthography to phonology: An attempt at an old interpretation. In: K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds), *Surface dyslexia* (335–359). Hillsdale, NJ: Erlbaum.

Patterson, K. E., Seidenberg, M., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In: R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 131–181). Oxford: Oxford University Press.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language*, 39, 151–174.

Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In: S. Kinoshita, & S. J. Lupker (Eds), *Masked priming: State of the art* (pp. 97–120). Hove, UK: Psychology Press.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1998). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331, 585–589.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, 1, 153–170.

Petersen, S. E., Fox, P. T., Snyder, A. Z., Raichle, M. E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science*, 249, 1041–1044.

Peterson, R. R., & Simpson, G. B. (1989). The effect of backward priming on word recognition in single-word and sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1020–1032.

- Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 139–156.
- Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In: Cognitive Science Society (Eds), *Proceedings of the 17th annual conference of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 765–805.
- Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, 107, 786–823.
- Plaut, D. C., & Gonnerman, L. M. (2000) Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15, 445–485.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Plourde, C. E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, 51, 181–194.
- Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 148–162.
- Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1142–1158.
- Pollatsek, A., Perea, M., & Carreiras, M. (2005). Does conal prime CANAL more than cinal? Masked phonological priming effects in Spanish with the lexical decision task. *Memory & Cognition*, 33, 557–565.
- Posner, M. I. (1986). *Chronometric explorations of mind*. Oxford: Oxford University Press.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In: R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.
- Price, C. (2004). The functional neuroanatomy of reading. In: R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, J. Ashburner, & W. Penny (Eds), *Human brain function* (2nd ed., pp. 547–562). San Diego, CA: Elsevier.
- Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory & Language*, 25, 461–475.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Shankweiler, D. P., Katz, L., Fletcher, J. M., & Gore, J. C. (1996). Cerebral organization of component processes in reading. *Brain, 119*, 1221–1238.

Purcell, D. G., Stanovich, K. E., & Spector, A. (1978). Visual angle and the word superiority effect. *Memory and Cognition, 7*, 3–8.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Review, 129*, 643–673.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory & Language, 31*, 33–53.

Rastle, K., & Coltheart, M. (1999). Lexical and nonlexical phonological priming in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 461–481.

Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory & Language, 42*, 342–364.

Rastle, K., & Coltheart, M. (2006). Is there serial processing in the reading system; and are there local representations? In: S. Andrews (Ed.). *From inkmarks to ideas: Current issues in lexical processing*. Hove, UK: Psychology Press.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11*, 1090–1098.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*, 385–408.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14*, 191–201.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology, 12*, 206–226.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 274–280.

Reynolds, M., & Besner, D. (2005a). Basic processes in reading: A critical review of pseudohomophone effects in naming and a new computational account. *Psychonomic Bulletin and Review, 12*, 622–246.

- Reynolds, M., & Besner, D. (2005b). Contextual control over lexical and sub-lexical routines in reading aloud. *Psychonomic Bulletin and Review*, *12*, 113–118.
- Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, *15*, 1160–1175.
- Rodd, J. M. (2004). The effect of semantic ambiguity on reading aloud: A twist in the tale. *Psychonomic Bulletin & Review*, *11*, 440–445.
- Roget, P. M. (1911). *Roget's thesaurus of English words and phrases*. Available from Project Gutenberg, Illinois Benedictine College, Lisle, IL.
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*, 487–494.
- Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C., & Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language*, *36*, 382–345.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.
- Rumelhart, D. E., & McClelland, J. L. (Eds). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- Rumsey, J. M., Horwitz, B., Donahue, C., Nace, K., Maisog, J. M., & Andreason, P. (1997). Phonologic and orthographic components of word recognition: A PET–rCBF study. *Brain*, *119*, 739–759.
- Sandra, D., & Taft, M. (Eds). (1994). *Morphological structure, lexical representation and lexical access: A special issue of Language and Cognitive Processes*. Hove, UK: Erlbaum.
- Schendel, J. D., & Shaw, P. (1976). A test of the generality of the word-context effect. *Perception and Psychophysics*, *19*, 383–393.
- Schiller, N. O. (2004). The onset effect in word naming. *Journal of Memory and Language*, *50*, 477–490.
- Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition*, *26*, 1270–1281.
- Schlaggar, B. L., Brown, T. T., Lugar, H. M., Visscher, K. M., Miezin, F. M., & Petersen, S. E. (2002). Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science*, *296*, 1476–1479.
- Schomaker, L., & Segers, E. (1999). Finding features used in the human reading of cursive handwriting. *International Journal on Document Analysis and Recognition*, *2*, 13–18.

- Schomaker, L. R. B., & Van Galen, G. P. (1996). Computer modeling of handwriting. In: A. F. J. Dijkstra, & K. J. M. J. De Smedt (Eds), *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 386–420). London: Taylor & Francis.
- Schreuder, R., Flores d'Arcais, G. B., & Glazenberg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research*, *45*, 339–354.
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory & Language*, *27*, 499–520.
- Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 876–900.
- Sears, C. R., Lupker, S. J., & Hino, Y. (1999a). Orthographic neighborhood effects in perceptual identification and semantic categorization tasks. *Perception & Psychophysics*, *61*, 1537–1554.
- Sears, C. R., Hino, Y., & Lupker, S. J. (1999b). Orthographic neighborhood effects in parallel distributed processing models. *Canadian Journal Experimental Psychology*, *53*, 220–229.
- Seidenberg, M., & Plaut, D. C. (1998). Evaluating word reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*, 234–237.
- Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1–30.
- Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 245–263). Hillsdale, NJ: Erlbaum.
- Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*, 238–242.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.
- Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review*, *97*, 447–452.
- Seidenberg, M., and Plaut, D. C. (1998). Evaluating word reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*, 234–237.
- Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1177–1196.
- Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. (1984a). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383–404.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984b). Pre- and post-lexical loci of contextual effects on word recognition. *Memory and Cognition*, *12*, 315–328.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In: *Proceedings of a symposium on the mechanization of thought processes* (pp. 511–526). London: H. M. Stationery Office.

Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American*, *203*, 60–68.

Selkirk, E. O. (1980). *On prosodic structure and its relation to syntactic structure*. Bloomington Indiana: Indiana University Linguistics Club.

Selkirk, E. O. (1982). The syllable. In: H. van der Hulst, & N. Smith (Eds), *The structure of phonological representations* (Part II pp. 337–383). Dordrecht: Foris.

Sereno, J. A. (1991). Graphemic, associative, and syntactic priming effects at a brief stimulus onset asynchrony in lexical decision and naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 459–477.

Sereno, S. C., & Rayner, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 173–184.

Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In: M. Coltheart, K. Patterson, & J. C. Marshall (Eds), *Deep dyslexia*, (pp. 119–145), London: Routledge & Kegan Paul.

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *The Quarterly Journal of Experimental Psychology*, *35A*, 111–138.

Shulman, G. L. (1990). Relating attention to visual mechanisms. *Perception & Psychophysics*, *47*, 199–203.

Simos, P. G., Breier, J. L., Fletcher, J. M., Foorman, B. R., Castillo, E. M., & Papanicolaou, A. C. (2002). Brain mechanisms for reading words and nonwords: An integrated approach. *Cerebral Cortex*, *12*, 297–305.

Simpson, G. B., & Kang, H. (1994). The flexible use of phonological information in word recognition in Korean. *Journal of Memory & Language*, *33*, 319–331.

Ślowiaczek, M. L., & Clifton, C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior*, *19*, 573–582.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214–241.

Smith, M. C. (1979). Contextual facilitation in a letter search task depends on how the prime is processed. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 239–251.



- Smith, M. C., Theodor, L., & Franklin, P. E. (1983). On the relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 697–712.
- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 8, 411–416.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in normal older adults and individuals with senile dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 461–479.
- Spoehr, K. T., & Smith, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology*, 5, 71–89.
- Spoehr, K. T., & Smith, E. E. (1975). The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 21–34.
- Stanners, R. F., Neiser, J. J., & Painton, S. (1979a). Memory representation for prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 18, 733–743.
- Stanners, R. F., Neiser, J. J., Herson, W. P., & Hall, R. (1979b). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, 18, 399–412.
- Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, 112, 1–36.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 34–78.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64, 1358–1368.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41–78.
- Stolz, J. A., Besner, D., & Carr, T. H. (2005). Semantic priming is robust but unreliable: Inherent noise and, strategic control in the activation and evaluation of semantic knowledge. *Visual Cognition*, 12, 284–336.
- Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, 36, 337–359.
- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1140–1154.
- Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling-sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 207–214.

- Taft, M. (1979a). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7, 263–272.
- Taft, M. (1979b). Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, 18, 21–39.
- Taft, M. (1985). The decoding of words in lexical access: A review of the morphographic approach. In: D. Besner, T. G. Waller, & G. E. MacKinnon (Eds), *Reading research: Advances in theory and practice* (Vol. 5, pp. 83–123). Orlando, FL: Academic Press.
- Taft, M. (1987). Morphographic processing: The BOSS Re-emerges. In: M. Coltheart (Ed.) *Attention and performance XII: The psychology of reading* (pp. 265–279). Hillsdale, NJ: Erlbaum.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638–647.
- Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15, 607–620.
- Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, 22, 259–282.
- Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing: An Interdisciplinary Journal*, 10, 165–200.
- Tan, L. H., & Perfetti, C. A. (1999). Phonological activation in visual identification of Chinese two-character words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 382–393.
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Press, Columbia University.
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- Treiman, R. (1989). The internal structure of the syllable. In: G. Carlson, & M. Tanenhaus (Eds), *Linguistic structure in language processing* (pp. 27–52). Dordrecht, the Netherlands: Kluwer.
- Treiman, R., & Chafetz, J. (1987). Are there onset and rime-like units in printed words? In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 281–298). Hove, UK: Erlbaum.
- Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, 27, 87–104.
- Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language*, 47, 448–468.
- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, 124, 107–136.

- Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language*, 27, 466–477.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6, 171–178.
- Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron*, 24, 105–110.
- Treisman, A. M. (1986). Features and objects and visual processing. *Scientific American*, 255, 114–125.
- Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, 16, 765–780.
- Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition*, 15, 181–198.
- Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 371–386.
- Venezky, R. L. (1970). *The Structure of English Orthography*. The Hague: Mouton.
- Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50A, 439–456.
- West, R. F., & Stanovich, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory and Cognition*, 14, 1040–1112.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59–85.
- Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review*, 8, 221–243.
- Whittlesea, B. W. A., & Cantwell, A. L. (1987). Enduring influence of the purpose of experiences: Encoding-retrieval interactions in word and pseudoword perception. *Memory and Cognition*, 15, 465–472.
- Wijk, A. (1966). *Rules of pronunciation for the English language*. Oxford: Oxford University Press.
- Wijk, A. (1969). Regularized English. The only practicable solution of the English spelling reform problem. In: W. Haas (Ed.), *Alphabets for English*. Manchester: Manchester University Press.
- Winnick, W. A., & Kressel, K. (1965). Tachistoscopic recognition thresholds, paired-associate learning, and free recall as a function of abstractness-concreteness and word frequency. *Journal of Experimental Psychology*, 70, 163–168.

- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.
- Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (in press). Single versus Dual Process Models of Lexical Decision Performance: Insights from RT Distributional Analysis. *Journal of Experimental Psychology: Human Perception & Performance*.
- Yates, M., Locker, L., & Simpson, G. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin and Review*, *11*, 452–457.
- Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1385–1397.
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.
- Zevin, J. D., & Balota, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 121–135.
- Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*, 1–29.
- Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory and Cognition*, *32*, 31–38.
- Zevin, J. D., & Seidenberg, M. S. (2006) Simulating consistency effects and individual differences in nonword naming. *Journal of Memory and Language*, *54*, 145–160.
- Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *Quarterly Journal of Experimental Psychology*, *53A*, 671–692.
- Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language*, *37*, 533–554.
- Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: resolving neighborhood conflicts in the lexical decision task. *Cognition*, *68*, 53–62.
- Zorzi, M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 847–856.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

This page intentionally left blank

## Chapter 10

### Lexical Processing and Sentence Context Effects

*Robin K. Morris*

Visual word recognition (lexical processing) provides the base for constructing meaning from text, as words are the primary meaning bearing element provided to the reader. Chapter 9 provides an extensive review of how people recognize words presented in isolation or in combination with a single-word partner. The evidence from isolated word recognition paradigms makes it clear that individual words contain a wealth of information of a variety of forms (e.g., visual features, orthography, phonology, morphology) and that people can utilize this information to recognize words rapidly and with relatively little conscious effort in the absence of sentence context. However, that is seldom the situation that we find ourselves in outside the laboratory. Far more often we find ourselves faced with recognizing words in the course of continuous silent reading. This is a task that may have word recognition at its base, but it is also a task in which the primary goal is comprehension of a larger meaningful message and as such is one that involves many additional processes that may not be engaged in act of recognizing a word that stands alone. This chapter addresses the issues related to lexical access during reading. Recognizing that the differential task demands of reading compared to isolated word recognition might affect the relative value of different sources of information (Balota, Paul, & Spieler, 1999), and given that reading places particular emphasis on processing for meaning, the first section examines the influence of lexical properties of words with particular emphasis on the meaning bearing properties of words. Although there is extensive evidence demonstrating that we are quite capable of recognizing words in the absence of context, there is also a large body of evidence demonstrating that context can influence word processing when it is present. The latter portion of the chapter takes up the issue of context effects on lexical access.

#### 1. LEXICAL PROPERTIES

There are many different sources of information that are realized within the printed letter string of an individual word that might influence lexical access. These are factors that have figured prominently in studies of visual word recognition that have measured overt responses to words presented in isolation or in single word (priming) contexts and have

been extensively reviewed in an earlier chapter of this volume (see Balota et al.). Many of these factors have similar effects whether one is faced with recognizing a word standing alone or one is reading for meaning. For example, there are clear effects of visual feature information in traditional word recognition paradigms and in studies of word recognition in the context of sentence reading. There is evidence of early phonological activation in reading from studies in which readers experience difficulty when they encounter a letter string with more than one possible pronunciation, for example, the letter string “wind” is pronounced differently when it refers to a weather condition than when it refers to a rotating action (e.g., Folk & Morris, 1995). Other studies have demonstrated that initial processing time on a word is affected by the existence of an unseen phonological partner, such as the pair “sale” and “sail” (Folk, 1999; Pollatsek, Lesch, Morris, & Rayner, 1996; Rayner, Pollatsek, & Binder, 1998), and there is evidence that these early phonological effects extend to scripts like Chinese that do not always make the script to sound relation transparent (see Pollatsek, Rayner, & Lee, 2000, for a review). However, the task demands of reading highlight the search for word meaning and the need to integrate that meaning with information gleaned from the text up to that point. And so, it is there that this review of lexical access in reading will begin. We will look at four aspects of word meaning and their respective roles in recognizing words in the course of reading for comprehension: morphology, word familiarity, word class, and lexical ambiguity.

### **1.1. Morphology**

Morphemes are the smallest meaning bearing units of a word. Most long words in English are composed of more than one morpheme. One of the central questions in psycholinguistic research on morphology concerns the way in which this information is represented in the lexicon. That is, does each complex word have its own unique lexical entry (e.g., Fowler, Napps, & Feldman, 1985) or are they represented as a root with links to possible affixes (e.g., Taft & Forster, 1976). A second, related set of questions asks about how and when morphological units are identified and what role if any do they play in lexical access.

Are morphemes active processing units in the recognition of morphologically complex words when people are engaged in continuous silent reading of connected text? Reading studies have exposed effects of morphological information in gaze duration on a word presented in sentence context. For example, Lima (1987) and Inhoff (1989a, 1989b) found differences in initial processing time between affixed and pseudoaffixed English words (e.g., *relive* and *relish*) and between compound and pseudocompound words (e.g., *cowboy* and *carpet*), respectively. In addition, there is evidence that initial fixation time on morphologically complex English words is influenced by the frequency of the morphological constituents that make up the word in addition to the frequency of the whole word form (Andrews, Miller, & Rayner, 2004; Niswander, Pollatsek, & Rayner, 2000) and that constituent frequency effects can be observed in cases in which the word form frequency is controlled (Juhász, Starr, & Inhoff, 2003). Evidence of constituent frequency effects have also been documented in Finnish reading (Pollatsek, Hyona, & Bertram, 2000; Bertram & Hyona, 2003). These results suggest that morphological constituents are activated in the course of retrieving lexical representations.

So, how is a word decomposed into its morphological constituents prior to the word being recognized? There has been a recent spate of research reports (primarily masked priming studies) suggesting that morphological decomposition in English may be carried out in very early stages of word processing on the basis of orthographic information. Finnish readers spend less initial processing time on long compound words when vowel quality differs across the constituents that make up the word than when the vowel quality is consistent, suggesting that Finnish readers use vowel quality as a morphological segmentation cue (Bertram, Pollatsek, & Hyona, 2004).

Parafoveal preview manipulations have been used to ask questions about the time-course of morphological influence and the results have varied depending on the properties of the languages being investigated. To date the studies that have attempted to demonstrate preview benefit for morphological units during reading in English have failed (Lima, 1978; Inhoff, 1989a, 1989b; Kambe, 2004). In contrast, morphological preview benefits have been observed in Hebrew (e.g., Deutsch, Frost, Pelleg, Pollatsek, & Rayner, 2003; Deutsch, Frost, Pollatsek, & Rayner, 2005). These differences in the pattern of processing observed in English and Hebrew have been attributed to the morphological richness of Hebrew in comparison to English. However, it appears that the story may not be that simple. Recent preview experiments conducted in Finnish, another morphologically rich language, have failed to find evidence of a morphological preview benefit (Bertram & Hyona, in press).

## 1.2. Word Familiarity

Word frequency effects have been demonstrated in virtually every standard measure of word recognition, including naming, lexical decision, tachistoscopic report, semantic categorization, same–different judgments, initial reading as measured by fixation duration, and measures of the early time course of brain activity as reflected in the early components of event-related potentials (ERPs). The ubiquity of frequency effects has been taken as evidence that word frequency is a basic dimension of lexical processing, with more frequent words processed quickly compared with less frequent words.

Lexical familiarity, as assessed by printed word frequency, age of acquisition (AoA), or subjective familiarity rating influences a reader's initial processing time on a word, as measured by first fixation duration or gaze duration. Readers spend more time on less familiar than on more familiar words of equal length (e.g., Inhoff & Rayner, 1986; Juhasz & Rayner, 2003; Rayner & Duffy, 1986; Schmauder, Morris, & Poynor, 2000; Williams & Morris, 2004). These differences are observed even when word length, number of syllables, and word initial bigram and trigram frequency are controlled. High-frequency words are also more likely to be skipped than low-frequency words (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998). In addition to these initial processing time differences, the duration of the first fixation after the low-frequency word is often inflated compared to the high-frequency case. This is thought to reflect the processing of the low-frequency word spilling over onto the next fixation (Rayner, Sereno, Morris, Schmauder, & Clifton, 1990).



Gernsbacher (1984) suggested that printed word frequency may not be the best index of a reader's familiarity and experience with a word, given that people are exposed to words through spoken language as well as print. Juhasz and Rayner (2003) and Williams and Morris (2004) have demonstrated that other measures of word familiarity derived from subjective familiarity ratings, and AoA norms may represent unique (but overlapping) sources of variance to word reading time. Subjective familiarity is thought to be an index of frequency of exposure that is somewhat less biased to print exposure. Williams and Morris (2004) reported familiarity effects above and beyond what could be accounted for by printed word frequency norms. These effects were similar to the effects that have been observed in naming. Juhasz and Rayner (2003) reported unique contributions of AoA in the eye movement patterns of skilled readers and suggested that this measure may reflect differences in the quality of the semantic representation to a larger extent than the other two measures (see Juhasz, 2005, for more extensive treatment of these issues).

### 1.3. Word Class

One way of investigating the role of word meaning in lexical access is to compare reading behavior on words that differ in the degree to which they convey semantic content. For example, we can consider the extent to which linguistic distinctions between content and function words influence processing of those words in text. Content words denote entities, actions, and properties. They are derivational, have compositional meaning, and participate in productive compounding. In contrast, function words are defined by the grammatical relations or syntactic functions in which they participate. They have little if any lexical-semantic content do not participate in productive compounding or contribute to meaning in a compositional way. Because of these linguistic differences, some scholars have suggested that these two classes of words may be accessed differently.

A number of studies have demonstrated that when searching for a target letter in text passages, participants tend to show more detection errors (failure to notice the presence of the target letter) for a letter presented in a function word than for the same letter occurring in a content word. While the results are quite consistent, the theoretical interpretations are widely varied. Some theoretical accounts of this effect have emphasized differences at the level of lexical representation and process between the two word types (e.g., Healy, 1994), while others have emphasized differences beyond the level of lexical processing (e.g., Koriat & Greenberg, 1994). More recently, Greenberg, Healy, and Koriat (2004) have integrated their seemingly divergent views into a single model (Greenberg, Healy, & Koriat, 2004), incorporating both lexical level and text integration-level accounts of the missing letter effect.

Function and content word-processing differences have also been documented in more naturalistic reading tasks. Haberlandt and Graesser (1989) reported processing time differences between the two word classes in a self-paced reading task in which participants were simply told to read for comprehension. The later a word occurred in a sentence, the longer the processing time on that word, and the increase was greater for content words

than for function words. This observation is consistent with the notion that there are observable processing differences between function and content words that occur beyond the lexical level.

Eye movement measures have revealed that readers are twice as likely to skip function words as they are to skip content words when reading technical prose (Carpenter & Just, 1983). However, differences in average word length and word frequency between the two word classes make it hard to know whether this result reflects differences in word class per se. Word skipping increases as word frequency increases (Reichle et al., 1998), even when word length and word class (nouns only) are controlled (Rayner, Sereno, & Raney, 1996). Word skipping is also known to increase as word length decreases (e.g., Rayner & McConkie, 1976) and as predictability increases (Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Rayner & Well, 1996). Function words that are high frequency, short, and predictable are often skipped (Holmes & O'Regan, 1981; O'Regan, 1979; but also see O'Regan 1980, 1992). These findings clearly demonstrated effects very early in lexical processing, but it is not clear whether they should be attributed to differences between the two word classes per se. Schmauder et al. (2000) had participants read sentences that contained a critical function or content word and looked at processing time as well as word skipping measures. In a second experiment words from the two word classes were presented in a primed naming task. Neither the naming task, nor the initial processing time measures revealed evidence of lexical processing differences between function and content words when word length, frequency, and sentence position were controlled. However, there were interactions between word frequency and word class in later processing measures and these were taken to reflect the unique roles these two types of words play in constructing meaning from text.

In summary, the literature contrasting function and content word processing provides a great deal of evidence that differences in the grammatical functions of these two word types translate to processing differences. These differences are most clearly documented in tasks and measures thought to reflect text integration processes. In contrast, there is little evidence that lexical access is different for the two word classes.

#### **1.4. Lexical Ambiguity**

Lexically ambiguous words possess multiple meanings associated with a single orthographic form and, as such, they afford unique opportunities to examine the role of word meaning in lexical access. The questions addressed in the lexical ambiguity resolution literature have historically been seen as central to understanding the nature of the language-processing system more generally. In particular, much of this research has been dedicated to addressing the extent to which language processing is modular (Fodor, 1983) versus interactive (e.g., McClelland, 1987).

Early studies provided evidence that all meanings of an ambiguous word are activated, regardless of the context in which the word occurs (e.g., Frazier & Rayner, 1987; Rayner

& Duffy, 1986; Swinney, 1979; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982). Models based on this result came to be referred to as exhaustive access models and embraced the assumptions of modularity (Fodor, 1983): in particular, the assumption that there is an autonomous lexical processor that recognizes words on the basis of lexical properties alone, without benefit of meaning or context. Under this view all known meanings of an ambiguous word are accessed regardless of context, hence the label exhaustive access models. Other studies using similar methodologies provided evidence that given the appropriate context conditions, participants were faster to respond to a probe related to the context appropriate meaning of an ambiguous word than to a probe related to the context inappropriate meaning (Simpson & Kreuger, 1991; Tabossi, 1988; Tabossi, Colombo, & Job, 1987). The models developed to capture this result have been referred to as selective access models and are consistent with an interactive view of the language-processing system in which information derived from the context may interact with lexical information at the earliest stages of lexical processing. Under this view, given sufficiently constraining context, access may be limited to the context appropriate meaning of the ambiguous word.

Perhaps the most compelling evidence of exhaustive access came from cross-modal priming studies in which participants listened to a sentence or short passage and responded to a printed letter string presented visually. Presentation of the visual probe was calibrated to the auditory occurrence of an ambiguous word. Participants saw one of three possible probes. There was a related probe for each meaning of the ambiguous word and there was an unrelated probe. The time between probe and target could be manipulated to assess the status of word meanings over time. When the probe occurred within 200 ms of the target, participants were faster to respond to either of the meaning-related primes than to an unrelated prime, suggesting that upon hearing the ambiguous word both meanings became available regardless of context (e.g., Onifer & Swinney, 1981; Seidenberg, et al., 1982; Swinney, 1979). As the time between the occurrence of the ambiguous word and the probe was lengthened, facilitation was limited to one meaning-related probe. Taken together, these findings suggested that readers initially access multiple meanings and rapidly select a single interpretation for incorporation into the ongoing text representation.

The timing of access also seems to depend on the relative frequency of the various meanings. In this chapter, we will use the term *meaning dominance* to refer to the extent to which one meaning is more likely to occur than another. The term *balanced words* refers to words with relatively equally likely interpretations, and *biased words* refers to words that have one interpretation that is much more likely than the other(s). Likelihood in these studies is typically operationally defined as the probability that a particular meaning is given as the first associative response to the word presented in isolation.

Meaning dominance effects observed in sentence priming studies demonstrated that for balanced ambiguous words the two meanings are activated close together in time (Seidenberg et al., 1982; Swinney, 1979). Biased words also showed evidence of multiple access (Onifer & Swinney, 1981), but the dominant meaning becomes available prior to the subordinate interpretation (Burgess & Simpson, 1988; Simpson, 1981; Simpson & Burgess, 1985), suggesting that access is frequency ordered.

Eye movement studies have also provided evidence of meaning dominance effects (Dopkins, Morris, & Rayner, 1992; Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rayner, Pacht, & Duffy, 1994; Sereno, Pacht, & Rayner, 1992). In these studies, readers' eye movements were monitored as they read sentences, or short paragraphs and fixation time on the ambiguous word or a control word that was matched in length and frequency was measured. In general the findings from these studies can be summarized as follows. When neutral context precedes the ambiguous word, readers fixate longer on balanced ambiguous words than on biased words or on an unambiguous control word. However, when the reader encounters information that disambiguates to the subordinate interpretation following a biased ambiguous word, they spend more time than when disambiguating information follows a balanced ambiguous word or an unambiguous control word.

Duffy et al. (1988) also manipulated the sentence context preceding the ambiguous word. In this case, the effects of meaning dominance on initial processing time were reversed. Gaze duration on the balanced ambiguous word did not differ from the unambiguous control, suggesting that activation of the context appropriate interpretation exceeded that of the alternative interpretation, thereby reducing the competition between the two interpretations. In the case of the biased ambiguous words, the preceding context supported the subordinate interpretation of the word and readers looked longer at the ambiguous word than at an unambiguous control, suggesting that context had boosted activation of the weaker meaning so that it now competed with the otherwise dominant interpretation. Subsequent studies have yielded similar patterns of data with effects of similar magnitude (e.g., Dopkins et al., 1992; Folk & Morris, 1995; Rayner et al., 1994). The latter effect (longer processing time on a biased ambiguous word in context that supports the less likely interpretation) has been termed the subordinate bias effect or SBE (Rayner et al., 1994) and will be referred to as such throughout this chapter.

Casting questions of lexical ambiguity resolution as tests of modularity, the concept of exhaustive access was tied to the notion of an access process that was impervious to contextual influences, while the selective access accounts were decidedly context dependent. As this brief review has illustrated, there is now a large body of evidence demonstrating that readers access multiple meanings, and that access is influenced by the relative strength of the respective meanings of the word and the context in which it occurs. The picture that has emerged calls for hybrid models that account for rapid activation of multiple meanings and the early influence of sentence context.

There are a number of models of this sort in the literature (e.g., Duffy et al., 1988; Kawamoto, 1993; Simpson & Burgess, 1985; Twilley & Dixon, 2000; Kawamoto & Zemblidge, 1992). As an example, Duffy et al. (1988) proposed the re-ordered access model of lexical ambiguity resolution. This model was later instantiated in a connectionist architecture (Duffy, Kambe, & Rayner, 2001). The model makes four basic assumptions: (1) access is exhaustive; (2) meaning dominance influences the relative activation of multiple meanings; (3) access is an interactive process in which disambiguating context preceding the ambiguous word may increase the activation of the context appropriate interpretation; and

(4) the level of lexical activation of the context inappropriate meaning of the ambiguous word is unaffected by this process. While the re-ordered access model is representative of contemporary hybrid models, the various models differ in the extent to which they characterize exhaustive access as mandatory or highly likely and they differ in their view of the fate of the meaning that is not selected for integration into the discourse representation. Finally, although they all acknowledge that context may influence the order in which meanings become available, they say remarkably little about the properties of the sentence context that give rise to these effects. We will now deal with each of these issues in turn.

### 1.5. Selective Access Re-visited

The SBE has served as a test case for the assumption that access is exhaustive. Under the re-ordered access model the SBE has been taken as evidence that both meanings are activated, and that context may increase the activation of the subordinate interpretation to the point that the two meanings compete for selection. Alternative explanations of the SBE have been proposed by proponents of selective access models (e.g., Simpson & Kreuger, 1991; Kellas & Vu, 1999).

Under a strong selective access account the processing underlying the SBE is fundamentally different than that proposed by the re-ordered access model. According to this account, each meaning of an ambiguous word is stored separately in the lexicon and access is interactive. In the presence of supportive context, a single meaning may be activated without activating any other meanings associated with that form. The frequency of occurrence of the intended meaning of the word (not its frequency in relation to an alternative) and the context in which it occurs are the critical factors influencing initial processing time under this view. In contrast, the re-ordered access model assumes that all possible meanings of an ambiguous word share a common lexical entry. Context may influence the relative activation of candidate meanings, but contrary to the selective access view all meanings are accessed when the form is activated. Under this account, printed word form frequency, meaning dominance (the relative frequency of the possible meanings associated with that form), and the context in which the word occurs, are the critical factors, and exhaustive access is unavoidable.

Under the selective view the SBE is a word frequency effect. The subordinate interpretation of a biased word is in essence a low-frequency word, and so, when compared to a control word matched to the form frequency of the letter string (typically a much higher frequency than the subordinate meaning frequency), it takes longer to process. There is no competition between meanings. It is meaning frequency, not meaning dominance (the relative strength of the subordinate meaning with respect to the dominant meaning) that should determine processing time. Sereno et al. (1992) examined the SBE from this perspective. In their experiment readers encountered a biased ambiguous word in a sentence context that supported the subordinate interpretation. The frequency of the subordinate interpretation was estimated as the proportion of form frequency equivalent

to the meaning bias. That is, an interpretation with a meaning with a bias of .15 was estimated to have a meaning frequency that was 15% of the form frequency of the ambiguous letter string. There were two control conditions in this experiment. One control was matched to the form frequency of the ambiguous word and the other was matched to the frequency of the context appropriate meaning. If access is selective, processing of a biased ambiguous word in a context that supports the subordinate interpretation and processing of an unambiguous word with a printed word frequency equivalent to the subordinate meaning frequency should not differ.

Sereno et al. (1992) obtained the typical SBE when the form frequency control condition was used as the comparison condition. Differences were also observed in three of the four processing measures that were reported when the meaning frequency control was used as the comparison condition. These results indicate that readers were not treating the subordinate interpretation of a biased ambiguous word like an unambiguous low-frequency word. The spillover and total time measures provide evidence of processing difficulty beyond that observed in the low-frequency unambiguous case, presumably due to the co-activation of the alternative interpretation of the ambiguous word (see also, Morris, *in press*; Sereno, O'Donnell, & Rayner, 2005). The fact that there are effects of meaning dominance above and beyond what is predicted by word frequency differences alone adds to the growing body of evidence that there is competition between simultaneously activated meanings.

Given the evidence that multiple meanings are accessed and compete for selection and the evidence that context can affect the status of that competition, one might ask about the limits of that contextual influence. Martin, Vu, Kellas, & Metcalf (1999) proposed that Duffy et al. (1988) and the numerous other eye movement studies that have demonstrated the SBE have lacked sufficient context strength to selectively access the subordinate meaning of an ambiguous word. Under this view exhaustive access and selective access are two extreme points on a common continuum, with selective access as an extreme case of contextual re-ordering. When the context is sufficiently constraining, only the context appropriate meaning of the ambiguous word should be accessed and no initial processing time cost should be observed. However, Binder and Rayner (1998) used Martin et al.'s materials (after eliminating some problematic items) and failed to replicate Martin et al.'s results whether they measured processing time in fixation duration or in self-paced reading time. Numerous other attempts to eliminate the SBE from the eye movement record of skilled readers by manipulating characteristics of the context have also failed (e.g., Binder, 2003; Dopkins et al., 1992; Kambe, Rayner, & Duffy, 2001; Morris & Binder, 2002; Rayner et al., 1994). Given this confluence of evidence, Binder and Rayner (1998) concluded that although context exerts an influence on the order and relative strength with which the possible meanings of an ambiguous word are accessed, even a strongly biasing subordinate context does not preclude activation of all candidate interpretations of a common word form. The evidence from eye movement monitoring studies is quite consistent on this point. But it is important to note that the data from sentence priming studies presents a somewhat different picture, with some studies showing evidence of selective access and others with seemingly similar

manipulations not showing evidence consistent with selective access accounts and there appears to be no obvious way of categorizing these differences methodologically (Simpson, 1994; cf. Tabossi & Sbisà, 2001).

### 1.6. Fate of the Unselected Meaning

The research of the last 30 years has established that, with few exceptions, multiple meanings of an ambiguous word are activated at access, and a single meaning is rapidly selected. Within less than half a second the context inappropriate meanings of the word no longer show signs of activation. Models of ambiguity resolution have dealt with the change in the state of the unselected meaning in different ways. The recorded access model assumes that lexical activations of unselected meanings passively decay. In contrast, the activation-suppression model (Neill & Valdes, 1996) assumes that the unselected meaning is actively suppressed. If there is an active suppression mechanism at work, is it triggered by selection of the context appropriate interpretation (e.g., Binder & Morris, 1995; Gernsbacher, 1990; Gernsbacher & Faust, 1995; Gernsbacher, Robertson, & Werner, 2001; Morris & Binder, 2001; Simpson & Kang, 1994; Simpson & Adamopoulos, 2001), or do readers re-instate a prior episode and retrospectively inhibit the previously rejected interpretation when the situation calls for it (e.g., Neill, 1989; Neill & Valdes, 1996)? This is an area of active debate where many questions are yet to be answered, in fact there is an entire book dedicated to these issues (Gorfein, 2001).

## 2. SENTENCE CONTEXT EFFECTS

Although there is still disagreement about the extent to which selective access is possible, and there is ongoing debate regarding the consequences of meaning selection on the status of the unselected meaning, there is agreement that context influences the status of the candidate meanings as they become available to the reader. There is also substantial evidence from research on the processing of unambiguous words indicating that readers are sensitive to contextual information. In eye movement studies these effects frequently emerge in first fixation and gaze duration, and this is consistent with the evidence from the lexical ambiguity resolution literature suggesting that context is influencing lexical access. One of the most consistent findings is that responses to words are faster when the word is preceded by a congruent context than when it is preceded by a neutral or incongruent context. For example, the word “treasure” is recognized more quickly in the sentence “The pirate found the treasure,” than in the sentence “The person liked the treasure,” or worse yet, “The house was destroyed by the treasure” (e.g., Balota et al., 1985; Ehrlich & Rayner, 1981; Fischler & Bloom, 1979, 1980; Foss, 1982; Hess, Foss, & Carroll, 1995; Schuberth, Spoehr, & Lane, 1981; Simpson, Casteel, & Burgess, 1989). The fact that context has an influence on word processing is clear, however, we are still far from reaching consensus on the processing mechanisms and/or the contextual factors that underlie these effects. The following section reviews the relevant evidence from studies examining the processing of ambiguous and unambiguous words encountered in the course of reading for comprehension.

## 2.1. Predictability

Some of the earliest studies of sentence context effects on word processing looked at the effect of predictability (Tulving & Gold, 1963; Tulving, Mandler, & Bauml, 1964; Morton, 1964). To predict is to declare or indicate in advance. In this section, the term “predictability” is used to refer to the extent readers might anticipate the identity of upcoming words based on the context in which they occur. This factor is typically operationalized either by measures of cloze probability in which participants are presented with a sentence fragment and asked to complete the sentence with the first word that comes to mind, or by rating tasks in which readers are asked to rate the likelihood that a sentence fragment would be continued with a particular word.

There are numerous studies showing that words that are predictable from the context in which they occur are processed more rapidly than words that are not predictable (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Binder, Pollatsek, & Rayner, 1999; Ehrlich & Rayner, 1981; Frisson, Rayner, & Pickering, 2005; Inhoff, 1984; Lavigne, Vitu, & d’Ydewalle, 2000; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner, & Well, 1996; Schilling, Rayner, & Chumbley, 1998; White, Rayner, & Liversedge, 2005; Zola, 1984). In addition, words are more likely to be skipped in a predictive context than in a neutral context (Altarriba et al., 1996; Balota et al., 1986; Dreighe, Brysbaert, Desment, & DeBaeke, 2004; Ehrlich & Rayner, 1981; Rayner et al., 2004).

On the basis of the findings summarized in the previous two paragraphs one might conclude that readers are anticipating the upcoming word based on context and retrieving that lexical item prior to processing the perceptual input. However, we know that perceptual processing can be accomplished very rapidly. Reading is unimpaired when words are masked 100 ms after the start of each fixation and only slightly impaired when the exposure duration is reduced to 50 ms (e.g., Rayner, & Slowiaczek, 1981; Slowiaczek, & Rayner, 1987). These might be taken as estimates of the minimum amount of time required to extract the visual information from the page and mentally represent it in some sensory form. But even if we look at the typical fixation duration in reading standard text, readers spend just over a quarter of a second on any given word. In addition, we know that sentence contexts in natural language use are seldom sufficiently constraining to allow readers to accurately anticipate the next word in the discourse (Gough, Alford, & Holley-Wilcox, 1981) and assuming that predicting is a conscious strategy that readers apply to the process, erroneous predictions should be costly. Stanovich and West (1981, 1983) provided evidence that words with high cloze probabilities were named faster following a related sentence context than following a neutral or incongruent context. They reasoned that if readers were using the context to predict the next word prior to processing the perceptual input, then they should be slower to name the target in an incongruent context than a neutral context because they would have to reject their predicted completion before they could accurately name the target. Contrary to this hypothesis, they found no sign of inhibition when the word was named in an incongruent context.



Taken together, these results suggest that although words that are highly likely to occur in a particular context enjoy a processing benefit, these effects cannot be fully accounted for by an anticipatory mechanism that predicts the upcoming word without benefit of perceptual input. Further evidence of this comes from studies in which words that are preceded by semantically related context are processed faster than words that are preceded by unrelated context, even when the word is not the predicted completion generated in a sentence completion task (Duffy, Henderson, & Morris, 1989; Morris, 1994; Morris & Folk, 1998). These results point to the need to postulate other mechanisms to account for sentence context effects.

## 2.2. Intralexical Priming

Lexical-level explanations of sentence context effects on word processing propose that contextual facilitation arises from word-to-word associations, or intralexical priming. According to this account, context effects are the result of activation spreading from related words in a context to the target word, speeding access to that word (e.g., Duffy et al., 1989; Fodor, 1983; Seidenberg et al., 1982), in much the same way that semantic-relatedness effects occur in word lists (e.g., Meyer & Schvaneveldt, 1976), or in the semantic priming experiments using single-word contexts that are reviewed in the visual word recognition chapter of this volume.

There is some evidence that when words that are semantically related occur very close to one another, the presence of the first word may prime the second. Sereno and Rayner (1992) demonstrated intralexical priming effects in sentence context using the fast-priming paradigm. This task paradigm combines properties of reading for comprehension with properties of traditional semantic priming paradigms minus the need for participants to make an overt response. Readers' eye movements are monitored as they read silently for comprehension. As the reader's eyes approach the target word, a random letter string occupies the target location. When the reader's eye lands on the target location, the letter string is replaced with a prime word for a brief period of time and then replaced with the target word. Sereno and Rayner found that readers spent less time on a target word presented in neutral sentence context when it was preceded by a semantically related prime presented for 30 ms than when it was preceded by an unrelated prime. There is also evidence of facilitation in gaze duration on a word when it is preceded by a close semantic associate (e.g., "king" preceded by "queen"), but only when the two words appear within a single clause (Carroll & Slowiaczek, 1986). There was no evidence of priming when a clause boundary was imposed between the two critical words. To the extent that there are intralexical priming effects in reading, they appear to be short lived.

Rayner et al. (1994) attempted to use intralexical priming to boost the activation of the subordinate interpretation of a biased ambiguous word so that it was accessed without competition from the dominant meaning. In one experiment the readers' first encounter with the subordinate sense of the ambiguous word was obtained through a paired associate task prior to a sentence reading task in which their eye movements were monitored.

In a second experiment, both exposures to the word occurred within the context of a short passage. Gaze duration on the second encounter of the biased ambiguous word or an unambiguous control word (matched for length and frequency) was measured in each of these experiments. There was no evidence that priming by a lexical associate or explicit repetition of the biased ambiguous word in context that supported the subordinate interpretation at both encounters was sufficient to mediate the SBE (see also Morris & Binder, 2001).

While context effects on word processing in reading may arise in part from intralexical priming, the domain of influence of intralexical priming effects in continuous reading appears to be quite restricted. In addition, there is evidence that words are processed faster in related context than in an unrelated context, in the absence of lexical associates (e.g., Foss & Ross, 1983; Sharkey & Mitchell, 1985; Sharkey & Sharkey, 1992), and there is evidence that lexical relatedness alone is not always sufficient to produce this processing advantage (Duffy et al., 1989; Hess et al., 1995; Masson, 1986; Morris, 1994; O'Seaghdha, 1989; Potter, Moryadas, Abrams, & Noel, 1993; Simpson et al., 1989; Williams, 1988).

### 2.3. Interactive Sentence Context Effects

Conscious prediction mechanisms and automatic spreading activation between lexical-semantic associates are insufficient to account for the full range of contextual facilitation that has been observed. According to interactive accounts of contextual facilitation, emergent properties of the discourse representation may influence the processing of individual words during reading. In order to construct productive models of these effects, we need to take inventory of the properties of the discourse representation that might play a role, and those that do not.

Evidence that lexical relatedness alone is not sufficient to produce a word processing advantage exists in several different forms. Simpson et al. (1989) showed that sets of words that produced facilitation when embedded in a sentence context failed to produce the same effects when the same words were presented in a scrambled order. Conversely, Williams (1988) showed that a set of words that produced facilitation in the absence of a sentence frame did not yield facilitation when they were embedded in sentences.

Morris (1994) demonstrated that the time to read a word varied as a function of the sentence context, even when the words that made up that context did not change across conditions. In those experiments, participants read sentences like the following:

1. The waiter watched as the accountant balanced the ledger the second time.
2. The waiter who watched the accountant balanced the ledger the second time.

Participants' eye movements were monitored, and the primary dependent measure was gaze duration on the target word (ledger). The results of a sentence completion task were

used to select target words that were produced as completions less than 15% of the time. Each sentence contained a verb, which was semantically related to the target (“balanced” in this example), and two nouns, each of which were related to the verb (waiter and accountant). One of the nouns, in conjunction with the verb, was highly related to the target word (accountant + balanced). The other noun, in conjunction with the verb, was related to a very different scenario (waiter + balanced). That is, when accountants balance, the sentence is about bookkeeping, and when waiters balance, the sentence is about restaurants. Control conditions were created by replacing the critical content words with neutral words (e.g., person, woman, saw, etc.). Readers spent less time on the target word (ledger) only when the accountant was balancing, and not when the sentence contained the same words, but the waiter was balancing. This effect was observed in both first fixation and gaze duration data. These results clearly implicate information beyond the lexical level in the word recognition process. However, there was also some evidence for intralexical priming, in this experiment. Processing time on the matrix verb was examined in some of the control conditions. In one set of controls, accountant or waiter appeared in conjunction with a neutral agent (e.g., man, woman, person) and in another set of controls, both potential subjects were represented by similarly neutral terms. Processing time on the verb was shorter following one semantically related noun and one neutral noun than following two neutral nouns. This processing time advantage was observed regardless of whether the semantically related noun was the agent of the action denoted by the verb or not, suggesting that it was the result of semantic association between the noun and verb, as opposed to properties of the discourse representation. This raises the possibility that discourse-level and intralexical-level accounts need not be mutually exclusive.

The experiments discussed thus far clearly demonstrate the need for an interactive account of sentence context effects and set some bounds on what does and does not interact. But they do not tell us much about how such interactive effects might occur. Schwanenflugel and LaCount (1988) defined contextual constraint in terms of semantic feature activation, and Tabossi has made a similar proposal regarding context effects on lexical ambiguity resolution. Schwanenflugel and LaCount proposed that high-constraint sentences impose greater feature restrictions on the spread of activation through the lexicon. Lexical access is facilitated to the extent that its semantic features are shared with the semantic features highlighted by the sentence context. If the sentence activates very general semantic features or a large number of specific features, then many words will be activated. This view makes the interesting prediction that words that are closely related to the predicted target word will only be activated to the extent that they also share semantic features with the context. So, for example, the word “shower,” processing of the word “shower” would not be facilitated in the sentence “The tired mother gave her dirty child a ...,” even though it is closely related to the predicted completion, “bath” and it is a plausible completion for this sentence frame. Schwanenflugel and colleagues reported evidence of this in a series of cross-modal priming studies in which readers listened to a sentence and then made a lexical decision response to a visually presented probe. However, the responses in these experiments occurred more than a second after the offset of the sentence leaving open the possibility that multiple candidates were activated by the

sentence context, and the best fit to the context was selected prior to making a response. Traxler and Foss (2000) showed that when an earlier processing measure is used, facilitation is observed for both predicted unambiguous sentence completions and for semantically related targets, contradicting the predictions of the feature activation account. Interestingly, although they found evidence that multiple lexical candidates were facilitated by the sentence context, they found no evidence of competition among the candidate completions.

The lack of evidence for competition in the Traxler and Foss studies is particularly interesting, given that the lexical ambiguity resolution literature provides numerous demonstrations of competition when multiple candidate interpretations are activated within a similar timeframe. The critical distinction between these two cases may be that in the unambiguous word experiments each candidate has a unique lexical entry (word form), while in the case of the ambiguous word there are two or more distinct semantic interpretations activated via a single word form. Perhaps the competition is specific to multiple candidates co-activated from a common form.

Dopkins et al. (1992) took a slightly different approach to the question of the role of contextual constraint. They manipulated the type of disambiguating context that occurred prior to a biased ambiguous word. In one condition, the context supported the subordinate interpretation (positive evidence), as in “Having been heavily praised by the drinkers, the port was soon guzzled to the last drop.” In this case, the context prior to the ambiguous word “port” supports the subordinate beverage meaning of port through its association with drinking. This is representative of the disambiguating context that occurs most often in other published reports on ambiguity resolution. In another condition, the context biased toward the subordinate interpretation by ruling out the dominant interpretation (negative evidence) as in “Having been carried for miles by mule-train, the port was soon guzzled to the last drop.” In this case, the dominant “harbor” interpretation is ruled out since harbors cannot be carried by mule-train, leaving the subordinate “beverage” interpretation of port as the only viable option. Initial processing time (as measured by gaze duration) on the target word revealed that both negative and positive evidence were effective in boosting the activation of the subordinate interpretation to compete for selection. However, context that ruled out the dominant interpretation was no more effective than context that supported the subordinate interpretation.

## 2.4. Properties of the Discourse Representation

The majority of the work reviewed thus far has looked at contextual influences on lexical access occurring within a single sentence. We now consider some properties of the discourse representation that have been shown to influence comprehension and memory for text and ask to what extent these properties influence lexical access.

It has been suggested that entities related to the discourse topic are maintained in an active state and likewise that a shift in the topic of a discourse renders information related

to the original topic less available for processing (e.g., Clifton & Ferreira, 1987; Gernsbacher, 1990; O'Brien, Duffy, & Myers, 1986). Binder and Morris (1995) examined the ability of discourse topic to influence lexical access by looking at the influence of topic on lexical ambiguity resolution. In the first experiment participants read short passages like the following that contained two instances of a balanced ambiguous word.

*Meaning consistent.* There was a lot of excitement at the bars downtown. Crowds of people were gathered outside the club (home) on the street. It appeared that someone had been hurt in the **club** that night. The police had been called.

*Meaning switched.* There was a lot of excitement at the bars downtown. Crowds of people were gathered outside the club (home) on the street. An hour earlier, a man was struck on the head with a **club** and robbed. The police had been called.

The context appropriate interpretation either remained consistent from first to second encounter or it was switched via information conveyed in the intervening context. Control conditions were created by replacing the first instance of the ambiguous word with an unambiguous control word that was matched for word length and word frequency with the ambiguous word. Initial processing time on the second occurrence of the balanced ambiguous word (henceforth the target word, in bold in the example) and on the context immediately following the ambiguity (the post-target region, underlined in the example) were measured. Looking at initial processing time on the target there was evidence that readers benefited from the repetition of the ambiguous word when the meaning remained consistent across encounters and there was no cost associated with the conditions in which the meaning switched.

Processing time in the post-target region showed the opposite pattern of results. There was a cost in initial processing time in the post-target region when the meaning switched and no benefit was observed in that region when the meaning was consistent with the first encounter. In another experiment, the meaning of the ambiguous word always switched from first to second encounter. But now, the intervening context either maintained the original topic of the passage or the topic shifted. A topic shift was operationalized as a change in the focal actor and a change in location. If the processing cost observed in the post-target region of the previous experiment is due to difficulty integrating the selected meaning into a discourse representation that includes the other meaning of that word, then that cost should diminish with a shift in discourse topic, as the previous concept is no longer in a highly active state in the discourse representation. The same logic applies to the initial processing time on the target word. If the benefit obtained in the meaning consistent condition emanates from the discourse representation, then it too should diminish with a shift in topic. Interestingly, the topic shift alleviated the processing difficulty in the post-target region in the inconsistent condition, but had no effect on the processing benefit observed on the target word in the consistent condition.

The initial processing time on the target word in Binder and Morris (1995) may not have been affected by the topic manipulation because accessing meaning is not affected

by such global discourse factors or it could be because this particular topic manipulation was not effective. Several recent investigations of the role of linguistic focus in word processing in reading have provided results consistent with the first explanation. The focus of a sentence is said to be the most prominent or emphasized constituent in that sentence (Halliday, 1967). The focus of a sentence may be indicated through the use of *wh*-questions (e.g., Birch & Rayner, 1997; Blutner & Sommer, 1988; Cutler & Fodor, 1979), *there*-insertion sentences (e.g., Birch & Garnsey, 1995), or *it*-cleft sentence constructions (e.g., Birch & Garnsey, 1995; Birch & Rayner, 1997). Like topic, focusing is thought to enhance the relative availability of concepts in memory in spoken language comprehension (Gernsbacher & Jescheniak, 1995; Gernsbacher & Shroyer, 1989) and in reading (Birch & Garnsey, 1995; Birch & Rayner, 1997; Carpenter & Just, 1977; McKoon, Ward, Ratcliffe, & Sproat, 1993; Morris & Folk, 1998; Singer, 1976; Ward & Sturt, in press).

Blutner and Sommer (1988) manipulated focus through the use of *wh*-questions. They measured lexical decision time on a previously focused concept and found that focused concepts were responded to faster than non-focused concepts, suggesting that focus facilitates access. However, a recent series of eye movement studies paint a different picture. Birch and Rayner (1997) failed to find initial processing effects for focused items while monitoring readers' eye movements in a silent reading task. In one experiment, Birch and Rayner syntactically directed focus through the use of *it*-cleft constructions, and in another experiment subjects read sentence pairs in which the first sentence was in the form of a *wh*-question that focused a particular entity in sentence two. In both cases they demonstrated differential effects of focus, but not in readers' initial processing time as measured by first fixation and gaze duration (see also, Morris & Folk, 1998; Ward & Sturt, in press). Although Morris and Folk (1998) also found no evidence of facilitation in initial processing time on the focused item, they observed a processing advantage for words that were semantically related to the focused item that occurred later in the clefted sentence, suggesting that the heightened prominence of the focused item in the discourse representation facilitated access of related words. These studies differed from Blutner and Sommer in that participants simply read the sentences as their eye movements were monitored. There was no overt response to the focused item required and the timeframe of processing captured in the first fixation and gaze duration measures of the eye movement record provide an earlier measure of processing than the lexical decision latency. Blutner and Sommer's results are consistent with the later processing advantage observed in the silent reading studies.

Several ambiguity studies have attempted to eliminate the SBE through manipulations of discourse-level variables (e.g., Binder, 2003; Kambe, Rayner, & Duffy, 2001; Morris & Binder, 2002; Wiley & Rayner, 2000). For example, in one experiment reported by Binder (2003), participants read passages that contained a biased ambiguous word. The sentence containing the ambiguous word (local context) supported the subordinate meaning and the discourse topic established in the first sentence of the passage (global context) was consistent, inconsistent, or neutral with the local context. Even when local sentence context and global discourse topic information converged in support of the subordinate interpretation, the SBE was not eliminated (see also Kambe et al., 2001; Morris

& Binder, 2001). All of these studies obtained evidence that the SBE endured even when higher order discourse variables (global context, topic, conceptual repetition) and local sentence context were brought to bear.

Hopefully, this small sample of the work on context effects on lexical access has illustrated the progress that is being made toward understanding the limits of these effects. For example, the results reviewed here would suggest that the semantic content of the sentence context exerts a powerful influence on lexical access relative to the content of the more extended discourse (although we still have much to learn about “the semantic context”). In addition, although there is evidence that the message that emerges from successful syntactic parsing influences lexical processing, there is little evidence to suggest that syntactic relations per se have any direct influence on lexical activation. Finally, there is little evidence that linguistic devices known to increase the salience of concepts that have been retrieved from long-term memory also influence initial access to that information. Teasing apart those aspects of linguistic representation and process that have a direct influence on lexical access from those that do not is critical to developing more specific processing models.

## REFERENCES

- Altarriba, J., Kroll, J.F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, *24*(4), 477–492.
- Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, *16*, 285–311.
- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364–390.
- Balota, D.A., Watson, J.M., Duchek, J.M., & Ferraro, F.R. (1999). Cross-modal semantic and homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals, *Journal of the International Neuropsychological Society*, *5*, 626–640.
- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, *48*, 615–634.
- Bertram, R., Pollatsek, A., Hyönä, J. (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language*, *51*, 325–345.
- Binder, K. S. (2003). Sentential and discourse topic effects on lexical ambiguity processing: An eye-movement examination. *Memory and Cognition*, *31*, 690–702.
- Binder, K.S., & Morris, R.K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1186–1196.

- Binder, K.S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 1162–1172.
- Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations. *Psychonomic Bulletin and Review*, 5, 271–276.
- Burgess, C., & Simpson, G. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language*, 33, 86–103.
- Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In: K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes*, (pp. 275–307). New York: Academic Press.
- Carroll, P., & Slowiaczek, M.L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition*, 14(6), 509–522.
- Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin and Review*, 10, 415–422.
- Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2005). Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Language and Cognitive Processes*, 20, 341–371.
- Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language*, 31, 461–477.
- Duffy, S.A., Henderson, J.M., & Morris, R.K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 791–801.
- Duffy, S. A., Kambe, G., & Rayner, K. (2001). The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements. In: D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, (pp. 27–43). Washington, DC: American Psychological Association.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Ehrlich, S.F., & Rayner, K. (1981). Contextual effects of word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 641–655.
- Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18, 1–20.
- Fischler, I., & Bloom, P. A. (1980). *Memory and Cognition*, 8, 216–225.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.



Folk, J.R. (1999). Phonological codes are used to access the lexicon during silent reading. Folk, Jocelyn R. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 892–906.

Folk, J.R., & Morris, R.K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1412–1429.

Foss, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology*, 14, 590–607.

Foss, D.J., & Ross, J.R. (1983). Great expectations: Context effects during sentence processing. In: G. Flores D'Arcais, & R.J. Jarvella (Eds), *The Process of Language Understanding*, (pp. 169–191). Chichester, England: Wiley.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition*, 13, 241–255.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, 26, 505–526.

Frisson, S., Rayner, K., & Pickering, M.J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 862–877.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, 113, 256–281.

Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Gernsbacher, M.A., & Faust, M. (1995). Skilled suppression. In: F.N. Dempster, & C.J. Brainerd (Eds), *Interference and inhibition in cognition* (pp. 295–327). San Diego, CA: Academic Press, Inc.

Gernsbacher, M.A., Robertson, R.W., & Werner, N.K. (2001). The costs and benefits of meaning. In: D.S. Gorfein, S. David (Eds), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, (pp. 119–137). Washington, DC: American Psychological Association.

Gorfein, D.S. (2001). An activation-selection view of homograph disambiguation: A matter of emphasis? In: D.S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, (pp. 157–173). Washington, DC: American Psychological Association.

Greenberg, S. N., Healy, A. F., & Koriati, A. (2004). The GO model: A reconsideration of the role of structural units in guiding and organizing text on line. *Psychonomic Bulletin & Review*, 11, 428–433.

- Haberlandt, K., & Graesser, A. C. (1989). Buffering new information during reading. *Discourse Processes*, 12, 479–494.
- Healy, A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin and Review*, 1, 333–344.
- Hess, D.J., Foss, D.J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124, 62–82.
- Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, 20, 417–430.
- Inhoff, A. (1989a). Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 544–555.
- Inhoff, A. (1989b). Lexical access during eye fixations in reading: Are word access codes used to integrate lexical information across interword fixations? *Journal of Memory and Language*, 28, 444–461. Both of these Inhoffs possible based on reference in text.
- Inhoff, A.W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning & Verbal Behavior*, 23(5), 612–624.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, 40, 431–439.
- Juhasz, B. (2005). Age-of-Acquisition Effects in Word and Picture Identification Juhasz, Barbara J. *Psychological Bulletin*. 131, 684-712.
- Juhasz, B., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1312–1318.
- Juhasz, B. J., Starr, M. S., & Inhoff, A. W. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94, 223–244.
- Kambe, G. (2004). Parafoveal processing of prefixed words during eye fixations in reading: Evidence against morphological influences on parafoveal preprocessing. *Perception and Psychophysics*, 66, 279–292.
- Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory and Cognition*, 29, 363–372.
- Kawamoto, A. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, 32, 474–516.
- Kawamoto, A.H., & Zemplidge, J.H. (1992). Pronunciation of homographs , *Journal of Memory and Language*. 31, 349–374.

- Kellas, G., & Vu, H. (1999). Strength of context *does* modulate the subordinate bias effect: A reply to Binder and Rayner. *Psychonomic Bulletin and Review*, 6, 511–517.
- Koriat, A., & Greenberg, S. N. (1994). The extraction of phrase structure during reading: Evidence from letter detection errors. *Psychonomic Bulletin and Review*, 1, 345–356.
- Lavigne, F., Vitu, F., & d'Y dewalle, G. (2000). The influence of semantic context on initial eye landing sites in words. *Acta Psychologica*, 104(2), 191–214.
- Lima, S. (1987). Morphological analysis in sentence reading. *Journal of Memory and Language*, 26, 84–99.
- Martin, C., Vu, H., Kellas, G., & Metcalf, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *Quarterly Journal of Experimental Psychology*, 52A, 813–839.
- Masson, M.E. (1986). Comprehension of rapidly presented sentences: The mind is quicker than the eye. *Journal of Memory and Language*, 25(5), 588–604.
- McClelland, J. L. (1987). The case of interactionism in language processing. In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, (pp. 3–36). Hillsdale, NJ: Erlbaum.
- Meyer, D.E., & Schvaneveldt, R.W. (1976), Meaning, memory structure, and mental processes *Science*, 192(4234), 27–33.
- Morris, R.K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20, 92–103.
- Morris, R.K., & Binder, K.S. (2001). What happens to the unselected meaning of an ambiguous word in skilled reading? In: D.S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, (pp. 139–153). Washington, DC: American Psychological Association.
- Morris, R. K., & Binder, K. S. (2002). What happens to the unselected meaning of an ambiguous word in skilled reading? In: D.S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*, (pp. 139–153). Washington, DC: American Psychological Association.
- Morris, R.K., & Folk, J.R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition*, 26, 1313–1322.
- Morton, J. (1964). The effects of context on the visual duration threshold for words. *British Journal of Psychology*, 55, 165–180.
- Neill, W.T., & Valdes, L.A. (1996). Facilitatory and inhibitory aspects of attention. In: A.F. Kramer, M.G.H. Coles, and G.D. Logan, (Eds), *Converging operations in the study of visual selective attention*, (pp. 77–106). Washington, DC: American Psychological Association.
- Niswander, E., Pollatsek, A., & Rayner, K. (2000). The process of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, 15, 389–420.

- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning, and contextual bias. *Memory and Cognition*, *9*, 225–236.
- O'Regan, J. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception and Psychophysics*, *25*, 501–509.
- O'Regan, J. K. (1980). The control of saccade size and fixation duration in reading: The limits of linguistic control. *Perception and Psychophysics*, *28*, 112–117.
- O'Regan, J. K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory. Special Issue: Object perception and scene analysis. *Canadian Journal of Psychology*, *46*, 461–488.
- Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 820–833.
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E.D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191–201.
- Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 779–790.
- Rayner, K., & McConkie, G. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829–837.
- Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language*, *33*, 527–544.
- Rayner, K., Pollatsek, A., & Binder, K. (1998). Phonological codes and eye movements in reading. *Journal of Experimental Psychology*, *24*, 476–497.
- Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1188–1200.
- Rayner, K., & Slowiaczek, M.L. (1981). Expectations and parafoveal information in reading: Comments on McClelland and O'Regan. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 645–651.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review*, *3*, 504–509.

- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.
- Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(4), 1270–1281.
- Schmauder, A. R., Morris, R. K., & Poynor, D. V. (2000). Lexical processing and text integration of function and content words: Evidence from priming and eye fixations. *Memory and Cognition*, *28*, 1098–1108.
- Schuberth, R. E., Spoehr, K. T., & Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory and Cognition*, *9*, 68–77.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*, 489–537.
- Sereno, S.C., O'Donnell, P.J., & Rayner, K. (2006). Eye Movements and Lexical Ambiguity Resolution: Investigating the Subordinate-Bias Effect. *Journal of Experimental Psychology: Human Perception and Performance*. *32*, 335–350.
- Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science*, *3*, 296–300.
- Sereno, S.C., & Rayner, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 173–184.
- Sharkey, A.J., & Sharkey, N.E. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language*, *31*(4), 543–572.
- Sharkey, N.E., & Mitchell, D.C. (1984). Word recognition in a functional context: The use of scripts in reading. *Journal of Memory and Language*, *24*(2), 253–270.
- Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*, 120–136.
- Simpson, G.B. (1994). Context and the processing of ambiguous words. In: M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, (pp. 359–374). San Diego, CA: Academic Press, Inc.
- Simpson, G.B., & Adamopoulos, A.C. (2001). Repeated homographs in word and sentence contexts: Multiple processing of multiple meanings . In: D.S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, (pp. 105–117). Washington, DC: American Psychological Association.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 28–39.

- Simpson, G.B., & Kang, H. (1994). Inhibitory processes in the recognition of homograph meanings. In: D. Dagenbach, & T.H. Carr (Eds), *Inhibitory processes in attention, memory, and language*, pp. 359–381. San Diego, CA: Academic Press, Inc.
- Simpson, G. B., & Kreuger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, 30, 627–643.
- Simpson, G.B., Peterson, R.R., Casteel, M.A., & Burgess, C. (1989). Lexical and sentence context effects in word recognition *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(1), 88–97.
- Slowiaczek, M.L., & Rayner, K. (1987). Sequential masking during eye fixations in reading. *Bulletin of the Psychonomic Society*, 25(3), May 1987, 175–178.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659.
- Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, 49, 161–167.
- Tabossi, P., & Sbisà, S. (2001). Methodological issues in the study of lexical ambiguity resolution. In: D.S. Gorfein (Ed.), (2001). *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, (pp. 11–26). Washington, DC: American Psychological Association.
- Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15, 607–620.
- Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, 66, 319–327.
- Tulving, E., Mandler, G., & Bauml, R. (1964). Interaction of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology*, 18, 62–71.
- Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin and Review*, 7, 49–81.
- White, S.J., Rayner, K., & Liversedge, S.P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination *Psychonomic Bulletin & Review. Special Issue: Memory strength and recency judgments*, 12(5), 891–896.
- Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory and Cognition*, 28, 1011–1021.
- Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and word acquisition. *European Journal of Cognitive Psychology*, 16, 312–339.
- Zola, D. (1984) Redundancy and word perception during reading. *Perception & Psychophysics*, 36(3), 277–284.

This page intentionally left blank

## Chapter 11

### Semantic Memory

*Beth A. Ober and Gregory K. Shenaut*

#### 1. INTRODUCTION

During the decade ending around 1972, Quillian and colleagues introduced a seminal computational model that they called *semantic memory*. This was less than a century after memory itself was introduced to modern psychology by, e.g., Wundt, Ebbinghaus, and James in the late 1800s, and semantics per se was brought into existence as a field of study within linguistics by Bréal. Quillian's semantic memory continues to influence psychology, psycholinguistics, and cognitive neuroscience. The PsycInfo database contains over 2400 publications containing the keyword "semantic memory," from 1966 (when Quillian's dissertation was completed) to 2005. The most important paradigms involving semantic memory today are the "neuro" paradigms, which involve the relationship of semantic memory to brain structure or function. Figure 1 illustrates the transition of semantic memory research from an almost purely normal-literature phenomenon to one increasingly dominated by neuro paradigms. Over half (54%) of all semantic memory publications have been neuro-related; for the past decade, almost 3/4 (72%) have been neuro-related.

This chapter aims to provide key background information on semantic memory theory and research in normal and neuropsychological populations. We begin with a description of the original Quillian model of semantic memory, including its structures and processes. Then, we describe what we have dubbed *generic semantic memory*, the post-Quillian conceptions in wide use today. Next, we give an overview of some theoretical extensions, consequences, and divergences over the past several decades regarding semantic memory. The balance of the chapter focuses on (generic) semantic memory within the neuro literature: (1) five major neuro-related models of semantic memory are critically reviewed, (2) the issue of storage versus access deficits of semantic memory in neuropsychological populations are discussed, (3) empirical work on the use of similarity judgments to assess semantic memory organization in neuropsychological populations, specifically Alzheimer's disease (AD) is presented, and (4) research on the use of semantic priming to assess semantic knowledge in neuropsychological populations



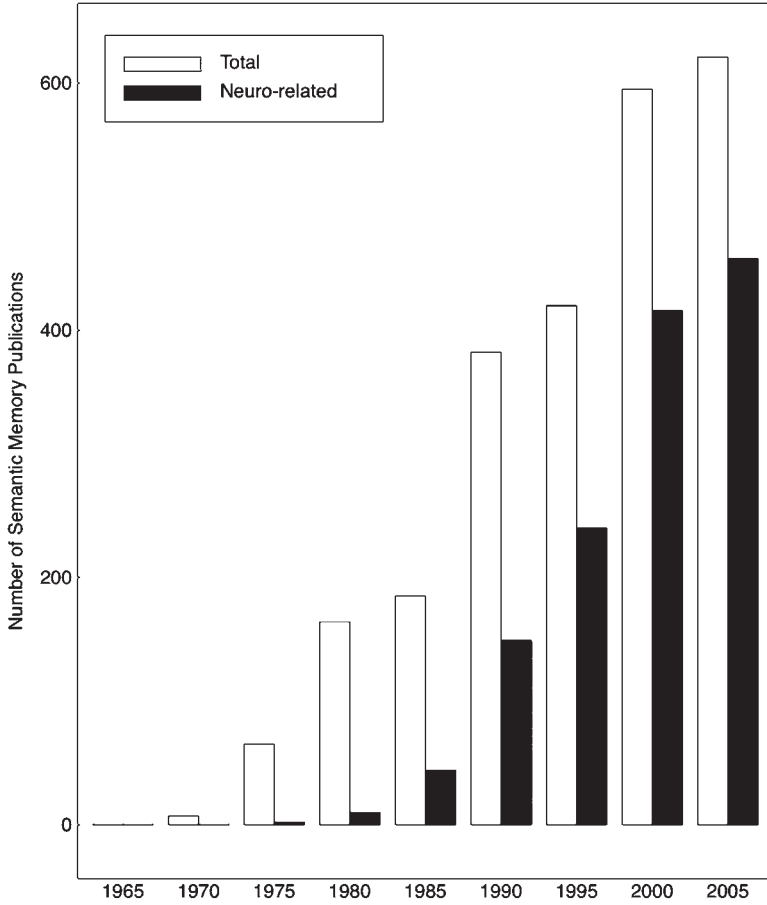


Figure 1. Number of publications in consecutive five-year periods ending in the years displayed on the abscissa, as determined by searches of the CSA/APA PsycInfo database. “Total” publications resulted from the search string “KW=(semantic memory)”; “Neuro-related” publications resulted from the search command “KW=(semantic memory) and KW=(defici\* or brain or neuro\* or impair\* or abnormal\* or fmri or mri or patient\* or dement\* or aphasi\* or evoked potential\* or erp or positron or pet).”

(specifically AD) is discussed. We conclude the chapter by suggesting some intriguing topics for future research on semantic memory, with a focus on further interaction and collaboration between researchers in the behavioral sciences and those in the neuro-related sciences.

## 2. QUILLIAN’S SEMANTIC MEMORY

Quillian’s semantic memory was first a theory of human long-term memory, and second a series of computer simulations of certain types of language processing. This

section contains our interpretative summary of semantic memory theory as reflected in Quillian's publications over a period of approximately a decade (Quillian, 1961, 1962, 1966, 1967, 1968, 1969; Collins & Quillian, 1972). Our goal in this section is to present an accurate picture of Quillian's framework, including certain elements that are not widely recognized today.<sup>1</sup>

During the late 1950s and early 1960s, it was widely conceded that lexical ambiguity was the greatest problem facing syntax-based natural language-recognition systems. The most promising of the potential solutions considered at the time was the "thesaurus method," which involved looking up each word of the sentence in an online thesaurus, performing set intersections on the resulting lists of words, and using the intersections to resolve ambiguities in the words of the sentence (e.g., Masterman, 1957). Quillian's model substantially extended the thesaurus method: he used semantic relations among the words in a text either to represent an "understanding" of the text virtually without syntax (earlier versions), or via parallel semantic and syntactic analysis (later versions). It was because of the need to capture the semantic relations of the English lexicon that the structures and processes of semantic memory were conceived. Because of the variable terminology used by Quillian and subsequent writers, and to avoid irrelevant theoretical connotations, our exposition will use the more neutral word *entry* in place of *token*, *concept*, *word*, *property*, *unit node*, *frame*, and so on, and *binding* in place of *connection*, *link*, *type*, *feature*, and so on. The intended metaphor is that of entries in a dictionary or encyclopedia bound together into intricate constellations of meaning as they recur in the bodies of definitions.

## 2.1. Structures

### 2.1.1. Entries and bindings

A Quillian semantic memory consisted of a set of entries, interconnected with arbitrarily complex bindings. Each entry corresponded to a conceptual notion, including but not limited to things like words and propositions. Each entry had associated with it a set of proximate bindings. All bindings had a unidirectional pointer to a predicate (*attribute*) and some number (usually zero or one) of unidirectional pointers to *values*. Note that the attribute was in effect a label on the binding, an idea probably influenced by Quillian's teachers, Simon and Newell, who in turn had been influenced by the work of the German psychologist Otto Selz (Simon, 1981). The predicates and values in bindings were all entries in the semantic memory themselves. For example, an entry for *canary* might have a binding with the attribute label *color* and the value *yellow*. An English-like approximation would be "The color of canary is yellow." In turn, the structure of attributes and values of bindings could

---

<sup>1</sup> We de-emphasize several terminological variants and changes in the model due to implementational constraints such as the core memory capacity of available computers and whether certain elements (e.g., property lists) were present in the available programming languages (e.g., Bobrow & Raphael, 1964; Bobrow & Murphy, 1968); we also will not focus upon the two specific computer simulations implemented by Quillian. The first simulation (Quillian, 1966) modeled inferences based on pairs of words (e.g., *lawyer*, *client*); the second (Quillian, 1969) modeled semantics-driven understanding of written text (children's stories).

have bindings themselves, for example, the attribute label *color* in the example might have a nested binding with the attribute *location* and the value *body-surface*, the value *yellow* might have a nested binding with the attribute *color-saturation* and the value *pale*: “The color of the body-surface of canary is yellow with a color-saturation of pale.” In addition to subnetworks of attributes and values, bindings also optionally had weights that indicated their physical strength, or intensity, as well as values indicating their importance to the entry of which they were a part in terms of *number* and *criticality*, notions borrowed from the classical categorization theory of Bruner, Goodnow, and Austin (1956).

Quillian emphasized several structural aspects of this view of semantic memory. First, each attribute and value of each binding was itself an entry, with its own set of proximate bindings. Second, there was no strict hierarchy in a semantic memory as a whole. Each entry was the root of its own hierarchy, but simultaneously was subsumed under an unlimited number of other hierarchies branching from other entries. Third, the chains of bindings leading away from a given root entry could refer back recursively to the root entry itself; that is, the entry *canary* could appear as an attribute or value of an indirect binding under the root entry *canary*. For example, starting with the verb *to fly*, we might find the following loop: *fly, wing, feather, bird, fly*. Similar longer or shorter recursive loops are quite common.

### 2.1.2. *Superset bindings and cognitive economy*

There was one kind of binding which was particularly important in Quillian’s model, the *superset* binding, also known as the *isa* binding. Superset bindings indicated set inclusion, so for example, *canary* would have a binding with the attribute *superset* and the value *finch* (a canary is a type of finch). Quillian’s model made substantial use of the transitivity of the superset relationship to implement the principle of *cognitive economy*. That is, if *bird* was a superset of *finch*, and *finch* a superset of *canary*, there was no a priori need for a superset binding corresponding to *bird* in the *canary* entry, since this relationship could be inferred using transitivity.<sup>2</sup> More than one superset binding could be present; for example, the entry for *canary* might have superset bindings for both *finch* and for *bird*. Multiple superset bindings were motivated by experimental results showing that high frequency or highly criterial attributes that were logically indirect (e.g., *bird*) could be accessed more quickly than less common, less criterial attributes that were logically more direct (e.g., *finch*; Collins & Quillian, 1972).

## 2.2. Processes

### 2.2.1. *Spreading activation*

One can intuitively visualize the very large constellation of entries and bindings in a semantic memory as being arranged in three-dimensional space such that the distance

---

<sup>2</sup> Logical operators were also allowed in bindings. For example, negation could be used to allow exceptions like flightless birds and legless lizards to be represented; disjunction and conjunction could be used to represent disjunctive and conjunctive criteria.

between entries is a function of how many stages of binding separate them.<sup>3</sup> When an entry is accessed, activation spreads out along its bindings, passing through connected entries and in turn out along their bindings. Quillian (1966, p. 72) expressed this vividly as “firing an activation sphere,” evoking an imaginary bubble expanding out from the original entry. The speed and priority of this breadth-first search could in principle be modulated by number (i.e., quantity) and criteriality tags in the bindings (although the extent to which Quillian applied this feature is not clear). The search would terminate after a certain number of entries had been examined. As activation reached an entry, the entry was given a temporary marker, or *activation tag*, which identified the origin of the current activation sphere and also contained a pointer back to the immediately preceding entry or binding. If a new entry reached by a thread of this activation sphere already had an activation tag from the sphere’s origin, the thread ended there.

### 2.2.2. *Intersections, paths*

As understanding of a text progressed, spheres of activation were rapidly fired from successive words (or other grammatical units) so that multiple parallel spheres were active at the same time. In this case, when activation reached an entry that had a tag from a different sphere of activation, the bidirectional path between the initial entries specified an *intersection*, corresponding to a possibly complex inference involving the entries. For example, in the sentence *John put the canary into a cage*, the binding of *canary* as a *bird*, and the binding of *bird* to *pet* (birds can be pets), plus the binding of *pet* and *bird* to *cage* (pet birds are kept in cages), would lead to the inference that the canary in the sentence could be a pet.

Since spreading activation as described by Quillian was an automatic process not influenced by syntax or other context, each new intersection was evaluated based on its context, including the surrounding syntax, but also in terms of *form tests* specific to the entries. Evaluations of intersections could be conscious—an “interrupt” occurred, in Quillian’s computational metaphor. For example, in the sentence *He saw a canary fly by the cage*, presumably the same intersection involving *canary*–(*bird*–*pet*)–*cage* would be found, but the inference that the canary could be a pet would not be activated, because it is inconsistent with the syntactic frame. If the evaluation of an intersection failed, then it was abandoned and the spreading activation/evaluation process continued until the processing limit was exceeded. As validated intersections were found, they were unified with previously activated intersections to create a situation-specific representation, in the same format as existing entries, which could become a new entry in semantic memory, for example, an entry corresponding to a text or to an episode.<sup>4</sup> There is a fairly good fit between Quillian’s notion of the conscious evaluation of intersections leading to the construction of situation-specific conceptual structures, with Chafe’s

<sup>3</sup> Note that this is strictly illustrative: such a spatial representation was never part of Quillian’s model.

<sup>4</sup> The automatic incorporation of new knowledge into a semantic memory was discussed, but not simulated by Quillian.

(1994) focal/active, peripheral/semi-active, and unconscious/inactive states of knowledge during discourse, with short *intonation units* corresponding to brief activations of ideas, which then quickly recede to semi-activation. This basic concept also can be aligned with Barsalou's (1999) idea that network-like conceptual representations are created dynamically and situation-dependently, thereby creating performance differences in various categorization tasks. For Barsalou, the dynamic structures are created from modality-specific perceptual codes rather than amodal structures, but there is nothing fundamentally incompatible between the formulations of Barsalou and Quillian, if the perceptual codes were organized, accessed, and activated as a semantic memory; in fact, the goal of representing perceptual processing and knowledge was stated several times by Quillian (1966, 1968). These commonalities in the thinking of Quillian, Chafe, and Barsalou suggest that, as Chafe put it, "In the long run, it may be less fruitful to think of something being *in* memory, or retrieving something *from* memory, than to view these phenomena in terms of activation" (p. 53, italics in original).

### 2.2.3. *Supersets, identicalness*

As newly activated intersections were verified and unified with previous material, it was necessary to determine which constituent entities were identical to each other, or in Quillian's terminology, whether two entities could be *identified*. For example, one way to understand the sentence *John bought a canary, but the bird died* is to infer that *canary* and *bird* can refer to the same, identical entity. Quillian's principal identity-verification strategy was based on the *superset intersection*, an intersection that consisted only of superset bindings. Note that if one entry was directly above another entry (*canary–bird*), then in most cases they could stand in the identity relationship. On the other hand, if they shared a common superset but one was not the superset of the other (*canary–pigeon*), then they may or may not be identical, depending on the context and the values of bindings with shared attributes. Much of the early reaction-time (RT) literature on semantic memory by Collins, Quillian, and others, was based on verifying sentences like *A canary is a bird*, *A canary is a robin*, and so on (see Collins & Quillian, 1972, for a discussion and overview). As a result of these influential experiments, and of discussions that emphasized superset intersections more than intersections involving other types of bindings, the Quillian semantic memory has sometimes been characterized mistakenly as if it had only superset bindings. In addition to descriptive bindings (e.g., *color*, *habitat*), there were several other bindings that could be used in logical inference in various ways (including probabilistically): *similarity*, *part*, *proximity*, *adjacency*, *consequence*, *precedence*, and *parent*. All of these bindings, and an unlimited number of others, were part of the basic structure of a semantic memory and were traversed by spreading activation.

## 3. GENERIC SEMANTIC MEMORY

Quillian's conception of a semantic memory was both a set of theoretical assumptions or hypotheses about human memory and language, and an evolving specification for computational simulations based on them. The two simulations implemented by Quillian,

of finding meaningful intersections between specific words, and of building up meaningful representations of English text, were illustrative, but did not exhaust the intended range of applications of the theory, which reached beyond the domain of language understanding. For example, “spatio-visual” memory, the recognition of objects through sensory perception, the generation and storage of imagery, the induction and learning of conceptual knowledge, and analogy and metaphor were all considered prime candidates for modeling via a semantic memory, at least insofar as “the same static store of information” could underlie such other mental phenomena, “rather than supposing that these rely on separate memory structures, even though, such a memory would then have to be richer in interlinkages than that we shall utilize here” (Quillian, 1966, p. 22; see also Collins & Quillian, 1972; Collins, 1975).

During the decade in which the Quillian semantic memory model was being actively developed, there were several extensions and modifications reflecting the results of psychological experiments. Since Quillian’s final publication on semantic memory (Collins & Quillian, 1972), it has been the basis, as was shown in Figure 1, of a continuously growing body of work. Several subsequent avenues of research were connected directly to the Quillian model, for example, its application to the problems of programmed learning and human reasoning (Carbonell & Collins, 1973; Collins, 1978); in addition, several computational models which were quite similar to the Quillian framework (e.g., Schank, 1975; Rumelhart, Lindsay, & Norman, 1972; Anderson & Bower, 1973) were developed during the same period. However, most subsequent work was based on particular components of Quillian’s semantic memory, often altered in meaning, scope, or form. This section contains a brief survey of the key influences on the “genericization” of semantic memory.

### 3.1. Consciousness

While Quillian discussed the interaction of semantic memory and consciousness, he did not emphasize it. Recall that the basic process of spreading activation was unconscious, but when an intersection was found, the subsequent evaluation of the intersection could involve some degree of consciousness. Tulving (1972) noticed that work stemming from or related to Quillian’s framework, while undeniably related to human memory, was quite different from the list-learning or paired-associate learning paradigms that had dominated human memory research up through the 1960s (but cf. Tulving, 1962; Bower, 1972). He proposed two qualitatively different kinds of human memory, one based on episodes, the other on knowledge. He delineated several differences between them, and he called them *episodic* and *semantic* memory. Tulving (1985) fractionated long-term memory further, adding *procedural* memory (but cf. Squire, 1994; Schacter & Tulving, 1994).

Tulving (1985) identified episodic, semantic, and procedural memory with three kinds of conscious experience: *autonoetic*, *noetic*, and *anoetic*, respectively. The three types of memory and their alternate conceptions as types of consciousness were hierarchical: to

say that a certain memory event was episodic, for example, was to say that it was associated with autoeotic consciousness (and possibly with noetic and anoetic consciousness); semantic memory events were associated with noetic consciousness (and possibly anoetic, but not, presumably, autoeotic consciousness). One result of Tulving's proposal has been an equivocation on "semantic memory." Under Tulving's formulation, the dynamics of the system were secondary; semantic memory was defined in contrast to other memory systems in terms of the associated personal, conscious experience. This conflicts directly with central elements of Quillian's framework: constellations of entries bound together, and the twin processes of unconscious spreading activation followed by conscious evaluation of intersections. This equivocation has resulted in some confusion in the literature (see Tulving, 1983, for further discussion). In summary, Tulving conceives of semantic memory as a subset of what Quillian called semantic memory, primarily in contrast to episodic memory; most versions of generic semantic memory adopt Tulving's semantic versus episodic distinction.

### 3.2. Spreading Activation

Recall that in Quillian's framework, spreading activation involved a complex activation marker spread among entries via their bindings. The marker contained information used to control the further spread of activation and to reconstruct the full paths of intersections. The entries and bindings in the resulting path were unified with the developing overall conceptual representation of a text, and could result in the creation of new conceptual entries. Collins and Loftus (1975) redefined spreading activation as a continuously variable process in which activation spread in a fluid-like manner through bindings whose capacity and length were a function of relatedness, criteriality, and so on. The amount of available activation was limited, and it drained away from reservoir-like entries over time. An entry was triggered into activity when it reached a threshold level of activation, and then it began to spread activation through its outputs. Because of the emphasis on bindings of different length and capacity, and on the build-up and fading away of activation, this kind of model was well-suited to describe temporal phenomena such as semantic priming (section 4.2). However, since the concept of a discrete path between entries was abandoned in favor of the activation of individual entries, new concepts could no longer be derived from meaningful, context-specific intersections of existing ones.

In a further fractionation of semantic memory, Collins and Loftus's (1975) revision also assumed two largely independent, specialized semantic memories, one for the lexicon, the other for nonlexical knowledge. This was based on the observation that priming can be found independently for similar-sounding words, conceptually related words, or both (but Collins, 1975, continued to have reservations about this—and any other—division of semantic memory). The distinction between lexical versus semantic memory is fairly common in the literature today; generic semantic memory usually but not always features at least one independent lexicon (see Coltheart, 2004, for a discussion). See Schank (1976) for yet another proposed apportionment of lexical, episodic, and other elements of semantic memory.

### 3.3. Clinical Assessment

In seminal work, Warrington (1975) applied certain aspects of Quillian's framework to clinical memory disorders. In her work, the superset binding hierarchy was emphasized over the connectivity among other types of bindings and the processes of spreading activation and evaluation, a tendency that has characterized neuropsychologists' subsequent interest in semantic memory. While a range of different methods have been used to test semantic memory in clinical settings, the tendency has been to examine the standard repertory of neuropsychological tests of language function, and to choose those that include tests of concept/word knowledge but do not emphasize syntax, phonology, communicative competence, personal memory, and so on. As a result of this selection process, clinical tests of semantic memory have been based to a large extent on two categories of tests: confrontation naming of pictures (usually the Boston Naming Test; Kaplan, Goodglass, & Weintraub, 1983), and word fluency, or "controlled word association," tests in which subjects must generate words that belong to a certain category, such as items found in supermarkets, or that begin with certain letters (FAS). Note that the category versus lexical distinction in fluency tests corresponds somewhat to Collins and Loftus's (1975) distinction between lexical and nonlexical semantic memory.

It has been pointed out that these tests examine only a small subset of semantic memory capabilities, and that they involve cognitive processes (such as attention, working memory, and strategy deployment) not specific to semantic memory (e.g., Rende, Ramsberger, & Miyake, 2002; Ober, 2002; Shenaut & Ober, 1996). In response to the limitations of naming and fluency in the assessment of semantic memory, more nuanced test batteries have been created, such as that of Hodges, Salmon, and Butters (1992), which was based on using the same set of 48 test items in naming, category fluency, sorting, picture-word matching, and verbal generation of definitions. Also, some alternative clinical instruments involving semantic memory have been developed, such as the Pyramids and Palm Trees test (Howard & Patterson, 1992). Still, as the importance of clinically based studies has increased in the semantic memory literature, clinical investigators have contributed to the process of genericization by frequently altering or simplifying the underlying theory and by relying on just a few limited and/or insufficiently specific neuropsychological tests.

## 4. THEORETICAL EXTENSIONS, CONSEQUENCES, AND DIVERGENCES

The decade during which Quillian was developing semantic memory theory ushered in a period of very active theoretical development in psychology and linguistics, and so it can be difficult to determine the original source of contemporary ideas. However, there is a group of important theoretical advances that are allied to Quillian's framework, even though some may not have cited Quillian's work explicitly, or may have referred to a more generic conception of semantic memory. This section briefly surveys a selection of them.



#### 4.1. Eco's Model Q

The semiotician Umberto Eco (1976) pointed out that meaning cannot adequately be represented in terms of decomposition into sets of elements or features, because in order for the features to have meaning within the system, they in turn would need to be decomposed into further sets of elements and features, and so on, in what he called *infinite semantic recursivity*. His solution to this problem was what he called "Model Q" (the "Q" is for "Quillian"), basically semantic memory à la Quillian (1968). Model Q's main advantage for Eco was that in contrast to other approaches to semantics, in which meanings were specified in terms of feature lists or strictly hierarchical trees, no formal distinctions were made in semantic memory among concepts, words, properties, features and so on; instead (section 2.1), each entry served both as a concept defined by the system and, through bindings referring to it, as part of the definition of an unlimited number of other entries (*unlimited semiosis*). He felt it especially important that in Model Q, new meanings were created when the system was "nourished by fresh information," and that "further data could be inferred from incomplete data" (pp. 122–125). This was based on the semantic memory processes which create new inferences from the intersections of two or more spheres of activation. Eco's approach to semiotics has been widely accepted (e.g., Malmkjær, 1991). Furthermore, his insight regarding infinite semantic recursion has great relevance to debates regarding the status of features and other decompositional objects still active in semantic memory research (sections 4.6.2, 9.2).

#### 4.2. The Semantic Priming Effect

In a sentence verification task, Collins and Quillian (1970) noted that RT decreased when a sentence was preceded by another sentence that inferred it. For example, *A canary is a bird* is faster when preceded by *A canary can fly* (which, in Quillian's framework, requires the inference *A canary is a bird*), than when it is preceded by *A canary is yellow*. Their explanation of this is that the lingering effects of prior activation (of the path from *canary* to *fly* through *bird*) facilitates the later verification of its subpath (*canary is a bird*); a similar effect was found when the same inference was shared implicitly by two sentences: *A canary has wings* facilitated *A canary has a beak*. They also noted that prior reading of sentences containing a given word (i.e., *canary*) sped RTs for subsequent sentences that also contained the same word (but that this effect was based on perceptual identity, not semantic relatedness).

Meyer and Schvaneveldt (1971) presented pairs of letter strings (simultaneously, but most later studies presented a prime string followed by a target string) and subjects were timed as they classified them as "words" or "nonwords" (*lexical decision*). They found that word decision times were faster when the prime and the target were semantically related, and they considered Collins and Quillian's (1970) concept of activation lingering after spreading through semantic memory as an explanation of the semantic priming effect. A number of studies followed, showing that the effect was robust, but that it could be affected by many experimental factors—for example, Meyer and Schvaneveldt (1976) demonstrated that degraded stimuli, slowing the reading time, increased the semantic

priming effect. It became clear that semantic priming is the locus of a variety of mental processes, some dedicated to semantic memory, others to more general cognitive processes; moreover, spreading activation has become the most widely accepted explanation (Collins & Loftus, 1975).

#### 4.2.1. *Controlled versus automatic processes*

Neely (1977) found that in addition to the kind of unconscious, automatic spreading activation postulated by Quillian, there was another kind of activation under conscious control that could have a substantial effect on RT and priming effects. This was characterized by increased priming (*facilitation*) from related primes, and negative priming from unrelated primes (*inhibition*), relative to a neutral prime (usually some kind of “ready” symbol or a nonword). For example, when there was a high proportion of related prime–target pairs and ample time between the presentations of the prime and target, subjects adopted an *expectancy* strategy, which involved constructing a mental list of possible associates to the prime; when the target was in the preparatory set, it could be identified as a word more quickly, and when it was not in the preparatory set, the unsuccessful search lengthened RTs. Neely (1991) proposed an additional strategy called *post-lexical semantic matching*, which was relevant only for lexical decision; it involved checking for a relationship between the prime and target after the target has appeared; if a relationship was noted, the subject was biased to make a word (as opposed to nonword) response. In general, the overall priming effect was greater when controlled priming processes came into play.

The presence of more than one type of semantic processing may be relevant to Collins and Loftus’s (1975) division of semantic memory into lexical and nonlexical components based on the presence of lexical priming without phonological or orthographic priming, since there is evidence that phonological priming is more dependent on the use of strategy and conscious awareness than associative priming (Ober & Shenaut, 1988; Norris, McQueen, & Cutler, 2002) or orthographic priming (Napps & Fowler, 1987). This suggests that independent priming from semantic versus phonological primes could be related to differences in strategies used by subjects in performing the task rather than to the existence of separate memory stores (cf. Coltheart, 2004).

#### 4.2.2. *Methodological implications*

Note that Neely’s process-based model represented a substantial shift in emphasis away from earlier models which focused on the structure of concept memory (i.e., features, semantic networks, spatial models) along with a unitary access process. This shift in emphasis became very important in the neuro-related literature, because a deficit in a controlled process such as expectancy or post-lexical congruency, even though it can produce abnormal semantic priming, is not evidence that the basic structure or contents of the knowledge base is impaired, whereas abnormal semantic priming using methodology that minimizes the use of controlled processing is stronger evidence of a true loss of conceptual knowledge. Neely (1991) identified several

methodological variables that affect the degree to which controlled processing is used. The most obvious one is the prime–target stimulus onset asynchrony (SOA): shorter SOAs, below 250 ms or so, do not allow time for expectancy to operate. Continuous priming, that is, a stream of letter strings that are classified or named, reduces post-lexical semantic matching, as does “go/no-go” lexical decision, in which subjects respond only to word targets. Masking the prime to below the threshold of identification has been tried as a method to reduce controlled processing, but there is convincing evidence that even with masking, subjects can identify the prime often enough to employ controlled strategies. The composition of the stimulus set also affects the use of controlled processing. Neely defines the *relatedness proportion* as the proportion of related word target trials to all word target trials, and the *nonword ratio* (which applies only to lexical decision) as the proportion of trials with nonword targets relative to all trials other than related prime–target trials. It turns out that when the relatedness proportion is high, subjects are more likely to generate expectancy sets, leading to more activation and more inhibition. If the nonword ratio is high, subjects are biased toward making nonword responses when the prime and target are unrelated, resulting in greater inhibition on unrelated trials. Therefore, various combinations of procedural manipulations such as short SOA, fairly low relatedness proportions and/or nonword ratios, and go/no-go responses, can be used to reduce the likelihood that subjects will make use of controlled processing in a semantic priming experiment.

### 4.3. Frames, Scripts, Schemata

A critical aspect of Quillian’s framework was that bindings had labels (section 2.1). He distinguished between the proximate bindings in an entry’s *plane* (analogous to a dictionary entry’s definition), and the more remote bindings reached through spreading activation. Planes had the same form as the structure built up dynamically through spreading activation, intersection, evaluation, and unification; when an entry is activated, activation spreads from the bindings in its plane. Note that the structured, labeled bindings in Quillian’s planes were equivalent to frames (Minsky, 1974), scripts, schemata, and constructions. As an example of the connection to frame theory, Collins and Quillian (1972) adopted the grammatical relations proposed in Fillmore (1968) for use as bindings identifying the roles of entries acting together in a proposition. Later Fillmore (1976) introduced *frame semantics*, based on a more highly specified structure related to Quillian’s planes (see Petruck, 1996, for a review). Quillian’s semantic memory is also connected to *cognitive linguistics*, an approach to linguistic theory that focuses on the connection between the interaction effect of cognition, the human body, and the environment on language (e.g., Lakoff, 1987; Croft & Cruse, 2004), and to the various construction grammars (e.g., Fillmore, Kay, & O’Connor, 1988).

### 4.4. Unification

Quillian implemented the process by which new entries resulting from language understanding became unified on an ad hoc basis in IPL-V and LISP 1.5, the principal artificial intelligence languages available at the time. However, during the same decade,

the general unification problem began to receive a great deal of attention (e.g., Robinson, 1965), and new programming languages were developed (most notably, PROLOG) that contained efficient, built-in unification functions. The availability of easy-to-use, efficient unification languages facilitated a number of theoretical advances, often involving data that corresponded more or less to the activated conceptual structures created by Quillian's semantic memory during language understanding. The structures were given labels such as planes, frames, schemata; they contained labeled bindings (similar to Quillian's), and unification operated by locating a binding in two or more structures with the same attribute, whose value was filled in one frame and either filled identically or empty in the others. In this case, the empty values were filled in with the value of the filled binding; when this resulted in all obligatory values being filled, with no contradictions, the entries were unified. As a result of this work, one way that recent models differ from Quillian's is their reliance on more evolved unification functions. See Knight (1989), for a survey of the history, principles, and applications of unification theory.

#### 4.5. Construction Grammars

Quillian's model was focused on semantic relations among entries, but he was also concerned to some degree with the role of syntax in language use. This was reflected in two ways in his work. In the earlier models, ad hoc form tests were part of the evaluation procedure: they rejected intersections that violated syntactic constraints. In later models, he attempted to develop a semantics system that worked in tandem with an independent syntactic network processor (Woods, 1970), such that the two systems interacted with each other as they found intersections or constituents. As mentioned above, subsequent natural language-processing systems such as frame semantics and cognitive linguistics combined semantic and syntactic processing in various ways, with a trend away from the syntax/semantics dichotomy. The ultimate development to date of this idea is the *construction* (Goldberg, 1995; Kay, 2002), which is a complex nexus of bindings containing syntactic and semantic information. A construction, in this context, can be a complex sign such as a word (*top*), a syntactic pattern (NP's NP), or an idiom (*X blew X's top*). In each case, the structure corresponding to the construction has both syntactic and semantic information, and all of the constructions in a system are organized as entries connected via labeled bindings in a semantic network similar to Quillian's semantic memory. Perhaps the construction grammar system most relevant to neuro-related applications is embodied construction grammar, which emphasizes the role of the human body, particularly such elements as situation, perception, and the body's configuration. For example, the fact that we have two hands and ten fingers has influenced syntax and semantics (e.g., Bergen & Chang, 2005). In embodied construction grammars, an interaction-activation process works along with unification-like evaluations of structural correspondences. Since the frame-like constructions contain syntactic, semantic, and physical information, these systems can simulate a range of phenomena, including language understanding, inference making, and even perception and manipulation of the environment, thereby achieving-surpassing-most of the goals stated but not implemented by Quillian (1969).

## 4.6. Concepts and Similarity

Empirical results in the domain of concept knowledge and similarity among concepts have been important influences on the formation of current conceptions of semantic memory. This section briefly considers some of the key phenomena of these overlapping domains.

### 4.6.1. *Prototypes and basic levels*

As mentioned in section 2.1, Quillian adopted the idea of criteriality from the classic model of Bruner et al. (1956). During the 1970s, several developments superseded the classic model. Probably the most critical work was done by Rosch (1975), who worked within a model similar to the generic “superset hierarchy” version of Quillian’s semantic memory. She demonstrated that many concepts appear to be represented in terms of their relation to a possibly abstract *prototype*, which in turn consisted of a set of features. Items that shared many features with the prototype were more *typical* exemplars of the concept than items that shared few features. Second, Rosch (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) demonstrated that not all entities were accessed with the same degree of naturalness. She showed that there were at least three distinct levels: the *basic level* (e.g., *chair*, *horse*), the most easily accessed, the *superordinate level* (e.g., *furniture*, *animal*), which for Rosch could include any entity with a direct or indirect superset relation to the basic level object, and the *subordinate level* (e.g., *bench*, *palamino*), which could include any entity of which the basic level object has a direct or indirect superset relationship. Note that Rosch’s approach introduced an equivocation regarding the terms superordinate and subordinate that has caused some confusion: in most semantic memory models, including Quillian’s, there is a constellation of multi-rooted hierarchical bindings such that there can be many levels of logical superordination or subordination connecting two entries; this conflicts with Rosch’s three-part division of the entire network into basic, superordinate, and subordinate levels. Another possible confusion is that Rosch’s prototypes have often been taken to be “best examples” of a category, as opposed to a kind of summary representation consisting of weighted, possibly contradictory features. For instance, there is no “best example” of a *dog* that could account for the full range of “dogginess” (various sizes, lengths of hair, ear shapes and sizes, and so on); instead, the prototype of *dog* could be represented in terms of an entry with variably weighted bindings encompassing the entire range of variation, but which would not correspond to a particular type of *dog*. This and other aspects of concept theory are discussed by Murphy (2002).

### 4.6.2. *Features, structural alignment*

A classic way to represent the mutual similarity among the items in a set, for example, a semantic category, is geometric, such that the distance in space between each pair of items corresponds to the “mental space” (or dissimilarity) between the items. For example, Rips, Shoben, and Smith (1973) used multidimensional scaling (MDS) to create two- or three-dimensional plots of items in categories such as bird and animal, and found that MDS distances correlated with category verification RTs. Furthermore, the first two or

three dimensions tended to be interpretable as attribute dimensions within the category (e.g., *wild* versus *tame* or *big* versus *small*). However, Tversky (1977) pointed out that MDS analyses frequently violated basic geometric axioms such as the triangle inequality (e.g., no two-dimensional plot can represent the items *bracelet*, *wristwatch*, and *clock* because *bracelet* is close to *wristwatch* but far from *clock*, while *clock* is close to *wristwatch* but far from *bracelet*). Furthermore, while adding more dimensions resolves this issue, it is extremely rare to be able to interpret more than three MDS dimensions in a meaningful way. Tversky proposed a nongeometric, set-theoretic process model of similarity based on asymmetric, weighted matching of their common and distinctive features. For Tversky, “features may correspond to components such as eyes or mouth; they may represent concrete properties such as size or color; and they may reflect abstract attributes such as quality or complexity” (pp. 15–16). While his feature-matching approach was successful at explaining many phenomena of similarity and conceptual knowledge, there were no appropriate bounds on which features were relevant in a certain comparison: in other words, the model was too powerful. An approach originally designed to account for certain phenomena of perceptual similarity (Medin, Goldstone, & Gentner, 1993) led to Markman and Gentner’s (1993) proposal that the cognitive process used in making a conceptual similarity judgment was bounded by the attributes of the items being compared, represented in a frame-like structure. In the structured representation approach, which is similar to unification, the structures of the items must first be aligned to the extent possible; comparisons are made only between corresponding portions of the items. For example, *car* and *truck* have many alignable attributes: trucks have two doors while cars have either two or four doors; both have engines, headlights, and steering wheels. However, *car* and *tree* have almost no alignable attributes except perhaps extremely general ones like hardness, color, ability to move, and size. It has been found (e.g., Markman & Gentner, 1996) that similarity ratings are often more affected by alignable than by nonalignable differences. There is evidence that there can be more than one possible alignment of concepts as a function of past experience, context, and task demands (e.g., Markman, 1999, pp. 289–294).

## 4.7. Connectionist Models

The Quillian formulation of spreading activation was an effective information-processing algorithm, but the process of passing a complex marker among entries was not biologically plausible. Newer, neurologically inspired models involved nodes which could be in various levels of activation, connected by links that could either increase (activate) or decrease (inhibit) a destination node as a function of the level of activation of the source node. There were two main families of models based on extensions of this idea: those where the internal nodes corresponded to individual entities (interactive activation models), and those where there was no isolated representation of individual entities (distributed representation models).

### 4.7.1. Interactive activation models

Recall that Collins and Loftus (1975) had proposed a continuous process of spreading activation; to this was added two additional properties, by analogy with neural systems.

First, there was a process of inhibition (negative activation), such that the activation of one entry caused the activation of another one to decrease. Second, similar entries could be organized into mutual inhibitory sets (*lateral inhibition*). When entries were connected with bindings that could either increase or decrease the activation of other entries with various strengths, it was found that activating certain entries externally caused the system to enter a state of disequilibrium which lasted for a time, but eventually resolved into a stable state corresponding to the result of a Quillian spreading-activation plus evaluation cycle. In particular, lateral inhibition caused all but one of a set of mutually interconnected similar entries to be suppressed, resulting in clear discrimination between activated and not activated states. “Interaction–activation” models were used initially to perform such tasks as letter recognition (McClelland & Rumelhart, 1981) and word–sense disambiguation (Cottrell & Small, 1984), and have been extended to many other domains.

#### 4.7.2. *Distributed representation models*

Quillian’s semantic memory was the first computer-based connectionist model applied to a wide body of problems. Yet there was some concern that because the individual entries and bindings were set up by the experimenter, an element of bias could be introduced. In response, a new class of connectionist models were developed whose inputs, outputs, and architecture were specified by the experimenter, but whose internal structure was a *tabula rasa*. These systems tended to resemble interactivation–activation models, with units organized into massively interconnected planes, and with feed-forward connections from one plane to the next. Some of these models had a feedback mechanism resembling a clocked state machine, where some of the outputs of cycle<sub>*n*</sub> were gated back to become inputs in cycle<sub>*n*+1</sub> (*recurrence*). The systems were programmed using a mechanism known as *back-propagation*: the experimenter determines for each possible input what outputs are correct, and cycles through a presentation of each input (or input sequence in the case of recurrent architectures); the degree to which the outputs differ from the correct output is propagated back along its inputs, at each point, the weights of the connections are changed slightly. This process is repeated until the outputs are close enough, as defined by the experimenter, to the correct outputs for each input configuration. Because of the iterative back-propagation programming sequence, the distinctions are continuous (graded) rather than discrete; this also mimics human performance. While the input and output units are programmed and read explicitly, the intermediate (“hidden”) units change their settings implicitly. These systems have been demonstrated to be extremely powerful, and furthermore to exhibit a critical element of human cognition, *graceful degradation*. See Rogers and McClelland (2004) for an extensive overview of distributed representation models.

Because Quillian-style models of semantic memory have symbolic entries corresponding to concepts, lexical items, and similar mental objects, they are known as *localist* connectionist models in contrast to distributed representation models, in which each interconnected unit is involved to some degree with the system’s response to each input. There has been considerable debate regarding the relative value of localist and distributed

representation models as vehicles for semantic analysis. Three issues are particularly important in this regard. First is the issue of systems that must be programmed versus systems that learn. While it is true that learning is central to back-propagation distributed representation systems, localist systems have been developed with the ability to allocate new entries and thereby to acquire new concepts (see Page, 2000, and commentaries for a thorough overview of this issue; see also Wasikan, 2001). Second, some have held that distributed representation systems are more like actual brain function, in fact, this is why distributed representation systems are also known as “neural networks.” However, the connection tends to be one of inspiration more than a realistic simulation of the brain—for example, no biological system uses anything like back-propagation. Third, since distributed representation systems learn only to produce outputs corresponding to inputs, it is difficult for them to bind together arbitrarily complex, nested propositional structures (the “binding problem”; see Roskies, 1999, and articles in the same volume). Finally, as Page (2000) points out, many purportedly distributed models have important localist structure such as in the arrays of input and output units; that is, they are actually localist-distributed hybrids.

Distributed representation models incorporate assumptions about how degradation of semantic memory could occur in brain-damaged populations, primarily by analogy with graceful degradation. McClelland (1987, p. 472) uses three methods to analyze how a distributed representation system is degraded due to damage: by randomly deleting input nodes, by randomly destroying connections from a unit, and by adding random noise to connection weights. Due to the fact that every unit participates in every input–output mapping to some degree, most of the programming is preserved even with fairly extensive damage. This has led to a large body of work in which various sets of knowledge are programmed (via back-propagation) into a distributed representation system, followed by network damage thought to resemble brain damage due to pathology; the resulting errors in input–output mapping are then compared to the performance of clinical subjects.

Some cognitive neuroscientists, including Tyler and Devlin (e.g., Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Tyler, Moss, Durrant-Peatfield, & Levy, 2000; Randall, Moss, Rodd, Greer, & Tyler, 2004) have developed distributed representation models of semantic/conceptual knowledge that make assumptions not only about the differences in feature representation for *living things* versus *nonliving things* (as per Farah & McClelland, 1991; Tippet & Farah, 1994), but also about the importance of distinctiveness of features and of the correlations among features. A selective review of this work is in section 5.4.

## 5. SEMANTIC MEMORY AND THE BRAIN

The remainder of this chapter will focus on neuro-related aspects and issues of (generic) semantic memory. In this section, we critically review five major models of the brain underpinnings of semantic memory. These models are based on data from case studies and group studies of neuropsychological patients, including those with category



specific impairment (CSI; see Table 1), AD (see Table 2), and semantic dementia (SD; see Table 3). Neuropsychological and structural neuroimaging findings with these patient groups, and more recently, functional imaging findings with normal controls, have provided the empirical database upon which these models have been constructed. Not surprisingly, more recently developed models have greater scope (in terms of phenomena to be explained) and specificity (in their predictions) than earlier models. Although some of these models have proposed multiple systems/subsystems of semantic memory, they all assume that semantic knowledge is organized on the basis of representational constraints imposed by the brain.

Table 1

Overview of category-specific impairment (CSI) and semantic memory.

---

*Description.* Patients with category-specific impairment (CSI) have a disproportionate impairment for particular semantic categories or domains (e.g., *living* versus *nonliving*). A landmark publication on CSI was authored by Warrington and Shallice (1984) who described four herpes encephalitis patients (young and middle-aged adults), with bilateral temporal damage, who had global amnesia, were generally impaired on picture naming and word definition tasks, but were very disproportionately impaired when the stimulus items represented animals or plants (i.e., *living things*) as compared to artifacts (e.g., *tools*, *musical instruments*). The CSI cases reported since 1984 have included some with an artifacts deficit, although a *living-things* deficit is much more common. There have also been occasional reports of patients with selective deficits for specific categories such as *fruits*, *vegetables*, or *animals* (e.g., Caramazza & Shelton, 1998; Hart & Gordon, 1992).

*Neuropsychological considerations & confirmation of a semantic deficit.* There are two methodological issues that are critically important in making the determination that a patient has CSI. First, it must be demonstrated that the performance deficits on the semantic tasks are not due to perceptual problems (e.g., deficits in perceptual-level processing of pictures or words) or word retrieval deficits; otherwise, it will be unclear as to whether the performance deficits are truly semantic/conceptual in nature, as opposed to being restricted to a particular input or output route to/from semantic memory. (This is also important in the evaluation of semantic/conceptual knowledge in AD or SD.) Second, the stimulus materials used from the various categories/domains being tested must be matched on variables that can affect performance (e.g., frequency of occurrence for word stimuli, visual complexity for picture stimuli). Otherwise, it could be the case that the category or domain for which performance is impaired happens to have stimuli that are more difficult to encode, recognize, or name.

*Neurological correlates.* Saffran and Schwartz (1994) describe the patients with disproportionate impairment of *living things* as most commonly having bilateral temporal lobe damage due to herpes encephalitis, and, less commonly, having temporal lobe (and sometimes, additionally, frontal lobe) damage due to infarct(s) or a degenerative disorder of unknown origin. In contrast, Saffran and Schwartz describe the patients with disproportionate impairment of *nonliving things* as most commonly having frontoparietal lesions of cerebrovascular origin. A recent review by Capitani et al. (2003) encompasses 61 CSI patients with deficits for one or more biological categories, and 18 CSI patients with deficits for one or more categories of *nonliving things*; the lesion sites were generally in temporal cortex for the former subgroup, and in frontoparietal or (less commonly) temporal cortex for the latter subgroup.

---

Table 2

Overview of Alzheimer's disease (AD) and semantic memory.

---

*Description.* Alzheimer's disease (AD) is the most common type of dementia, accounting for about one-half of all dementia cases. "Probable AD" and "possible AD" are diagnosed in vivo by universally accepted exclusionary criteria (McKhann et al., 1984). A definitive diagnosis of AD can only be made on the basis of autopsy findings, when brain tissue samples show the requisite concentration of senile plaques and neurofibrillary tangles; when the criteria for "probable AD" are followed, the autopsy data confirm this diagnosis in about 95% of cases (Morris, McKeel, Fulling, Torack, & Berg, 1988; Tierney et al., 1988).

*Neuropsychological profile.* Episodic memory impairment is the hallmark feature of AD; memory for recent events is impaired even in the earliest stages of the disease. Anomia is also evident, as are difficulties in attentional functioning, in the early stages. As the disease progresses, widespread cognitive impairments occur.

*Semantic memory impairment.* The nature and extent of deficits in semantic memory associated with mild-to-moderate AD have been the subject of much debate (see Ober, 1999, and the seven thematic articles that follow, and which comprise this *Journal of the International Neuropsychological Society* symposium, for a broad-ranging discussion of the issues). AD patients are impaired in their performance on standard neuropsychological tests of semantic memory, such as confrontation naming and verbal fluency (e.g., Martin & Fedio, 1983; Ober, Dronkers, Koss, Delis, & Friedland, 1986; Thompson-Schill, Gabrieli, & Fleischman, 1999). However, there are laboratory tests of semantic memory that show normal performance in AD (these findings are reviewed in this chapter).

*Neurological correlates.* The structural and functional brain abnormalities associated with the earlier stages of AD are overwhelmingly in posterior (temporal–parietal) as opposed to anterior neocortex, and are particularly evident in the hippocampus and the medial temporal lobe (e.g., Braak & Braak, 1991; Jack et al., 1997; Parks, Haxby, & Grady, 1993). Only as the disease progresses into the moderate and then severe stages, will significant atrophy and decreased function occur in more widespread areas of the neocortex, including frontal, anterior temporal, and lateral temporal areas (e.g., Parks et al., 1993; Scahill, Schott, Stevens, Rossor, & Fox, 2002). The contrasting brain pathology for AD compared to SD in the early stages of the disease, is consistent with episodic memory being affected disproportionately to semantic memory in AD, whereas the reverse holds for SD.

---

## 5.1. Sensory–Functional Theory

Sensory-functional theory was formulated by Warrington and colleagues (Warrington & Shallice, 1984; Warrington & McCarthy, 1987), who observed that the pattern of intact and impaired categories in CSI patients sometimes, but not always, conformed to a *living things* versus *nonliving things* dissociation. For example, some CSI patients showed impairment for *musical instruments* as well as for *living things*. The assumptions of sensory-functional theory are: (1) semantic memory is organized into modality-specific subsystems (e.g., visual/perceptual, functional/associative); and (2) the ability to recognize (and name) *living things* is relatively more dependent on

Table 3  
Overview of Semantic dementia (SD) and semantic memory.

---

*Description.* In a landmark study, Warrington (1975) documented a pattern of neuropsychological test performance, for three patients with degenerative brain disease which led her to conclude that these patients were suffering from semantic memory deficits. These patients had significant problems with word retrieval and word comprehension, and exhibited impoverished knowledge of many semantic domains. Other aspects of language and cognition, however, including day-to-day event memory were relatively well preserved. Warrington drew upon Tulving's (1972) distinction between episodic and semantic memory, in describing the deficits of these patients as semantic memory deficits. Similar cases were reported in subsequent years; and Snowden, Goulding, and Neary (1989) first coined the term *semantic dementia* for these patients. SD is rare, in comparison to most other types of dementia.

*Neuropsychological profile.* John Hodges and his colleagues have published extensively on SD. Their research has included thorough neuropsychological testing of language, memory, and visuo-spatial functioning (e.g., Hodges, Patterson, Oxbury, & Funnell, 1992; Hodges, Patterson, & Tyler, 1994). SD presents with a language disorder that worsens gradually over time, and is usually not accompanied by other cognitive or behavioral deficits for two or more years from onset in contrast to other, more typical dementias (Mesulam, 2001). It has, in fact, been considered by many clinicians to be a fluent form of primary progressive aphasia (see Grossman & Ash, 2004, for a review of fluent versus nonfluent primary progressive aphasia, with the former being equated to SD).

*Semantic memory impairment.* Detailed assessments of semantic knowledge, using picture as well as word stimulus materials, and assessing visually based semantic knowledge (e.g., Hodges et al., 1994, 1992) support the idea that SD is not just a language disorder, but an actual disorder of semantic memory.

*Neurological correlates.* SD has been termed a temporal variant of frontotemporal dementia, in that SD patients have extensive atrophy (greater on the left than right) in the polar, lateral, and inferior regions of the temporal lobe; this is in contrast to the marked atrophy in frontal regions, without specific semantic memory impairment, seen in the frontal variant of frontotemporal dementia (Galton et al., 2001).

---

visual information, whereas the ability to recognize (and name) *nonliving things* is relatively more dependent on functional information. These assumptions lead to three predictions: (1) dissociations will not occur within the domain of *living things*, since the same semantic subsystem (visual) is critical for all *living things*; (2) patients with category-specific deficits will also have deficits for the modality/type of information which is critically involved in recognizing items from the impaired category (e.g., *living things* deficits should be accompanied by visual-perceptual knowledge deficits across all categories/domains); and (3) patients with disproportionate deficits for a given modality/type of knowledge will also have a disproportionate deficit for the category/domain that depends on that type of knowledge. None of these predictions, however, have been fulfilled (for reviews of the relevant CSI data see Capitani, Laiacona,

Mahon, & Caramazza, 2003; Caramazza & Mahon, 2003; and Caramazza & Shelton, 1998). The original formulation of sensory-functional theory has been largely abandoned. However, revised formulations of sensory-functional types of theories have been developed; one of these is Damasio's (1989) convergence zone theory, upon which Simmons and Barsalou's (2003) conceptual topography theory is partly based (section 5.5).

## 5.2. Sensory–Motor Theory

The sensory–motor theory of Martin, Ungerleider, and Haxby (2000) proposes that conceptual knowledge is represented in the brain according to the features that define the object concepts (e.g., *tools*, *animals*) under study. Moreover, this theory assumes that semantic memory is functionally unitary and distributed over modality-specific representations. This is in contrast to the assumption of sensory-functional theory that conceptual knowledge and modality-specific representations are functionally and neurologically dissociable, and that there are subsystems of semantic memory. Sensory–motor theory is based, in large part, on neuroimaging data with normal subjects in which: (1) retrieval of the color or action associated with given objects activated inferior and superior, respectively, regions of the temporal lobe, areas known to mediate color versus action/motion perception (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995); and (2) naming of *tools* (relative to *animals*) activated an area of left temporal lobe overlapping with that activated by action naming in the Martin et al. (1995) study as well as a region of the left premotor cortex known to be involved in imaging of actions, whereas naming of *animals* (relative to *tools*) activated inner regions of occipital cortex, bilaterally (Martin, Wiggs, Ungerleider, & Haxby, 1996). Additional support for models in which partial representations of a given object are purportedly stored in or near the primary sensory and motor areas that are involved in perception and learning of the object's features comes from numerous functional imaging studies in which retrieving specific types of object attributes (e.g., color, action, and visual form) activated the same brain areas that have been shown to mediate perception of those attributes (e.g., Chao & Martin, 1999; Howard et al., 1998; Kable, Lease-Spellmeyer, & Chatterjee, 2002; Oliver & Thompson-Schill, 2003). Moreover, Martin et al. (2000) and other cognitive neuroscientists (e.g., Badre & Wagner, 2002) have proposed that specific regions of the left prefrontal cortex and the anterior temporal cortex have particular roles in retrieving, maintaining, and selecting semantic information (i.e., in “working with semantic representations”). For a review of the specific brain regions that are engaged in access, selection, and retrieval of semantic representations, see Thompson-Schill, Kan, and Oliver (2006).

## 5.3. The Domain-Specific Hypothesis

The domain-specific hypothesis, a multiple systems/subsystems model, holds that knowledge of a category is not distributed among the sensory–motor systems involved in processing category exemplars, but in category-specific brain systems that are “downstream” from sensory–motor processing (Caramazza & Shelton, 1998). This type of

organization is said to be due to evolutionary pressures for rapid and efficient processing of selected semantic domains, such as *animals*, *plants*, *conspicifics* (members of the same species), and possibly, artifacts such as *tools*. The domain-specific hypothesis (as delineated in Caramazza & Mahon, 2003) predicts that (1) conceptual deficits should typically affect just one of the (evolutionarily significant) categories/domains, and if the system for one domain is damaged it will not be possible for the function of this system to show “recovery” based on the functioning of another such system; (2) there is no necessary association between a deficit for a type/modality of knowledge and a conceptual deficit for a specific category/domain; and (3) perceptual (i.e., pre-conceptual/semantic) stages of object recognition may be functionally organized via domain/category constraints, as are conceptual/semantic stages of object recognition. As per Caramazza and Mahon, evidence in support of the first prediction can be found in Farah and Rabinowitz (2003); evidence in support of the second prediction comes from *living things* as well as *nonliving things* deficit cases showing equivalent impairments of visual/perceptual and associative/functional knowledge (e.g., Caramazza & Shelton, 1998; Laiacona & Capitani, 2001; Moss & Tyler, 2000; Samson, Pillon, & Wilde, 1998); evidence in support of the third prediction comes from cases showing equivalent deficits in visual/perceptual and functional/associative knowledge of *living things*, in the face of visual agnosia for *living things* (e.g., Caramazza & Shelton, 1998; Laiacona, Barbarotto, & Capitani, 1993). Caramazza and colleagues state that there is much indirect evidence for a very limited range of categories (as described above) being affected by domain-specific deficits; this evidence is said to be the pattern of category deficits across patients, as reviewed in Capitani et al. (2003).

#### 5.4. Conceptual Structure Theory

The conceptual structure theory (CST) of Tyler and colleagues (Tyler & Moss, 2001; Randall et al., 2004) is a distributed, connectionist model of semantic knowledge. CST assumes that category-specific semantic deficits (i.e., in CSI patients) as well as the non-specific deficits seen in other neuropsychological populations (including AD and SD) are the result of random damage to a conceptual/semantic system which is not organized a priori by object domain or feature modality. CST incorporates the following specific assumptions: (1) *living things* have more shared features (and thus, fewer distinctive features) than *nonliving things*; (2) for *living things*, biological function information is highly correlated with shared perceptual properties (e.g., *can see – has eyes*); (3) for artifacts, function information is highly correlated with distinctive perceptual properties (e.g., *cuts* as the function of *knife*, via its *blade*); (4) semantic categories, within (*living* versus *nonliving*) domains, differ in their structure (e.g., *vehicles* are less typical than other *nonliving* categories in having more properties overall and more shared versus distinctive properties, than do *tools*); and (5) features that are highly correlated with other features will be more resistant to damage than features that are not highly correlated. Assumptions 1–4 have received support from property generation norming, property verification, and other experimental work with large groups of young normals (YN). Assumption 5 has been evaluated with several semantic-deficit case studies and with computational models, with mixed support. (For reviews of this work see Randall et al., 2004; Tyler

et al., 2000; and Tyler & Moss, 2001.) There are other similarity and correlation-based distributed models of conceptual knowledge that have many of the same assumptions as CST (e.g., Devlin et al., 1998; McRae, de Sa, & Seidenberg, 1997); CST is the only such model, however, that incorporates assumptions about the interaction between distinctiveness and correlation within *living* and *nonliving* domains. A key prediction of CST, stemming from the assumptions that *living things* have relatively more correlated features and highly correlated features are more resistant to disruption, is that a disproportionate deficit for *living things* will be observed when the damage to the semantic system is mild, whereas a disproportionate deficit for *nonliving things* will arise only if the damage is severe enough that all that is left are some of the highly correlated, shared features of *living things*. In contrast, the correlated-structure model of Devlin et al. (1998) assumes that disrupting access to a given feature will disrupt access to highly correlated features, thus predicting that a disproportionate deficit for *living things* will occur at severe levels of damage. The evidence on this issue has been obtained mainly from patients with AD, and is quite mixed/inconclusive (e.g., Gonnerman, Andersen, Devlin, Kempler, & Seidenberg, 1997; Garrard et al., 2001; Zannino, Perri, Carlesimo, Pasqualetti, & Caltagirone, 2002). One limitation of CST is that it does not account for patients with disproportionate deficits for *nonliving things* in the face of relatively intact performance for *living things* (i.e., in patients who are in the earlier versus later stages of disease). Although this pattern of deficits is far less common than the pattern involving an early *living things* deficit, it does occur (e.g., patient “JJ,” described by Hillis & Caramazza, 1991); the domain-specific hypothesis can, of course, account for *nonliving things* deficits, even in mild-to-moderately (as opposed to severely) impaired patients.

A series of neuroimaging studies by Tyler and colleagues was designed to reveal patterns of brain activity specific to category and/or domain (*living* versus *nonliving*) in normal subjects, when stimulus and task characteristics were carefully controlled across these domains; such patterns would support the domain-specific hypothesis as opposed to the CST, whereas the absence of such patterns would support CST (or at least a unitary, distributed model of semantic memory). These papers used both fMRI and PET, lexical-decision as well as category-judgment tasks, picture as well as word stimuli, verb as well as noun word stimuli, and several *living* (e.g., *animals*) versus *nonliving* (e.g., *tools*) categories (Devlin et al., 2002; Pilgrim, Fadili, Fletcher, & Tyler, 2002; Tyler et al., 2003). Findings across these studies showed robust activations for the semantic/categorization tasks as compared with baseline tasks; these activations were mainly in left frontal (particularly inferior frontal) and left temporal regions. However, in none of these studies were there differential activations for the processing of the different categories (*living* versus *nonliving*; different categories within the *living* versus *nonliving* domains). Tyler and colleagues interpret their overall findings as supporting a model in which conceptual knowledge is represented within a unitary, distributed semantic/conceptual system, and as inconsistent with either the domain-specific hypothesis or with the sensory-motor theory hypothesis that different neural networks are responsible for processing different types of information that are associated with different categories of knowledge (e.g., *tools* versus *animals*). It is important to note that Mahon and Caramazza (2003) take issue with the argument that functional neuroimaging results indicating

category-differential (as well as category-identical) patterns of activation, as opposed to those yielding category-selective patterns of activations, are not interpretable within a domain-specific framework (as argued by Moss & Tyler, 2003; Tyler & Moss, 2001). Mahon and Caramazza make the opposing argument that functionally discrete processes do not have to be carried out by non-overlapping neural regions; rather, they can be carried out by overlapping regions (e.g., Martin & Weisberg, 2003). A potentially influential distributed representation model of semantic deficits is that of Rogers et al. (2004). This interdisciplinary model, which is focused on SD, has an emphasis on: (1) the manner in which semantic representations emerge from the interactions among modality-specific representations of objects, and (2) accounting for multiple aspects of normal versus impaired performance on tests of semantic/conceptual knowledge.

### 5.5. Conceptual Topography Theory

Simmons and Barsalou (2003) have developed a theory of conceptual knowledge, conceptual topography theory (CTT), which integrates aspects of both neural-structure and correlated-structure theories. Simmons and Barsalou propose that sensory-functional information, evolutionarily significant categories, and statistical relationships between categories and their properties are all important aspects of the organization of conceptual knowledge. They utilize convergence zone theory (Damasio, 1989; see also Damasio & Damasio, 1994) as a basis for synthesizing aspects of sensory-functional, domain-specific, and conceptual structure theory. Convergence zone theory begins with the widely accepted assumption that when an object is perceived, it activates feature detectors in relevant sensory–motor areas (in the case of vision, these could be, e.g., for shape, orientation, or color); these systems of detectors (called feature maps) are organized hierarchically for vision as well as for other modalities. The key innovation of convergence zone theory is its explanation of how the states of activation within feature maps are stored. Damasio proposes that the neurons in a nearby association area (*conjunctive neurons*) bind the pattern of activated features for use later. These association areas are referred to as *convergence zones* and they are proposed to be organized in multiple hierarchical levels, with the convergence zones that are located near specific sensory–motor areas capturing patterns of activation relevant to that modality (e.g., visual, auditory, motor areas), and the convergence zones that are located away from specific sensory–motor areas are involved in capturing increasingly higher levels of pattern activation, including convergence zones that integrate information across the highest levels of modality-specific convergence zones. Once the feature maps have been established in the convergence zones, the conjunctive neurons in these zones can re-enact the patterns of activation for instances of given concepts without bottom-up sensory stimulation, that is, via recollection/imagery (note the similarity to Quillian-style activation and subsequent unification of subnetworks). Simmons and Barsalou (2003) explain how the addition of two principles to Damasio's (1989) convergence zone theory can result in CTT, which can explain what is known about conceptual deficits, while synthesizing the three dominant theories (sensory-functional, domain-specific, and conceptual structure). It should be noted that the Simmons and Barsalou article is an extension of Barsalou (1999) in which convergence zone theory was applied to Barsalou's perceptual systems theory

of normal conceptual/semantic knowledge; the Simmons and Barsalou extension is meant to explain conceptual/semantic deficits in neuropsychological populations. The two new principles are (1) the similarity-in-topography (SIT) principle, and (2) the variable dispersion principle (which is a corollary of the SIT). Per Simmons and Barsalou (2003, p. 457) “The SIT principle concerns the organization of the conjunctive neurons in CZs (convergence zones). Essentially the SIT principle claims that categorical structure in the world becomes instantiated in the topography of the brain’s association areas. Specifically, the SIT principle states that: The spatial proximity of two neurons in a CZ reflects the similarity of the features they conjoin. As two sets of conjoined features become more similar, the conjunctive neurons that link them lie closer together in the CZ’s spatial topography.” The variable dispersion principle assumes that conjunctive neurons for a category are dispersed in clumps, with a given clump containing conjunctive neurons that are utilized for more than one category. In other words, there is low dispersion for categories with instances that have high similarity (e.g., *mammals*) and high dispersion for categories with instances that have relatively low similarity (e.g., *artifacts*). This relates to category-specific semantic deficits, because a lesion in an area where the clusters for a category are tightly localized will be more likely to lead to disruptions in performance for that category, than when the lesion occurs in an area where there is not this type of localization (i.e., where there is a cluster that is one of many widely distributed clusters for a category/domain). Simmons and Barsalou argue that CTT is quite successful at modeling a wide range of conceptual deficits in neuropsychological populations, as well as accounting for a wide variety of phenomena in conceptual/semantic processing in normal subjects. They do note, however, that additional direct evidence for the SIT principle, that is, data concerning the relation between conceptual similarity and topography within convergence zones, is required. This type of data will most likely have to come from the next generation of high-resolution, event-based neuroimaging studies of conceptual processing, with normal as well as neuropsychological populations. Neither critiques of, or data contradictory to, the CTT have yet appeared in the literature.

An intriguing source of empirical support for Simmons and Barsalou’s (2003) SIT principle is based on the observation that certain “mirror neuron” cells known to be involved in performing certain actions become activated when the subject observes others perform them. For example, the performance of goal-directed actions by humans as well as monkeys activates a network including premotor, motor, and posterior parietal regions; the observation of another individual performing those actions activates the same premotor and posterior parietal regions, but not the motor regions (e.g., Iacoboni et al., 1999). Another example is that when humans observe videos of others inhaling odorants that produce feelings of disgust, the same regions of the anterior insula and anterior cingulate are differentially activated (compared to neutral odorants), as when the disgust-producing odorants are actually inhaled (e.g., Wicker et al., 2003). A final example is the recent (human) finding, that a mirror neuron system in the inferior frontal cortex differentially responds to the observation of grasping actions in an appropriate, meaningful context from the observation of such actions in the absence of context (Iacoboni et al., 2005). All of these findings converge on the notion of highly overlapping brain regions being involved in the perceptual–motor as well as conceptual/semantic aspects of knowledge.



## 6. STORAGE VERSUS ACCESS DEFICITS IN NEUROPSYCHOLOGICAL POPULATIONS

The mixed findings with neuropsychological populations, in terms of whether normal or abnormal semantic memory performance is obtained, depending on the type of task and/or stimuli, have resulted in an ongoing debate as to whether storage versus access deficits are present in these patient populations. CSI, AD, and SD patients (Tables 1–3) sometimes show normal performance on implicit tests of semantic knowledge (e.g., semantic priming, which will be the focus of section 8), in the face of abnormal performance on many types of explicit tasks (e.g., confrontation naming, sorting, verbal fluency). It becomes increasingly difficult to argue for a complete loss of a given type of semantic knowledge, if an implicit test reveals the storage/availability of such knowledge. Hence, some researchers have suggested that task-dependent discrepancies may not simply be due to access/retrieval deficits (with intact storage of semantic knowledge), but rather, may reflect *partial degradation* of semantic knowledge. In other words, task-based performance dissociations have been discussed as consistent with loss of some (as opposed to all) semantic knowledge about given concepts or categories, and with the idea that the remaining knowledge is sufficient to support (for example) semantic priming, but not sufficient to support most explicit tasks. In this section, we will review notions about partial degradation, the “classic” Warrington and Shallice criteria regarding storage versus access deficits, the importance of methods of assessment/experimentation, and theories regarding the storage versus access debate.

### 6.1. Partial Degradation

The concept of partial degradation builds most prominently on the work of Martin and Fedio (1983) and Chertkow, Bub, and Seidenberg (1989) who hypothesized that AD patients, in the earlier stages of the disease, gradually lose specific semantic features, those that enable similar concepts (such as category comembers) to be distinguished from one another. In later stages of the disease, non-distinctive features are gradually lost as well. This type of partial degradation was thought to provide a reasonable explanation of (1) the relative preservation of superordinate category knowledge in comparison to knowledge of basic object concepts (“bottom-up” loss; e.g., Martin & Fedio, 1983; Chertkow et al., 1989; Hodges et al., 1992) as well as (2) greater-than-normal semantic priming (*hyperpriming*; e.g., Chertkow et al., 1989; Martin, 1992). A number of more recent papers have employed this conception of partial degradation to explain AD semantic memory data that do not fit with a simple storage loss or access deficit explanation (e.g., Alathari, Ngo, & Dopkins, 2004; Garrard, Ralph, Patterson, Pratt, & Hodges, 2005; Giffard et al., 2001; Salmon, Heindel, & Lange, 1999). It is important to note, however, that there is a major unresolved question regarding this concept of partial degradation: What, exactly, is the nature of the distinctive/specific semantic features that are lost?

There is an alternate conception of partial degradation that has generated increasing interest over the past 10–15 years: random loss of semantic features, with the degree of loss

gradually increasing over time. This type of degradation was first applied to the study of patients with semantic memory impairment by Farah and McClelland (1991), who showed, via a simple distributed representation model, that the often-reported *living-things* deficit could be simulated with random loss of semantic features. Their model made the following assumptions, based on norming studies: (1) overall, there are more visual features than functional features in semantic memory; and (2) *living things* depend more on visual features, whereas *nonliving things* depend more on functional features. Random loss of “input” features (in the simulation) resulted in a *living-things* deficit on the “output” end. Later, this type of model was applied to AD (e.g., Tippett & Farah, 1994; Devlin et al., 1998).

## 6.2. The Warrington and Shallice Criteria

Warrington and Shallice (1979) laid out proposed performance criteria for distinguishing between patients with storage versus access deficits of semantic memory. It was claimed that these criteria were developed pre-theoretically, i.e., not based on assumptions or predictions from any particular theories of semantic memory structures and processes. The criteria for a storage (as opposed to an access) deficit were stated as (1) consistency in performance on individual items (across tasks and over time); (2) relatively better performance on superordinate than subordinate information about items for which performance is disrupted (e.g., *A canary is a bird* versus *A canary is yellow*); (3) better performance for high frequency as compared to low frequency items; and (4) absence of improvement in performance based on cuing or priming. A fifth criterion concerning rate of presentation, was added by Warrington and McCarthy (1983); it was argued that an improvement in performance when the rate of stimulus presentation was decreased (allowing more processing time) would be indicative of an access, rather than a storage deficit. In an important paper, Rapp and Caramazza (1993) make the case that “... the inferences required to *relate* patterns of impaired performance to a distinction between impairments of access or storage are not theory independent but, instead, rely crucially and necessarily upon assumptions regarding the nature of access and storage” (p. 114). Rapp and Caramazza go on to review and critique the empirical validity of the patterns of deficits described for patients with purported semantic storage versus semantic access deficits, concluding that the evidence for two distinct patterns of performance is not compelling. Moreover, for each of the five criteria, Rapp and Caramazza review the rationale that had been provided by Warrington and colleagues, and then provide illustrations of how different assumptions (than those used by Warrington and colleagues) about the nature of semantic memory representations and the various processes used to access these representations can lead to different interpretations of the patients’ performance. For example, regarding the priming (or cuing) criterion, in which the presence of semantic priming is taken as support for an access (rather than storage) deficit, Rapp and Carramazza note that the underlying assumption for this criterion is all-or-none loss of semantic representations for objects. Alternatively, partial degradation can be posited, as discussed above, and which allows for normal priming in some, but not all circumstances.

It is important to note that, while theoretical assumptions matter a great deal, the methods that are employed are very important as well, in terms of whether data supports storage versus access deficits in neuropsychological populations. Using an example that has been “front-and-center” in the AD literature, and in our lab, the overarching methods issue for semantic priming is the degree to which the priming paradigm utilizes primarily automatic priming processes versus controlled priming processes (sections 4.2, 8.1). Briefly, when abnormalities in controlled semantic priming are obtained in patient populations, it is very difficult to attribute these abnormalities to storage deficits; however, when abnormalities in automatic semantic priming are obtained, storage deficits can be seriously considered. When there are discrepancies among automatic priming experiments, specific methodological factors need to be considered and evaluated experimentally, such as the type of stimuli, the timing parameters, and the subject characteristics.

## **7. SIMILARITY JUDGMENTS AND SEMANTIC MEMORY ORGANIZATION IN ALZHEIMER’S DISEASE**

In spite of showing deficits on clinical tests of semantic memory, AD patients have tended to perform more normally on alternate tests of similar material based on methods that make fewer demands on attentional resources. This has resulted in a debate as to whether AD causes a true structural deficit in semantic memory, or whether process deficits simply make it difficult to access the structure/knowledge that is there. Most clinical and experimental tests (e.g., picture naming, attribute verification, semantic priming) do not provide a clear picture of the structural organization of semantic memory. In an attempt to resolve the issue of a structure versus processing deficit, several AD studies have used inter-item similarity judgments within semantic domains large enough to allow the assessment of conceptual knowledge of these domains in terms of network organization.

### **7.1. Key Methods and Findings**

The two primary methods of analysis used are MDS (section 4.6.2) and Pathfinder analysis (Schvaneveldt, 1990), an alternative way to represent inter-item similarities graphically while avoiding violations of the triangle inequality. Pathfinder provides a simple metric of comparison for Pathfinder networks (*PFnets*) known as *PF closeness* (PFC; see Ober & Shenaut, 1999, for a discussion). Some studies of similarity networks in AD compared to normal elderly have yielded evidence for abnormalities in the AD networks (e.g., Chan, Butters, Paulsen et al., 1993, using a verbal fluency task with *animals*; Chan, Butters, Salmon, & McGuire, 1993, using a triadic comparison task with *animals*); these authors have argued that degradation of semantic memory in AD causes abnormal network organization. In contrast, some studies have shown normal similarity networks in AD (e.g., Bonilla & Johnson, 1995, using an item arrangement task with *animals* and *occupations*; Ober & Shenaut, 1999, using an item arrangement task with *animals* and *musical instruments*); these authors have concluded that semantic memory organization is relatively well preserved in AD (at least in the earlier stages). Differences in the stimuli and/or in the

nature of the similarity task can have a considerable impact on whether the AD similarity networks will be normal or not. Regarding the semantic domain, Chan, Salmon, and De La Pena (2001) found that the AD group's data matched that of controls for *tools* but differed for *animals*, and took this as support for the hypothesis of relative preservation in AD of knowledge about man-made things, with degradation of *living-things* knowledge (Silveri, Daniele, Guistolisi, & Gainotti, 1991, but cf. Tippett, Grossman, & Farah, 1996). Regarding the specific items within a semantic domain, Ober and Shenaut (1999) found nearly identical AD and elderly normal (EN) networks for a slightly different set of *animals* than those used by Chan et al. (2001) and much less agreement between the AD and EN networks for *musical instruments*.

Regarding the tasks that are used to elicit similarity data, in addition to the small differences in the composition of the set of *animals*, Chan, Butters, Salmon et al. (1993) used triadic comparison, while Ober and Shenaut (1999) used an item arrangement task. Evidently, the differences in the choice of items in the domain and/or in the task used to elicit similarity knowledge had a considerable impact on the outcomes of those two studies. In a study using five different elicitation tasks on items from the *railroad locomotive* domain with an expert participant, Gammack (1990) found that different tasks can produce different empirically derived network structures in a normal individual. Moreover, a recent study by Rich, Park, Dopkins, and Brandt (2002) examined semantic memory in AD compared to elderly controls on three different tasks, with a domain of nine *animals*; these authors found that AD patients' performance differed from that of controls as a function of the format and the degree of structure in the tasks, with more normal performance being exhibited with highly structured tasks. Indeed, Storms, Dirikx, Saerens, Verstraeten, and De Deyn (2003) recommended the use of several different similarity elicitation tasks, with patient populations such as AD, in order to determine whether abnormalities are task-specific or persistent; their position is that abnormalities present with some tasks but not with others are generally consistent with preserved knowledge and task-specific access impairments (see also Warrington & Shallice, 1979, 1984).

## 7.2. How Important are Experimental Tasks?

We have tested, in a general manner, the hypothesis presented in several papers (e.g., Chan et al., 1995; Fung et al., 2001; Hodges et al., 1992), that semantic organization, at least in some, if not all domains, is abnormal in AD. Specifically, we recently completed a series of 40 experiments (Shenaut & Ober, 2006) in which we elicited information regarding subjects' knowledge about similarity of items in a variety of domains, using a number of different elicitation tasks with 51 AD, 30 EN, and 30 YN subjects. There were five different tasks, using either implicit (RT) or explicit judgment, comparing 2–12 stimuli at a time, from a given semantic domain. The two implicit (RT) tasks were (1) *Definitions* (Task D), in which subjects were presented with an adapted dictionary definition (e.g., “a small yellow citrus fruit”) and two test items (e.g., “LEMON PINEAPPLE”) and asked to decide as quickly as possible which of the two words met the definition (with slower RTs indicating more similarity); and (2) *Questions* (Task Q), in which subjects were presented with a constant, domain-specific question (e.g., “Which

of these is more useful to humans?”), and two test items (e.g., “DEER ZEBRA”). As for Task D, the RT was used as the indicant of similarity; the subject’s answer to the question was not used in the analysis. The three explicit judgment tasks were (1) the “classic” *Triadic Comparison* (Task T), in which subjects indicated which two out of three items are more similar to each other, with each triplet of items in a domain being compared; (2) *Flags* (Task F), in which subjects arranged all 12 stimuli from the given semantic domain at one time, in terms of similarity; and (3) *Buttons* (Task B), in which subjects rated pairs of items in terms of how similar the items were to each other, using buttons labeled “Very Similar,” “Similar,” “Different,” “Very Different.” There were eight different stimulus domains differing in terms of dimensionality, typicality, and hierarchical structure: *animals*, *bodyparts*, *books*, *clothing*, *colors*, *fruit*, *milestones*, and *tools*.

We briefly describe just one result of this large study in order to contrast the predictions of the feature-matching and structural alignment theories described in section 4.6.2. From the standpoint of feature-matching theory, while of course there could be differing sources of error due to task (e.g., fatigue or practice effects, degree of processing), all of them should converge on the same network representation, since the items had the same features in every case. However, from the standpoint of structural alignment theory, the five tasks also differed in terms of the complexity of structural alignments that they required. Assuming that making a similarity assessment among  $N$  items requires that all  $N$  items be aligned structurally, Task B required only that two items be aligned, Task T required that three items be aligned, while Task F required that all 12 items be simultaneously aligned. Furthermore, on successive trials within a semantic domain, subjects can create new, specific alignments. For example, the alignment of *hammer* with *mallet* would differ from that of *hammer* with *plane*, because they have different commonalities. This suggests two contrasting sources of variability in similarity networks: tasks that compare relatively few items may be more variable due to the lack of a globally aligned view of all items in the domain, and tasks that compare larger numbers of items may be more variable due to processing limitations. These potential task effects are in addition to cognitive factors and to factors affecting individual domains (e.g., how alignable the items in the domain are with each other, and whether there is an obvious structure within the set).

Feature-matching theory predicts that the primary differences will be found between groups, and they will be due to performance issues resulting from age and AD. On the other hand, structural representation theory suggests the prediction of a major effect of task, due to their differing alignment requirements, and possibly also a Group  $\times$  Task interaction, if, for example, degraded semantic knowledge causes less accurate or more variable performance as a function of age or AD. To test these contrasting predictions, we first computed overall PFnets based on the group-size adjusted mean standard distances for each domain (yielding eight baseline PFnets). Then, we computed a similar overall PFnet within each Task  $\times$  Group  $\times$  Domain cell (a total of 120 single-cell PFnets). Finally, we compared each of the single-cell PFnets to its domain’s baseline PFnet, yielding 120 per-cell PFCs. Then, using the eight domains as the random factor, a 5 Task  $\times$  3 Group ANOVA was computed. There was a main effect of task,  $F(4, 28) = 15.0$ ,  $p < .001$ , but the effect of group and Group  $\times$  Task were nonsignificant. Furthermore, all three groups

had the same rank order of task (D Q F B T), and Games–Howell tests found significant differences for all three groups between mean PFCs for Tasks Q and F,  $p < .05$  for AD and EN,  $p < .001$  for AD, but no significant difference between any other pair of tasks for any group. This indicates that the two RT-based tasks, D and Q, produced PFnets that corresponded less well to the overall per-domain baseline PFnets than did the three explicit judgment tasks, F, B, and T. The identity of rank order for tasks indicates that the three groups responded to them in a similar manner. The presence of a large task effect is more compatible with a structural alignment model than with a feature comparison model, and the lack of a group effect indicates that per-group means of similarity judgments are relatively unaffected by age and AD. Returning to the question of overall semantic memory performance in AD, the finding of no group difference coupled with a robust effect of task appears to favor accounts based on processing deficits rather than alternates based on the loss or degradation of structure.

## 8. SEMANTIC PRIMING IN ALZHEIMER'S DISEASE

The first published paper on semantic priming in AD is that of Nebes, Martin, and Horn (1984); they showed normal priming for AD patients with a continuous, word-pronunciation task using strongly associated words, and they concluded that the structure of the associational network was at least grossly intact in AD. A number of semantic priming experiments with AD patients followed; different types of semantic relationships were assessed, with word pronunciation as well as lexical decision-priming paradigms. Some of these studies found normal AD priming, and their authors argued that the structure of the semantic memory network was intact, as was the automatic spread of activation within the network (e.g., Ober, Shenaut, Jagust, & Stillman, 1991; Ober, Shenaut, & Reed, 1995; Nebes, Boller, & Holland, 1986, with sentence frames as primes). Other studies found significantly greater-than-normal AD priming and attributed AD hyperpriming either to a degraded semantic memory network, with degraded concepts having more to gain via spread of activation than intact concepts (e.g., Chertkow et al., 1989; Martin, 1992, with object-decision priming), or to an attentional deficit (Hartman, 1991).

### 8.1. Controlled versus Automatic Semantic Priming in AD

A descriptive meta-analysis and theoretical evaluation of 21 AD semantic priming experiments (Ober & Shenaut, 1995) led to the conclusion that AD hyperpriming was attentionally based, in that greater-than-normal priming effects occurred only with controlled paradigms. A within-subjects assessment of automatic versus controlled priming, which used highly associated stimulus–response pairs as the related prime–target pairs in a lexical decision, pairwise presentation, priming paradigm, with a 250-ms SOA for one half of the trial blocks (i.e., automatic priming) and a 1000-ms SOA for the other half of the trial blocks (i.e., controlled priming) showed normal AD priming at the short SOA and greater-than-normal priming at the long SOA (Shenaut & Ober, 1996). Thus, the same AD subjects showed normal priming as well as hyperpriming, in the same experimental session, simply by preventing versus encouraging the opportunity for attentionally based priming

processes to play a role in task performance. Shenaut and Ober concluded that semantic memory structure was relatively intact in mild-to-moderate AD, and that any abnormalities in priming effects had to do with non-semantic memory processes (attention, strategy, decision-making, etc.). This conclusion was rebutted by Milberg, McGlinchey-Berroth, Duncan, and Higgins (1999) who obtained partial support for their gain/decay hypothesis of semantic activation in AD; this hypothesis states that a reduction in the time constant of spreading activation in AD produces dynamic changes in the availability of semantic representations, and predicts hyperpriming at short SOAs, but hypoprimering at longer SOAs. Milberg et al.'s priming experiment utilized masked lexical decision, and two SOAs: 250 and 1000 ms. They found normal AD priming at the short SOA (same as Shenaut & Ober, 1996), and less-than-normal priming at the long SOA (different from Shenaut & Ober, 1996); thus, hyperpriming was not obtained. Milberg and colleagues have argued, based in part on their priming data, that when abnormal priming is obtained in AD, it is due to abnormal activation processes, rather than to abnormal attentional processes.

## 8.2. Recent Hyperpriming Findings

An in-depth review of semantic priming research in AD, covering the literature through June 2001, is provided by Ober (2002). There is a particularly interesting study concerning hyperpriming in AD, by Giffard et al. (2001), not covered in our prior review, to which we now turn. The authors used highly associated category coordinate (e.g., *tiger-lion*) versus object-attribute (e.g., *zebra-stripe*) related pairs, in an automatic semantic priming paradigm, with a reasonably large sample of AD patients. The same stimulus materials were employed in a multipart, explicit semantic knowledge assessment (involving naming of drawings of the objects, attribute knowledge, and superordinate category knowledge). The AD group as a whole showed hyperpriming for the coordinate pairs and normal priming for the object-attribute pairs; the subgroup of AD patients who exhibited deficits in semantic knowledge showed significantly more coordinate priming than the subgroup of AD patients who did not show such deficits, and the latter subgroup showed significantly more coordinate priming than the control subjects. Giffard and coauthors interpreted their behavioral findings as consistent with the hypothesis of Martin (1992) that hyperpriming for conceptually similar concepts (belonging to the same superordinate category) reflects a loss of some specific attributes for concepts, which, in turn, results in semantic priming for similar concepts becoming more like repetition priming (i.e., priming between repeated instances of the same concept). Hyperpriming for category coordinates is viewed by Giffard and colleagues as consistent with the hypothesis of bottom-up loss of semantic knowledge in AD (and other relevant patient groups), which states that specific attributes of object concepts are lost first, with superordinate category knowledge being preserved until later in the disease process (e.g., Hodges et al., 1992; Martin, 1992). (Also see Giffard et al., 2002, for a longitudinal study of semantic priming, with a subset of the AD patients from their 2001 study.) It should be noted, however, that the hypothesis of bottom-up loss has received mixed support when tested empirically; evidence for AD patients showing comparable performance on superordinate and lower-level (e.g., features of basic level

objects) knowledge has been reported, for example, by Cox, Bayles, and Trosset (1996) and Nebes and Brady (1988).

How can we reconcile the hyperpriming obtained by Giffard et al. (2001) with the normal priming obtained by Shenaut and Ober (1996) in supposedly automatic priming conditions (i.e., 250-ms SOA, low proportion of related pairs)? Both studies used highly associated pairs in the related prime–target condition; however, two-thirds of the related pairs were category coordinates and one-third were object–attribute pairs in the Giffard et al. study, whereas many types of associative relationship were included among the related pairs in the Shenaut and Ober study. One key issue that was not addressed in either the Giffard et al. or Shenaut and Ober study is the degree to which hyperpriming is due to increased facilitation from related primes versus increased inhibition from unrelated primes. Neither study included a neutral prime condition, which would have enabled the partitioning of the overall priming effect into facilitation versus inhibition. The partial degradation explanation of hyperpriming predicts increased facilitation for AD semantic priming; in contrast, the attentional deficit explanation predicts increased inhibition for AD semantic priming. (For a more detailed discussion of the neutral prime issue see Ober, 2002.)

### 8.3. How Generalizable is Hyperpriming?

In an unpublished study (conducted in preparation for Shenaut & Ober, 1996), 21 AD, 15 EN, and 19 YN subjects participated in a two-choice, lexical decision-priming experiment that included an automatic priming condition (250-ms SOA; relatedness proportion of .15) and a controlled priming condition (1000-ms SOA; relatedness proportion of 0.50). There were two sets of highly associated prime–target pairs: 36 category coordinates (CC) and 36 non-category coordinates (NC), with CC and NC pairs matched for association strength. Examples of the CC pairs are: *emerald ruby*, *leopard tiger*. The NC pairs encompassed many types of semantic relationships including part–whole (e.g., *seed plant*) and object–function (e.g., *chair sit*). The same related pairs were used in both conditions, via counterbalancing across subjects. We found overall slowing of RT with aging and AD, significant overall priming, and significantly greater priming for CC compared to NC pairs. All other main effects and interactions, including any interactions involving subject group were nonsignificant. Thus, in these data, there is no trace of AD hyperpriming, for automatic or controlled conditions, for CC or NC (this held for an analysis based on percent priming). These results, particularly the data from the category coordinate condition, are at odds with the Giffard et al. (2001) finding of hyperpriming for CC in AD. What might account for the fact that Giffard and colleagues obtained automatic hyperpriming for CC, whereas we did not? Some possible explanations are (1) differences in the AD subject samples such that relatively more of the Giffard et al. subjects than our subjects had semantic memory impairment (which would presumably be linked to dementia severity), and (2) unknown differences in the nature of the critical stimulus items and/or the filler items. Since it is not yet possible to describe the circumstances under which AD hyperpriming occurs, its theoretical implications are unresolved.



## 9. CONCLUSIONS AND FUTURE DIRECTIONS

In this chapter, we have provided a detailed description of Quillian's original semantic memory model and its transition into a somewhat diverse set of generic elements used in subsequent theoretical developments. We have also provided an overview of neuro-related theories of semantic memory, which have incorporated some aspects of the generic semantic models, but which, until the mid-1990s, were driven mainly by findings from particular neuropsychological populations. Over the past 10 years, the neuro-related models of semantic memory have become increasingly informed by functional neuroimaging findings with both normal and patient populations. Finally, we discussed the storage versus access issue in neuropsychological populations, and then focused on findings from two specific paradigms that we and others have used extensively to evaluate the status of semantic memory in AD: similarity judgments (for items within a given semantic domain) and semantic priming. Here, we conclude the chapter with: (1) an attempt to reconcile the present status of semantic memory research with its origins; and (2) some observations regarding the need for further communication and collaboration between cognitive scientists (including, of course, psycholinguists) who are developing models based solely on normal behavioral data, and neuropsychologists or cognitive neuroscientists who are developing models based on behavioral data from brain-impaired populations and, increasingly, functional imaging data from normal as well as brain-impaired populations.

### 9.1. Relevance of Quillian's Original Model

To what extent is Quillian's original semantic memory model still relevant? If it were being developed today, it would be characterized as a unitary, localist, connectionist model of human long-term memory. Many of its original features could be retained, such as infinite recursiveness, spreading activation, labeled binding structures, and unification, although they could be updated and extended based on post-Quillian technical developments. Several other features, such as the model's use of consciousness, its application to perception systems, and its ability to consolidate new information, largely unexploited by Quillian, currently are of central importance in the literature and would undoubtedly be emphasized more than they were in the 1960s. We have also mentioned several new ideas that seem to us to be highly compatible with a modernized version of the original model: construction grammars unifying syntax and semantics systematize an interaction that was ad hoc in the early models; feed-forward networks, interactive activation, and lateral inhibition are all important extensions to the classic marker-passing spreading activation process used by Quillian. In addition, there have been several proposals regarding cognitive processes within semantic memory other than spreading activation (e.g., controlled processing in priming, structural alignment in similarity judgments and verification); clearly, the updated model would need to accommodate all of these ideas. Another important new idea is that the consequences of the fact that human memory is situated in a human body must be reflected in an adequate, embodied model. This includes both body-oriented elements of meaning, such as the importance of modal simulation in meaning, and also neuro-related restrictions on the structures and processes of the system as a

whole. We believe that several recent developments and discoveries are particularly compatible with a semantic model of human memory that accounts for many of these elements.

## 9.2. Alternate Models

There are two specific theoretical orientations that could be seen as alternates to a Quillian-style semantic memory: fractionated memory models, because Quillian's model was unitary, and distributed representations, because Quillian's model was localist. In fact, there is considerable empirical support for both approaches. As we reviewed earlier, a variety of different divisions of long-term memory have been suggested. In fact, from the perspective of modern construction grammars, which unite syntax and semantics, the first fractionation was due to Quillian himself in that he proposed a semantic processing system that interacted with a largely independent syntax. This was followed by Tulving's proposed division into semantic versus episodic versus procedural fractions, based importantly on phenomena of consciousness; Squire's closely related division into declarative (or explicit) versus nondeclarative (or implicit) memory, based on whether memories can be expressed verbally; Collins and Loftus's division into lexical versus conceptual memory, based on evidence suggesting these types of knowledge could be accessed and primed separately; Warrington's subdivision of Tulving's semantic fraction into functional/associative versus visual/perceptual, based mainly on data from CSI patients; Caramazza's subdivision, again of Tulving's semantic fraction, into separate regions for specific categories, also based primarily on CSI data; Hodges's support for Tulving's episodic versus semantic distinction primarily based on data comparing AD and SD; and various other permutations and consolidations. While qualitatively different phenomena of memory strongly suggest that Quillian's unitary semantic memory framework is insufficient, no single fractionization has accommodated them all satisfactorily. Furthermore, the relative complicatedness and mutual incompatibility of these non-unitary models have led to the search for a simple set of underlying principles that can account for the phenomena of fractionation within a unified memory framework.

The most widely developed set of underlying explanatory principles stems from distributed representation theory. As explained in section 4.7.2, these systems differ from Quillian-style localist semantic networks in that they represent information in terms of stable patterns of activation of visible output nodes in response to a given pattern of activation of the visible input nodes. The input–output mapping is accomplished by one or more interconnected banks of hidden nodes. The mapping can be programmed either by hand-adjusting the weights of the interconnections, or by the use of back-propagation from a random starting point to gradually approach a criterial minimization of the error. Distributed representations have at least three extremely powerful explanatory attributes. First, although back-propagation does not resemble any known neurological process, the fact that complex mappings of inputs to outputs can be achieved without human intervention establishes that automatic learning of complex patterns can be achieved using a simple algorithm in a massively interconnected network. Second, distributed representations share a property with many larger-scale computer networks

such as the Internet: they degrade gracefully when they are damaged (e.g., nodes going offline, connections broken). Memory deficits, especially of dementia patients, have been shown to follow a pattern of partial degradation, which supports the distributed representation framework. Third, some of the most impressive neuro-related support for distributed representations come from work showing an interesting relationship between the configurations of hidden unit weights in back-propagation networks programmed to map distinctive features to concepts. As reviewed in section 6.1, it turns out in this case that the patterns of loss found for living versus nonliving things in patients, are found when lesions are simulated in a distributed representation system. That is, the partial degradation found as a result of those lesions is not general to the entire network, but is greater for units that respond to relatively less correlated features than it is for units that respond to relatively more correlated features. Therefore, no fractionation of the memory system is required to account for the differential degree of loss of living versus non-living knowledge. Perhaps other proposed fractionations could be the result of similar factors; if so, this could lead to a simpler, more explanatory theory.

However, there is still a question as to whether purely distributed representations are adequate to represent all of human memory. First, there is the binding problem, which is related to Eco's infinite semantic recursivity (also known as the *dispersion problem*). One key to the success of Quillian's model was that because it was a localist system with labeled bindings, it is straightforward to do two things that are difficult in distributed systems: to represent complex propositional structures (such as entire discourses, whole texts, or complicated spatial or physical structures), and to represent recursive or reused representations. Various proposals have been made in an attempt to extend distributed representation systems to these phenomena. For example, *temporal binding*, the quasi-simultaneous representation of several network configurations via different frequencies of activation has been suggested as a way to include several different concepts into a single activated structure. However, while this may be adequate for fairly simple propositions in active consciousness, it is much less clear how the proposition could become part of long-term memory. Furthermore, the limit on the number of concepts that can plausibly be temporally bound is far too low (around five at once) to represent even many of the sentences in this paragraph. Another proposed solution is duplication of repeated material in the network. That is, when (e.g.) a certain concept must occur in more than one position in an output configuration, the representation for that concept occurs more than one time in the network. In fact, this is adequate in many cases, but is doomed to fail to represent recursive conceptual structures, as is discussed by Eco. Page points out that many distributed representation systems contain at least some localist elements, such as the input and output units. He proposes a localist system that contains distributed representation within local regions, and with a learning algorithm that involves allocation of new units, similar to the proposals of Quillian, to allow the incorporation of new information. This sort of system promises to offer many of the explanatory benefits of distributed representation systems, while retaining the infinite semantic recursivity and complex bindings possible with localist systems. Recent neurological evidence has been found that calls into question the neuro-plausibility of purely distributed systems. In a series of studies on invariant visual representation, Quiroga, Reddy, Kreiman, Koch, and

Fried (2005) found that single neurons or contiguous neural units responded to a wide range of perceptually different stimuli all referring to a certain concept. Most famously, a single neural unit responded to representations of the actress Halle Berry, including pictures and drawings of her; to pictures where she was dressed as Catwoman; and to the letter string “Halle Berry.” Yet, it did not respond to pictures of other female celebrities, or to other pictures of Catwoman who were not Halle Berry. Neural units also responded specifically to other celebrities (Jennifer Aniston, Julia Roberts, Kobe Bryant), to landmark buildings (Tower of Pisa, Sydney Opera House), animals (spiders, seals, horses), and specific food items. In summary, in spite of their attractive features, at least some common conceptual structures as well as some concept-related phenomena of the brain do not appear compatible with purely distributed representations.

### 9.3. Localist Connectionism and Conceptual Topography

Of the available simulation models available, the localist model with distributed elements described by Page seems to fit most closely with Quillian’s original idea, while incorporating many of the subsequent technical advances and empirical constraints. However, a localist connectionist framework without a connection to the human body is unsatisfactory. We feel that Barsalou’s proposed conceptual topography theory is largely compatible with Page’s conception, in particular because it proposes a neurologically - motivated process for the allocation of new entries (or clusters of neurons) in the network. According to Barsalou’s SIT principle, neurons are allocated such that those representing similar information are close together; this principle recurs in several physically determined layers, from sensory to the highest levels of thought and action. Interconnections are greater within clusters of relatively similar entries, and less among the clusters. This property appears to regionalize the various sensory modalities, but also regionalizes conceptually similar material at all levels. Note that the feature correlation phenomena can be accounted for by SIT. Also, both graceful and sudden loss of knowledge can be accounted for. As far as we know, there has been no effort made to unify a neurologically based allocation strategy with a localist semantic network, but it seems to us that this is a very promising avenue for future research.

### 9.4. Consciousness

The earliest and most widely accepted division of Quillian’s semantic memory was that proposed by Tulving, in which episodic memory is distinct from (or a subset of) semantic, procedural, and other systems of memory. Tulving has characterized this distinction largely in terms of consciousness, and the distinction has given rise to a body of research that utilizes subjects’ “remember” versus “know” judgments to distinguish between whether a certain memory task is largely episodic or semantic. However, relatively little experimental or clinical work has been done on the demarcation of the phenomena of consciousness from those of memory. If memory is to be divided into separate systems in such a way that each system is characterized by and defined in terms of a distinct conscious experience, then perhaps it makes sense to reconsider the notion of a unitary

memory system which interacts in some way with consciousness. We feel that some of the most exciting work on this issue is due to Dalla Barba (2002). Approaching the question from a clinical and a philosophical perspective, Dalla Barba offers a multi-point hypothesis regarding consciousness: consciousness is not unitary, but is a set of modes with which we address an object; among the modes are *knowing*, *imaginative*, and *temporal* consciousness; knowing consciousness underlies imaginative and temporal consciousness; all consciousness is transitive, that is, we are always conscious *of* an object. He identifies one dimension of temporal consciousness (personal past) with Tulving's episodic memory, but points out that deficits affecting past temporal consciousness (such as amnesia) also tend to affect consciousness of the personal present and future. Since there is little connection between memory per se and these latter dimensions of consciousness, Dalla Barba suggests that some neurological deficits that have been characterized as memory disorders may in fact be disorders of consciousness, rather than of memory. This line of research could affect the clinical view of memory, and would lead to the requirement that a model such as that of Quillian and its successors include the simulation of certain phenomena of consciousness, including those discussed by Quillian, but also those mentioned by Chafe, Tulving, and Dalla Barba. Such a model will differ vastly from Quillian's version, probably in ways we have not anticipated here due to advances in technology, but will undoubtedly retain the essential qualities of a semantic memory.

### 9.5. The Role of Cognitive Processes

The majority of neuro-related work on generic semantic memory focused on the structure of memory/knowledge rather than on processes that acted upon it (e.g., spreading activation, strategies, unification, structural alignment). Important structural issues included such questions as: Is knowledge about living things less well preserved in AD than knowledge about non-living things? Are functional features more available than visual features in CSI? Is abstract information lost from storage before concrete information in brain-injured patients with lexical-semantic deficits? Are there specific brain areas which are critically involved in storing information about tools, human faces, or animals? To what extent are there actually individual neurons that are "tuned" to respond to specific concepts/entities? In the section of the chapter which focused on the storage versus access deficits in semantic memory, we emphasized the point that much of what has been attributed to structural deficits (that is, loss from storage) in patient populations may, in fact, be shown to be due to access deficits, or perhaps, access deficits in combination with partial, but not complete loss, of the knowledge being tested. We noted that, in addition to the theoretical assumptions one has regarding evidence for storage versus access deficits, the methods that one uses to assess semantic memory play a very important role in whether performance for a brain-impaired individual will be normal or not. We provided some examples of task effects on semantic memory performance in AD: semantic priming effects as well as the similarity-based network organizations can be normal or abnormal depending on the specific task parameters, even within subjects and with the same stimulus materials. Thus, there seems to be a need for further advances in understanding the processes that normally occur in

the utilization of semantic memory/knowledge, and in understanding how these processes might be disrupted in particular brain-impaired populations, rather than assuming that performance deficits in neuropsychological populations signal loss of knowledge.

## 9.6. General Observations

Finally, we have two very general comments based on the literature review we undertook in preparation for writing this chapter. First, there is great value in having an awareness of the earlier/historical work in understanding the issues, limitations, and gaps in the current literature. This is increasingly difficult to do in any subspecialty within cognitive science and/or cognitive neuroscience, given the vast increase in journals and books in this field over the past several years. Nonetheless, the benefits are potentially great. This leads us to our second observation, regarding the need for much more cross-fertilization between the cognitive and neuro-related researchers. For those of us who are trying to understand semantic memory functioning in AD, CSI, SD, and/or other patient populations, our research can be significantly informed by the earlier as well as the more recent theoretical work in cognitive science and psycholinguistics. For those of us developing models of various aspects of normal semantic memory structure and/or function, an understanding of the nature of the deficits in particular patient populations can, in turn, validate or invalidate components of the models. There are research groups making excellent progress in applying the theoretical frameworks developed from work on normal semantic memory to research with patient populations and to functional imaging work with normals (some of this work was highlighted earlier). A particularly promising and interesting example of what we hope portends the future of semantic memory research is the theoretical work of Simmons and Barsalou (sections 5.5, 9.3). Simmons and Barsalou's conceptual topography theory is well-grounded in both the psychological and neuro-related literature on conceptual knowledge, and was developed to account for a wide range of findings with patients as well as normals. The principles of conceptual topography theory remain to be validated, mainly via functional neuroimaging studies. Much more of this type of work is essential, however, if significant progress is to be made in understanding issues such as: the normal structure and function of semantic memory, how it is disrupted in the various types of brain-impaired populations, the brain underpinnings of semantic memory in all its various manifestations, and the similarities and differences between episodic and semantic memory.

## ACKNOWLEDGMENT

The authors' research has been supported by U.C. Davis, the Department of Veterans Affairs Medical Research Service, the NIH Alzheimer's Disease Research Center (AG-10129), and the California Department of Health Services Alzheimers Disease Centers (Martinez, 03-75266; Sacramento, 98-14970).

**REFERENCES**

- Alathari, L., Trinh Ngo, C., & Dopkins, S. (2004). Loss of distinctive features and a broader pattern of priming in Alzheimer's disease. *Neuropsychology, 18*, 603–612.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
- Badre, D., & Wagner, A. D. (2002). Semantic retrieval, mnemonic control, and prefrontal cortex. *Behavioral and Cognitive Neuroscience Reviews, 1*, 206–218.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.
- Bergen, B. K., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In J.-O. Östman, & M. Fried (Eds), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 147–190). Amsterdam: John Benjamins.
- Bobrow, D. G., & Murphy, D. L. (1968). A note on the efficiency of a LISP computation in a paged machine. *Communications of the ACM, 11*, 558.
- Bobrow, D. G., & Raphael, B. (1964). A comparison of list-processing computer languages: Including a detailed comparison of COMMIT, IPL-V, LISP 1.5, and SLIP. *Communications of the ACM, 7*, 231–240.
- Bonilla, J. L., & Johnson, M. K. (1995). Semantic space in Alzheimer's disease patients. *Neuropsychology, 9*, 345–353.
- Bower, G. H. (1972). A selective review of organizational factors in memory. In: E. Tulving, & W. Donaldson (Eds), *Organization of Memory* (pp. 93–137). New York: Academic Press.
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica, 82*, 239–259.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology, 20*, 213–261.
- Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences, 7*, 354–361.
- Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10*, 1–34.
- Carbonell, J. R., & Collins, A. M. (1973). Natural semantics in artificial intelligence. *Third international joint conference on artificial intelligence*, August 20–23, 1973, Stanford University, Stanford, CA.
- Chafe, W. L. (1994). *Discourse, consciousness, and time*. Chicago, IL: University of Chicago Press.

- Chan, A. S., Butters, N., Paulsen, J. S., Salmon, D. P., Swenson, M. R., & Maloney, L. T. (1993). An assessment of the semantic network in patients with Alzheimer's disease. *Journal of Cognitive Neuroscience*, 5, 254–261.
- Chan, A. S., Butters, N., Salmon, D. P., Johnson, S. A., Paulsen, J. S., & Swenson, M. R. (1995). Comparison of the semantic networks in patients with dementia and amnesia. *Neuropsychology*, 9, 177–186.
- Chan, A. S., Butters, N., Salmon, D. P., & McGuire, K. A. (1993). Dimensionality and clustering in the semantic network of patients with Alzheimer's disease. *Psychology and Aging*, 8, 411–419.
- Chan, A. S., Salmon, D. P., & De La Pena, J. (2001). Abnormal semantic network for “animals” but not “tools” in patients with Alzheimer's disease. *Cortex*, 37, 197–217.
- Chao, L. L., & Martin, A. (1999). Cortical regions associated with perceiving, naming, and knowing about colors. *Journal of Cognitive Neuroscience*, 11, 25–35.
- Chertkow, H., Bub, D. N., & Seidenberg, M. (1989). Priming and semantic memory loss in Alzheimer's disease. *Brain and Language*, 36, 420–446.
- Collins, A. (1975). The trouble with memory distinctions. In: B. L. Nash-Webber, & R. Schank (Eds), *Proceedings of the 1975 workshop on theoretical issues in natural language processing, June 10–13, 1975, Cambridge, Massachusetts* (pp. 52–54). Morristown, NJ: Association for Computational Linguistics.
- Collins, A. (1978). Fragments of a theory of human plausible reasoning. In: D. L. Waltz (Ed.), *Proceedings of the July 27, 1978 workshop on theoretical issues in natural language processing Urbana-Champaign, IL* (pp. 194–201). Morristown, NJ: Association for Computational Linguistics.
- Collins, A. M., & Quillian, M. R. (1970). Facilitating retrieval from semantic memory: The effect of repeating part of an inference. In: A. F. Sanders (Ed.), *Attention and performance III* (pp. 304–314). Amsterdam: North-Holland Publishing Company.
- Collins, A. M., & Quillian, M. R. (1972). How to make a language user. In: E. Tulving, & W. Donaldson (Eds), *Organization of memory* (pp. 309–351). New York: Academic Press.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Coltheart, M. (2004). Are there lexicons? *The Quarterly Journal of Experimental Psychology*, 57A, 1153–1171.
- Cottrell, G. W., & Small, S. L. (1984). Viewing parsing as word sense discrimination: A connectionist approach. In: B. G. Bara, & G. Guida (Eds), *Computational models of natural language processing* (pp. 91–119). Amsterdam: Elsevier Science Publishers.
- Cox, D. M., Bayles, K. A., & Trosset, M. W. (1996). Category and attribute knowledge deterioration in Alzheimer's disease. *Brain and Language*, 52, 536–550.



- Croft, W. A., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Dalla Barba, G. (2002). *Memory, consciousness, and temporality*. Boston, MA: Kluwer Academic Publishers.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*, 25–62.
- Damasio, A. R., & Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In: C. Koch, & J. L. Davis (Eds), *Large-scale neuronal theories of the brain. Computational neuroscience*. Cambridge, MA: MIT Press.
- Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, *10*, 77–94.
- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., & Tyler, L. K. (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia*, *40*, 54–75.
- Eco, U. (1976). *A theory of semiotics*. Bloomington, IN: Indiana University Press.
- Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *20*, 339–357.
- Farah, M. J., & Rabinowitz, C. (2003). Genetic and environmental influences on the organization of semantic memory in the brain: Is “living things” an innate category? *Cognitive Neuropsychology*, *20*, 401–408.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. In: S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds), *Origins and evolution of language and speech* (Vol. 280, pp. 20–32). Annals of the New York Academy of Sciences. New York: New York Academy of Sciences.
- Fillmore, C. J. (1968). The case for case. In: E. Bach, & R. T. Harms (Eds), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Rinehart, and Winston.
- Fillmore, C. J., Kay, P., & O’Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, *64*, 501–538.
- Fung, T. D., Chertkow, H., Murtha, S., Whatmough, C., Péloquin, L., Whitehead, V., & Templeman, F. D. (2001). The spectrum of category effects in object and action knowledge in dementia of the Alzheimer’s type. *Neuropsychology*, *15*, 371–379.
- Galton, C. J., Patterson, K., Graham, K., Lambon-Ralph, M. A., Williams, G., Antoun, N., Sahakian, B. J., & Hodges, J. R. (2001). Differing patterns of temporal atrophy in Alzheimer’s disease and semantic dementia. *Neurology*, *57*, 216–225.

- Gammack, J. G. (1990). Expert conceptual structure: The stability of Pathfinder representations. In: R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 213–226). Norwood, NJ: Ablex Publishing Corporation.
- Garrard, P., Lambon Ralph, M. A., Patterson, K., Pratt, K. H., & Hodges, J. R. (2005). Semantic feature knowledge and picture naming in dementia of Alzheimer's type: A new approach. *Brain and Language*, *93*, 79–94.
- Garrard, P., Lambon Ralph, M. A., Watson, P. C., Powis, J., Patterson, K., & Hodges, J. R. (2001). Longitudinal profiles of semantic impairment for living and non-living concepts in dementia of the Alzheimer's type. *Journal of Cognitive Neuroscience*, *13*, 892–909.
- Giffard, B., Desgranges, B., Nore-Mary, F., Lalevée, C., Beauieux, H., de la Sayette, V., Pasquier, F., & Eustache, F. (2002). The dynamic time course of semantic memory impairment in Alzheimer's disease: Clues from hyperpriming and hypoprimering effects. *Brain*, *125*, 2044–2057.
- Giffard, B., Desgranges, B., Nore-Mary, F., Lalevée, C., de la Sayette, V., Pasquier, F., & Eustache, F. (2001). The nature of semantic memory deficits in Alzheimer's disease: New insights from hyperpriming effects. *Brain*, *124*, 1522–1532.
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gonnerman, L. M., Andersen, E. S., Devlin, J. T., Kempler, D., & Seidenberg, M. S. (1997). Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language*, *57*, 254–279.
- Grossman, M., & Ash, S. (2004). Primary progressive aphasia: A review. *Neurocase*, *10*, 3–18.
- Hart, J., & Gordon, B. (1992). Neural subsystems for object knowledge. *Nature*, *359*, 60–64.
- Hartman, M. (1991). The use of semantic knowledge in Alzheimer's disease: Evidence for impairments of attention. *Neuropsychologia*, *29*, 213–228.
- Hillis, A. E., & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, *114*, 2081–2094.
- Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*, 1783–1806.
- Hodges, J. R., Patterson, K., & Tyler, L. K. (1994). Loss of semantic memory: Implications for the modularity of mind. *Cognitive Neuropsychology*, *11*, 505–542.
- Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia*, *30*, 301–314.
- Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees test. A test of semantic access from words and pictures*. Bury St. Edmunds: Thames Valley Test Company.

- Howard, R. J., ffytche, D. H., Barnes, J., McKeefry, D., Ha, Y., Woodruff, P. W., Bullmore, E. T., Simmons, A., Williams, S. C. R., David, A. S., & Brammer, M. (1998). The functional anatomy of imagining and perceiving colour. *Neuroreport*, *9*, 1019–1023.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *Public Library of Science: Biology*, *3*, 529–535.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526–2528.
- Jack, C. R., Petersen, R. C., Xu, Y. C., Waring, S. C., O'Brien, P. C., Tangalos, E. G., Smith, G. E., Ivnik, R. J., & Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, *49*, 786–794.
- Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience*, *14*, 795–805.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lea and Febiger.
- Kay, P. (2002). An informal sketch of a formal architecture for Construction Grammar. *Grammars*, *5*, 1–19.
- Knight, K. (1989). Unification: A multidisciplinary survey. *ACM Computing Surveys*, *21*, 93–124.
- Laiacona, M., Barbarotto, R., & Capitani, E. (1993). Perceptual and associative knowledge in category specific impairment of semantic memory: A study of two cases. *Cortex*, *29*, 727–740.
- Laiacona, M., & Capitani, E. (2001). A case of prevailing deficit of non-living categories or a case of prevailing sparing of living categories? *Cognitive Neuropsychology*, *18*, 39–70.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.
- Mahon, B. Z., & Caramazza, A. (2003). There are facts...and then there are facts. *Trends in Cognitive Sciences*, *7*, 481–482.
- Malmkjær, K. (1991). Semiotics. In: K. Malmkjær (Ed.), *The linguistics encyclopedia* (pp. 399–401). London: Routledge.
- Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, *32*, 517–535.
- Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition*, *24*, 235–249.

- Martin, A. (1992). Semantic knowledge in patients with Alzheimer's disease: Evidence for degraded representations. In: L. Bäckman (Ed.), *Memory functioning in dementia* (pp. 119–134). Amsterdam: Elsevier Science Publishers.
- Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language*, *19*, 124–141.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*, 102–105.
- Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In: M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1023–1036). Cambridge, MA: MIT Press.
- Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, *20*, 575–587.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*, 649–652.
- Masterman, M. (1957). The thesaurus in syntax and semantics. *Mechanical Translation*, *4*, 35–43.
- McClelland, J. L. (1987). Resource requirements of standard and programmable nets. In: D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds), *Parallel distributed processing: Volume 1: Foundations* (pp. 460–487). Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254–278.
- Mesulam, M. M. (2001). Primary progressive aphasia. *Annals of Neurology*, *49*, 425–432.
- Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. *Science*, *192*, 27–33.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

- Milberg, W., McGlinchey-Berroth, R., Duncan, K. M., & Higgins, J. A. (1999). Alterations in the dynamics of semantic activation in Alzheimer's disease: Evidence for the gain/decay hypothesis of a disorder of semantic memory. *Journal of the International Neuropsychological Society*, 5, 641–658.
- Minsky, M. (1974). A framework for representing knowledge. In: P. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.
- Morris, J. C., McKeel, D. W., Fulling, K., Torack, R. M., & Berg, L. (1988). Validation of clinical diagnostic criteria for Alzheimer's disease. *Annals of Neurology*, 24, 17–22.
- Moss, H. E., & Tyler, L. K. (2000). A progressive category-specific semantic deficit for non-living things. *Neuropsychologia*, 38, 60–82.
- Moss, H. E., & Tyler, L. K. (2003). Weighing up the facts of category-specific semantic deficits. *Trends in Cognitive Sciences*, 7, 480–481.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research*, 16, 257–272.
- Nebes, R. D., Boller, F., & Holland, A. (1986). Use of semantic context by patients with Alzheimer's disease. *Psychology and Aging*, 1, 261–269.
- Nebes, R. D., & Brady, C. B. (1988). Integrity of semantic fields in Alzheimer's disease. *Cortex*, 24, 291–299.
- Nebes, R. D., Martin, D. C., & Horn, L. C. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology*, 93, 321–330.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106, 226–254.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: D. Besner, & G. W. Humphreys (Eds), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Norris, D., McQueen, J. M., & Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory & Cognition*, 30, 399–411.
- Ober, B. A. (1999). Semantic memory in Alzheimer's disease: Loss of knowledge or deficits in retrieval? Introduction from the symposium organizer. *Journal of the International Neuropsychological Society*, 5, 623–625.
- Ober, B. A. (2002). RT and non-RT methodology for semantic priming research with Alzheimer's disease patients: A critical review. *Journal of Clinical and Experimental Neuropsychology*, 24, 883–911.

- Ober, B. A., Dronkers, N. F., Koss, E., Delis, D. C., & Friedland, R. P. (1986). Retrieval from semantic memory in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology*, *8*, 75–92.
- Ober, B. A., & Shenaut, G. K. (1988). Lexical decision and priming in Alzheimer's disease. *Neuropsychologia*, *26*, 273–286.
- Ober, B. A., & Shenaut, G. K. (1995). Semantic priming in Alzheimer's disease: Meta-analysis and theoretical evaluation. In: P. Allen, & T. R. Bashore (Eds), *Age differences in word and language processing* (pp. 247–271). Amsterdam: Elsevier Science B.V.
- Ober, B. A., & Shenaut, G. K. (1999). Well-organized conceptual domains in Alzheimer's disease. *Journal of the International Neuropsychological Society*, *5*, 676–684.
- Ober, B. A., Shenaut, G. K., Jagust, W. J., & Stillman, R. C. (1991). Automatic semantic priming with various category relation in Alzheimer's disease and normal aging. *Psychology and Aging*, *6*, 647–660.
- Ober, B. A., Shenaut, G. K., & Reed, B. R. (1995). Assessment of associative relations in Alzheimer's disease: Evidence for preservation of semantic memory. *Aging and Cognition*, *2*, 254–267.
- Oliver, R. T., & Thompson-Schill, S. L. (2003). Dorsal stream activation during retrieval of object size and shape. *Cognitive, Affective & Behavioral Neuroscience*, *3*, 309–322.
- Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*, 443–512.
- Parks, R. W., Haxby, J. V., & Grady, C. L. (1993). Positron emission tomography in Alzheimer's disease. In: R. W. Parks, R. F. Zec, & R. S. Wilson (Eds), *Neuropsychology of Alzheimer's disease and other dementias* (pp. 459–488). New York: Oxford University Press.
- Petruck, M. R. L. (1996). Frame semantics. In: J. Verschueren, J.-O. Östman, J. Blommaert, & C. Bulcaen (Eds), *Handbook of pragmatics* (pp. 1–13). Philadelphia: John Benjamins.
- Pilgrim, L. K., Fadili, J., Fletcher, P., & Tyler, L. K. (2002). Overcoming confounds of stimulus blocking: An event-related fMRI design of semantic processing. *NeuroImage*, *16*, 713–723.
- Quillian, M. R. (1961). Simulation of human understanding of language. *Communications of the ACM*, *4*, 406.
- Quillian, M. R. (1962). A revised design for an understanding machine. *Mechanical Translation*, *7*, 17–29.
- Quillian, M. R. (1966). *Semantic memory*. AFCRL-66-189. Bedford, MA: Air Force Cambridge Research Laboratories, Office of Aerospace Research, United States Air Force. Doctoral dissertation, Carnegie Institute of Technology.

- Quillian, M. R. (1967). Word concepts: A theory and simulation of contemporary models of semantic capabilities. *Behavioral Science, 12*, 410–430.
- Quillian, M. R. (1968). Semantic memory. In: M. L. Minsky (Ed.), *Semantic information Processing* (pp. 216–255, 267–270). Cambridge, MA: MIT Press.
- Quillian, M. R. (1969). The teachable language comprehender: a simulation program and theory of language. *Communications of the ACM, 12*, 459–476.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*, 1102–1107.
- Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 393–406.
- Rapp, B., & Caramazza, A. (1993). On the distinction between deficits of access and deficits of storage: A question of theory. *Cognitive Neuropsychology, 10*, 113–141.
- Rende, B., Ramsberger, G., & Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology, 16*, 309–321.
- Rich, J. B., Park, N. W., Dopkins, S., & Brandt, J. (2002). What do Alzheimer's disease patients know about animals? It depends on task structure and presentation format. *Journal of the International Neuropsychological Society, 8*, 83–94.
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*, 1–20.
- Robinson, J. A. (1965). A machine-oriented logic based on the resolution principle. *Journal of the ACM, 12*, 23–41.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111*, 205–235.
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: The MIT Press (A Bradford Book).
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–233.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.
- Roskies, A. L. (1999). The binding problem—introduction. *Neuron, 24*, 7–9.

- Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. (1972). A process model for long-term memory. In: E. Tulving, & W. Donaldson (Eds), *Organization of memory* (pp. 197–246). New York: Academic Press.
- Saffran, E. M., & Schwartz, M. F. (1994). Of cabbage, and things: Semantic memory from a neuropsychological perspective – A tutorial review. In: C. Umiltà, & M. Moscovitch (Eds), *Attention and Performance XV. Conscious and nonconscious information processing* (pp. 507–536). Cambridge, MA: MIT Press.
- Salmon, D. P., Heindel, W. C., & Lange, K. L. (1999). Differential decline in word generation from phonemic and semantic categories during the course of Alzheimer’s disease: Implications for the integrity of semantic memory. *Journal of the International Neuropsychological Society*, 5, 692–703.
- Samson, D., Pillon, A., & De Wilde, V. (1998). Impaired knowledge of visual and non-visual attributes in a patient with a semantic impairment for living entities: A case of a true category-specific deficit. *Neurocase*, 4, 273–290.
- Scahill, R. I., Schott, J. M., Stevens, J. M., Rossor, M. N., & Fox, N. C. (2002). Mapping the evolution of regional atrophy in Alzheimer’s disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Science*, 99, 4703–4707.
- Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In: D. L. Schacter, & E. Tulving (Eds.), *Memory Systems 1994* (pp. 1–38). Cambridge, MA: MIT Press.
- Schank, R. C. (1975). The structure of episodes in memory. In: D. G. Bobrow, & A. Collins (Eds), *Representation and understanding* (pp. 237–272). New York: Academic Press.
- Schank, R. C. (1976). Existe-t-il une mémoire sémantique ? [Does a semantic memory exist?] In: S. Ehrlich, & E. Tulving (Eds), *La mémoire sémantique [numéro spécial]* (pp. 26–33). Paris: Bulletin de Psychologie.
- Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex Publishing Corporation.
- Shenaut, G. K., & Ober, B. A. (1996). Methodological control of semantic priming in Alzheimer’s disease. *Psychology and Aging*, 11, 443–448.
- Shenaut, G. K., & Ober, B. A. (2006). Assessment of semantic memory organization in Alzheimer’s disease and normal aging: The role of experimental task and stimulus domain. Manuscript in preparation, University of California, Davis.
- Silveri, M. C., Daniele, A., Guistolisi, L., & Gainotti, G. (1991). Dissociation between knowledge of living and nonliving things in dementia of the Alzheimer type. *Neurology*, 41, 545–546.
- Simon, H. A. (1981). Otto Selz and information-processing psychology. In: N. H. Frijda, & A. D. de Groot (Eds), *Otto Selz: His contribution to psychology* (pp. 147–163). The Hague: Mouton.



Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, *20*, 451–486.

Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioral Neurology*, *2*, 167–182.

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In: D. L. Schacter, & E. Tulving (Eds), *Memory systems 1994* (pp. 203–232). Cambridge, MA: MIT Press.

Storms, G., Dirikx, T., Saerens, J., Verstraeten, S., & De Deyn, P. P. (2003). On the use of scaling and clustering in the study of semantic deficits. *Neuropsychology*, *17*, 289–301.

Thompson-Schill, S. L., Gabrieli, J. D. E., & Fleischman, D. A. (1999). Effects of structural similarity and name frequency on picture naming in Alzheimer's disease. *Journal of the International Neuropsychological Society*, *5*, 659–667.

Thompson-Schill, S. L., Kan, I. P., & Oliver, R. T. (2006). Functional neuroimaging of semantic memory. In: R. Cabeza, & A. Kingstone (Eds), *Handbook of functional neuroimaging and cognition*, 2nd ed. (pp. 149–190). Cambridge, MA: MIT Press.

Tierney, M. C., Fisher, R. H., Lewis, A. J., Zoritto, M. L., Snow, W. G., Reid, D. W., & Nieuwstraten, P. (1988). The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: A clinicopathologic study of 57 cases. *Neurology*, *38*, 359–364.

Tippett, L. J., & Farah, M. J. (1994). A computational model of naming in Alzheimer's disease: Unitary or multiple impairments? *Neuropsychology*, *8*, 3–13.

Tippett, L. J., Grossman, M., & Farah, M. J. (1996). The semantic memory impairment of Alzheimer's disease: Category-specific? *Cortex*, *32*, 143–153.

Tulving, E. (1972). Episodic and semantic memory. In: E. Tulving, & W. Donaldson (Eds), *Organization of memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*, 344–354.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Tyler, L. K., Bright, P., Dick, E., Tavares, P., Pilgrim, L., Fletcher, P., Greer, M., & Moss, H. (2003). Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system. *Cognitive Neuropsychology*, *20*, 541–559.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*, 244–252.

- Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language*, *75*, 195–231.
- Tyler, L. K., Stamatakis, E. A., Dick, E., Bright, P., Fletcher, P., & Moss, H. (2003). Objects and their actions: Evidence for a neurally distributed semantic system. *NeuroImage*, *18*, 542–557.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*, 635–657.
- Warrington, E. K., & McCarthy, R. (1983). Category specific access dysphasia. *Brain*, *106*, 859–878.
- Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge. Further fractionations and an attempted integration. *Brain*, *110*, 1273–1296.
- Warrington, E. K., & Shallice, T. (1979). Semantic access dyslexia. *Brain*, *102*, 43–63.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*, 829–854.
- Waskan, J. A. (2001). A critique of connectionist semantics. *Connection Science*, *13*, 277–292.
- Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.
- Woods, W. A. (1970). Transition network grammars for natural language analysis. *Communications of the ACM*, *13*, 591–606.
- Zannino, G. D., Perri, R., Carlesimo, G. A., Pasqualetti, P., & Caltagirone, C. (2002). Category-specific impairment in patients with Alzheimer's disease as a function of disease severity: A cross-sectional investigation. *Neuropsychologia*, *40*, 2268–2279.

This page intentionally left blank

## Chapter 12

### Syntactic Parsing

*Martin J. Pickering and Roger P. G. van Gompel*

When people hear connected speech or read texts, they start processing immediately. In a classic demonstration, Marslen-Wilson (1973, 1975) had listeners shadow speech, and found that their errors were constrained by prior semantic context even when the shadowing lag was only around 300 ms. This indicates that sentence interpretation can occur extremely rapidly. Many subsequent studies have demonstrated that lexical, syntactic, and aspects of semantic processing occur without appreciable delay, for example during reading (e.g., Just & Carpenter, 1980) or spoken language comprehension in the presence of a visual array (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Furthermore, recent evidence suggests that people may even anticipate properties of upcoming words in the sentence (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

The phenomenon that each word in the sentence is interpreted immediately when it is encountered is referred to as *incrementality*. One aspect of this is that people appear to compute the grammatical structure of sentences incrementally. Most of the evidence for incremental syntactic processing or *parsing* comes from the study of potentially ambiguous sentences. For example, in sentence (1), people experience difficulty as soon as they encounter the phrase *by the lawyer* (Clifton et al. 2003; Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983; Trueswell, Tanenhaus, & Garnsey, 1994):

1. The evidence examined by the lawyer turned out to be unreliable.

After *the evidence examined*, the sentence is temporarily (or locally) ambiguous between the correct “reduced relative” analysis in which *examined* is a past participle, and a “main clause” analysis in which *examined* is a past tense verb. But the phrase *by the lawyer* provides very good evidence that the main clause analysis is wrong and that the reduced relative analysis is correct. Difficulty reading *by the lawyer* therefore suggests that people initially select or favor the main clause analysis, and experience difficulty when they realize that this analysis is probably wrong. In psycholinguistic terminology, they are “garden pathed” (Bever, 1970).

Such findings show that people do not wait until the end of the sentence before syntactically analyzing it. They are compatible with a *serial* account, in which people initially adopt one analysis (in this case, the main clause analysis). On this account, difficulty occurs when people realize this inconsistency and *reanalyze* (i.e., adopt a different analysis). They are also compatible with a *ranked parallel* account, in which people adopt more than one analysis, but rank one higher than any others. Difficulty occurs when later information causes people to re-rank their analyses. All current accounts assume that syntactic processing is either serial or ranked parallel.

Hence syntactic and semantic processing begin without any appreciable delay. This might suggest that the processor has an architecture in which different sources of information are integrated immediately. However, this assumption has proved extremely controversial, as we shall now see. We therefore first review theories and data relating to initial processing. Following this, we turn to questions of reanalysis and processing complexity, and then address a range of newer topics in parsing research that provide important links between the field and other areas of psycholinguistics and cognition more generally.

## 1. TWO-STAGE ACCOUNTS

For a long time, the study of parsing was almost entirely dominated by the question of how initial parsing decisions are made. The “big” question underlying this research was the question of *encapsulation*: the extent to which different knowledge sources are formally separated. Encapsulation is a critical property of *modularity*, which roughly amounts to the thesis that the mind consists of separate, specialized components that exist independently of a central store of general knowledge (J. A. Fodor, 1983). Although Fodor treated language as one module, parsing research has tended to ask whether specific aspects of language, such as syntax, are modular.

In practical terms, the investigation of encapsulation has addressed the issue of ambiguity resolution during initial processing. The earliest accounts of syntactic ambiguity resolution assumed that decisions were based on strategies such as interpret a string of words as agent verb patient (the *NVN* strategy) if possible (Bever, 1970; cf. Kimball, 1973). Such accounts are modular, because they make reference to syntactic information alone, and do not, for example, pay attention to the plausibility of the alternative analyses. However, people eventually make use of information like plausibility in choosing an analysis. Thus any modular account is *two-stage*, with initial processing being modular, but subsequent processing being usually not modular. This dichotomy is most clearly described in Rayner et al. (1983), where a separate (*thematic*) processor plays no role in the initial choice of analysis, but is used during reanalysis.

So what guides initial processing? Although Bever (1970) and Kimball (1973) proposed sets of principles, it was Frazier (1979, 1987a) who established by far the most

influential account, which came to be known as the “Garden-Path” theory.<sup>1</sup> In common with almost all two-stage accounts, it assumes serial processing. A single analysis is chosen on the basis of principles defined in terms of a phrase-structure tree. *Minimal Attachment* stipulates that an ambiguous phrase is attached to the preceding tree structure using the fewest number of nodes. Consider (2):

2. The spy saw the cop with binoculars but the cop didn’t see him.

Frazier assumed that VP-attachment in the first clause (i.e., the spy used binoculars to see the cop) involved a flat tree structure [V NP PP]<sub>VP</sub> whereas NP-attachment (i.e., the cop had binoculars) involved an embedded complex noun phrase [V [NP PP]<sub>NP</sub>]<sub>VP</sub>. Since VP-attachment involves one fewer node than NP-attachment, it is the minimally attached analysis and is therefore adopted initially. Rayner et al. (1983) tested this prediction by contrasting it with (3), where the VP-attached analysis is implausible:

3. The spy saw the cop with a revolver.

Using eye tracking, they found that readers had more difficulty with sentences such as (3) than with sentences such as (2), and suggested that they attach to the VP in both cases, but subsequently revise their initial decision in (3). This reanalysis makes use of the thematic processor.

Minimal attachment was used to explain a number of other types of locally ambiguous sentences, including reduced relatives (4), object/complement ambiguities like (5), and ambiguities caused by head-final verbs in languages like Dutch (6):

4. The florist sent the flowers was very pleased.
5. The man realized his goals were out of reach.
6. ... dat het meisje van Holland glimlachte.  
“... that the girl from Holland smiled.”

In all of these cases, minimal attachment posits that the processor initially adopts the ultimately incorrect analysis. Notice, however, that it can explain preferences in globally ambiguous sentences like (2) as well as locally ambiguous sentences. In all the cases, some experimental evidence provided support for the principle (e.g., Frazier & Rayner, 1982; Frazier, 1987b; Rayner et al., 1983; Ferreira & Clifton, 1986; Ferreira & Henderson, 1990), but other evidence supports alternative accounts (see below).

---

<sup>1</sup> Note the capitalization. In our usage, garden path is a descriptive term. In a serial account, people garden path if they choose a syntactic analysis that later turns out to be incorrect. In a ranked parallel account, people garden path if their highest ranked analysis later turns out to be incorrect. Garden Path theory, on the other hand, refers to the two-stage serial account proposed by Frazier (1979, 1987a).

Late closure stipulated that incoming material was to be attached into the constituent currently being processed. It appeared to deal with two rather different types of ambiguity. First, it predicted that a phrase would form part of the current constituent rather than start a new constituent, as in (7). Second, it predicted that a phrase would form part of the most recent constituent possible, as in (8):

7. When Mary was knitting the socks fell to the floor.
8. John said that Sue left yesterday.

In (7), *the socks* can serve as the object of *knitting*, in which case it forms part of the verb phrase headed by *knitting*; or as the subject of a new clause, as turns out to be the case. Late closure predicts that people would initially treat *the socks* as the object of *knitting*. But this analysis becomes impossible after *fell* (because *fell* would have no subject). In accord with this, Frazier and Rayner (1982) found that people experienced difficulty after reading *fell*. In the globally ambiguous (8), late closure predicts that *yesterday* should modify *left* (i.e., Sue left yesterday) rather than *said* (i.e., John spoke yesterday). This amounts to a preference for “low” attachment over “high” attachment (Kimball, 1973). According to Frazier (1987a), minimal attachment takes precedence when the principles are in conflict. For example, the processor adopts VP-attachment in (2), in accord with minimal attachment, even though late closure would support NP-attachment. But in (7) and (8), Frazier claims that the two analyses do not differ in the number of nodes, and so late closure applies.

In fact, minimal attachment and late closure require certain additional assumptions to make the above predictions for (1)–(8). First, they depend on specific assumptions about tree structure. For example, minimal attachment only predicts VP-attachment of (2) if the VP-analysis has a flat structure, an assumption which is not adopted within modern transformational frameworks (e.g., Kayne, 1984; Chomsky, 1995). Second, the processor sometimes has to postulate structure associated with words it has not encountered yet. In (6), the noun phrase *van Holland* could constitute an argument of a following verb, such as *houdt* (liked), because such arguments precede past tense verbs in Dutch subordinate clauses. Minimal attachment predicts that *van Holland* is an argument rather than a modifier of *het meisje* so long as the processor postulates the verb at the noun phrase.

In later years, Garden-Path theory introduced another principle known as the *active-filler strategy* (Frazier, 1987b; Frazier & Clifton, 1989; Frazier & Flores D’Arcais, 1989) or the *minimal chain principle* (De Vincenzi, 1991). This was designed to deal with unbounded dependencies, such as *wh*-questions and relative clauses, in which the verb and its argument can be separated by clause boundaries:

9. Which girl do you believe John loves a lot?

In (9), *which girl* is the object of *loves*, but is found at the beginning of the sentence. Garden-Path theory assumed that such sentences involved transformations, so that the *filler* (*which girl*) moves from its canonical location after *loves* to the front of the sentence, and leaves a *gap* (known as a *wh-trace*) at its canonical location (e.g., Chomsky,

1981). Psycholinguistic accounts that assume transformational grammar (e.g., J. D. Fodor, 1978) claim that the processor associates the filler and the gap (*gap filling*), and then integrates the filler with the verb. The active-filler strategy assumes that the processor favors any analysis that allows gap filling over any analysis that does not. In (9), it is locally possible that *which girl* serves as the object of *believe*, if it is used transitively, so Garden-Path theory predicts that this analysis is initially adopted. It becomes impossible as soon as *John* is encountered, and therefore causes what has been termed a “filled gap” effect (Stowe, 1986).

The active-filler strategy makes an interesting prediction for word order ambiguities and relative clause ambiguities in languages like Dutch and German (Frazier, 1987b; Frazier & Flores D’Arcais, 1989). For example, it predicts that temporarily ambiguous Dutch subject relative clauses such as (10a) are easier to process than object relatives such as (10b).

10a. Karl hielp de mijnwerker die de boswachters vond.

“Karl helped the miners who found the forester.”

10b. Karl hielp de mijnwerker die de boswachters vonden.

“Karl helped the miners who the forester found.”

Because languages such as Dutch and German are assumed to have an underlying subject-object-verb order (e.g., Koster, 1975), the subject gap in (10a) caused by the extraction of the relative pronoun filler (*die*) precedes *de boswachters*, whereas the object gap in (10b) follows it. As a result, the subject gap can be filled earlier than the object gap, so subject relatives should be easier to process than object relatives. For similar reasons, the active-filler strategy also predicts that temporarily ambiguous sentences with a subject-verb-object (SVO) order in Dutch and German should be easier to process than sentences with an object-verb-subject (OVS) order. These predictions are supported by several studies (e.g., Bader & Meng, 1999; Frazier, 1987b; Frazier & Flores D’Arcais, 1989; Hemforth, 1993; Kaan, 2001; Mak, Vonk, & Schriefers, 2002; Schriefers, Friederici, & Kühn, 1995).

Notice that the active-filler strategy assumes the existence of gaps in accord with most versions of transformational grammar (e.g., Chomsky, 1981). However, alternative linguistic theories eschew gaps (e.g., Pollard & Sag, 1994, Chapter 9; Steedman, 2000), and in accord with such theories, Pickering and Barry (1991) proposed that the processor associates fillers directly with the verb (or other subcategorizer) without going via a gap. Such an account makes the same predictions as the active-filler strategy for sentences like (8). It does not account for ambiguity resolution preferences in word order ambiguities and relative clauses in Dutch and German, so they would have to be due to other factors, for example a preference for a particular information structure (e.g., Kaan, 2001). However, it does explain why people experience difficulty with (11) at the verb *shot*, even though the gap would follow *the hapless man* (Traxler & Pickering, 1996):

11. That is the very small pistol in which the heartless killer shot the hapless man yesterday afternoon.



The active-filler strategy can only predict this finding if the processor can postulate gaps before they occur (Gibson & Hickok, 1993).

Although Garden-Path theory has been very much the dominant two-stage account, there have been many alternative accounts in which initial decisions are based on some sources of information to the exclusion of others. Most of these accounts propose that the processor initially adopts an analysis in which the verb (or other element) can assign a thematic role to a new constituent (Abney, 1989; Crocker, 1995; Pritchett, 1992). This can also be interpreted as an initial preference for adjuncts over arguments. For example, Schütze and Gibson (1999) found a preference for NP-attachment in sentences similar to (2) when a prepositional phrase was ambiguous between being an argument of the noun phrase and an adjunct of the verb phrase (cf. Clifton, Speer, & Abney, 1991). Such research could be used to discriminate between different modular accounts, but in practice far more attention has been paid to the question of whether the processor is modular or not. Before reviewing relevant evidence, we briefly outline the main characteristics of interactive (i.e., non-modular) accounts.

## 2. INTERACTIVE ACCOUNTS

In contrast to modular sentence processing accounts such as the Garden-Path model, interactive accounts assume that all potentially relevant sources of information can be used immediately during sentence processing and can affect initial processing decisions. Current interactive sentence processing models, which have been developed from earlier interactive models (e.g., Bates & MacWhinney, 1989; Tyler & Marslen-Wilson, 1977; Taraban & McClelland, 1988), are often called constraint-based (or constraint-satisfaction) models (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Kello, 1993; Trueswell et al., 1994); though there are also hybrid models in which some information can be delayed (e.g., Boland & Blodgett, 2001; Boland & Boehm-Jernigan, 1998). Interactive models generally assume that the processor activates all possible analyses of a sentence in parallel, and that the activation of the analyses depends on the amount of support they receive from the various sources of information. When one analysis receives much more support than its alternatives, processing is easy, but when two or more analyses receive about equal support, processing difficulty occurs. Additionally, they tend to be lexicalist, in that they assume that most or all syntactic information is stored with individual lexical items (e.g., MacDonald et al., 1994; Trueswell, 1996).

One of the difficulties with constraint-based models is that they tend not to be very predictive unless they have identified the full set of constraints that affect processing and have a precise model of how these constraints affect processing. However, some recent accounts have used computational modeling to derive more precise predictions. For example, Spivey and Tanenhaus (1998) and McRae et al. (1998) reported a model that explains how various sources of information affect phrase-by-phrase processing of reduced-relative ambiguities. In this model, all syntactic analyses of an ambiguous structure are activated in parallel, and their activation is determined by various constraints. At each word in the sentence, the activation of the analysis that receives most support from

the constraints increases until it reaches a threshold level and the processor moves to the next word. Reading times are modeled as the number of cycles the processor has to go through before it reaches this threshold level of activation. Hence, if two or more analyses receive about equal support from the various constraints, reading times should be long, but if one analysis receives much more support than its alternatives, reading times should be short.

Tabor, Juliano, and Tanenhaus (1997) and Tabor and Tanenhaus (1999) developed a learning-based model to predict sentence processing difficulty. During learning, the model maps language input into a multidimensional space. Sentence fragments that have similar continuations (because they are syntactically and semantically similar) occupy positions close together in the space and form clusters that function as attractors. These attractors can be considered as different analyses of ambiguous sentence fragments. Reading times during subsequent sentence processing are modeled as the time it takes for a sentence fragment to reach one of the attractors. When a sentence fragment is very similar to a single cluster of previously encountered fragments (i.e., it has a similar syntax and semantics), it starts at a position very close to the cluster, so it will quickly reach the attractor and reading times are predicted to be fast. In contrast, when a sentence fragment is similar to more than one attractor, that is, the syntax and semantics of the fragment is consistent with more than one analysis, reading times are long.

Computational models make constraint-based accounts more testable and have highlighted the different sources of information that appear to affect sentence processing. We now turn to these sources of information.

### 3. FREQUENCY EFFECTS

An obvious possibility is that choice of analysis is affected by frequency. It is therefore not surprising that many early accounts suggested that the processor initially favors frequent analyses over infrequent ones (Clifton, Frazier, & Connine, 1984; J. D. Fodor, 1978; Ford, Bresnan, & Kaplan, 1982; Mitchell & Holmes, 1985). Such accounts almost inevitably clash with Garden-Path theory, because principles like minimal attachment make no reference to frequency. On the other hand, frequency plays a very natural role within interactive accounts, so evidence for early effects of frequency would support them.

However, any frequency-based account needs to answer the question: frequency of what? The early accounts largely assumed that the processor counted the frequency with which particular verbs (or other elements) were used in a particular construction. Some verbs are most commonly used transitively (e.g., *read*), whereas others are most commonly used intransitively (e.g., *sing*), and the processor might simply adopt the most frequent analysis for each verb in cases of ambiguity. Alternatively, the processor might ignore the frequency of individual verbs, and simply adopt the most frequent construction. For example, verbs are more commonly used transitively than intransitively (in English at least) and so the processor might always adopt the transitive analysis, even when the verb is most commonly used intransitively. Between these extremes, it could

“group” similar verbs together (e.g., assessing the frequency with which verbs of physical transfer are used transitively). It also has to resolve the question of how “construction” is defined (e.g., do transitives in main clauses and subordinate clauses get “counted” together?). Mitchell, Cuetos, Corley, and Brysbaert (1995) refer to such issues as questions of *grain size*. Fine-grained accounts pay more attention to individual properties of sentences, but may face the “sparse data problem” if many categories are counted. Coarse-grained models make predictions that are simply very inaccurate for some ambiguities.

Mitchell et al. (1995) assumed that the processor would make a specific choice about grain size, and proposed that it used only coarse-grained information during initial processing (i.e., information independent of lexical items). In contrast, constraint-based theories claim that the processor employs fine-grained information associated with individual lexical items (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald et al., 1994; Trueswell et al., 1993; Trueswell, 1996). In fact, they also tend to assume that the processor uses various types of frequency information during syntactic ambiguity resolution including coarse-grained frequency information that is independent of lexical items, lexically specific frequency information, and perhaps even more fine-grained information associated with combinations of lexical items (e.g., McRae et al., 1998).

To explore the role of frequency, let us consider the resolution of complex noun-phrase ambiguities like (12), in which the Spanish sentence (12b) is a translation of the English sentence (12a):

- 12a. The journalist interviewed the daughter of the colonel who had the accident.  
 12b. El periodista entrevistó a la hija del coronel que tuvo el accidente.

This construction has been particularly important in assessing the Garden-Path theory. According to late closure, people should initially assume that *who had the accident* attaches “low” to *colonel* (i.e., so that he had the accident) rather than “high” to *daughter* (i.e., so that she had the accident). Although Cuetos and Mitchell (1988) found evidence for low attachment in (12a), they found evidence for *high* attachment in (12b). More recent evidence suggests that English shows either a weak preference for low attachment or no clear preference at all (e.g., Carreiras & Clifton, 1993, 1999; Traxler, Pickering, & Clifton, 1998) and there is some evidence for a similar low attachment preference in Italian (De Vincenzi & Job, 1995; but cf. Frenck-Mestre & Pynte, 2000), but other languages such as French, German, and Dutch show a high attachment preference (e.g., Brysbaert & Mitchell, 1996; Hemforth, Konieczny, & Scheepers, 2000; Zagar, Pynte, & Rativeau, 1997).

Several accounts explaining cross-linguistic differences in relative clause attachment have been proposed. Frazier and Clifton (1996) proposed Construal theory, which is a hybrid processing model in that it claims that structural parsing principles such as minimal attachment and late closure operate for structures involving primary syntactic relations (roughly, involving arguments), whereas for non-primary relations, as found in relative clause attachment, the processor immediately uses non-syntactic information.

Frazier (1990) proposed that low attachment in (12a) is preferred in English because the Saxon genitive (the colonel's daughter) is more common, so when the Norman genitive (the daughter of the colonel) is used, it indicates that relative clause attachment is low. By contrast, Spanish does not have a Saxon genitive.

Gibson, Pearlmutter, Canseco-Gonzalez and Hickok (1996) and Gibson, Pearlmutter, and Torrens (1999) argued that the same parsing strategies are used in English and Spanish. They investigated relative clauses that could be attached to one of the three noun phrases and observed that both in English and Spanish, attachment to the most recent, third noun phrase was easiest to process, followed by an attachment to the first noun phrase, while attachment to the middle noun phrase was very hard to process. They argued that attachment to the third noun phrase is preferred as a result of recency, while attachment to the first noun phrase is relatively easy due to *predicate proximity*: a preference to attach as close to the head of a predicate phrase as possible. The latter principle is assumed to be different in strength across languages.

Mitchell et al. (1995) and Cuetos, Mitchell, and Corley (1996) argued instead that the observed cross-linguistic differences in relative clause attachment are due to differences in the frequency of occurrence of high and low attachment between languages. They suggested that low attachment is preferred in English for sentences such as (12), because across all relative clauses involving two potential attachment sites, low attachment is frequent than high attachment. In contrast, they argued that high attachment is more preferred in other languages, because in these languages, it is more frequent than low attachment. However, such a coarse-grained frequency account is inconsistent with a number of studies. First, several studies (e.g., Gilboy, Sopena, Clifton, & Frazier, 1995; Traxler et al., 1998) have shown that relative clause attachment preferences depend on the preposition in the complex noun phrase. In particular, when the preposition is *with* rather than *of* (as in 12), the preference for low attachment is much stronger. If this is due to frequency information, this indicates that the processor takes into account information from (closed class) lexical items (but see Frazier & Clifton, 1996 for a different account explaining relative clause attachment preferences). Second, Brysbaert and Mitchell (1996) showed a preference for high attachment in Dutch, even though low attachment is more frequent (Mitchell & Brysbaert, 1998). While Desmet and Gibson (2003) criticized some of the evidence against coarse-grained frequency accounts on methodological grounds (Gibson & Schutze, 1999), Desmet, De Baecke, Drieghe, Brysbaert and Vonk (in press) argued for a more fine-grained account to explain relative clause attachment preferences in Dutch. They showed that relative clause attachment preferences were affected by the animacy and concreteness of the noun phrases: Attachment to animate and concrete noun phrases was preferred to attachment to inanimate and abstract noun phrases. These online attachment preferences corresponded to relative clause attachment corpus frequencies when the animacy and concreteness of the noun phrases was taken into account (see also Desmet, Brysbaert, & De Baecke, 2002).

Other research has more directly pitted lexical frequency-based explanations against Garden-Path theory. For example, Trueswell et al. (1993) investigated sentences such as

(13), which are temporarily ambiguous because the main verb (*forgot/hoped*) can occur with a direct object (e.g., *the man forgot the solution*) or a sentence complement, which is the correct analysis.

13a. The student forgot the solution was in the book.

13b. The student hoped the solution was in the book.

Trueswell et al. compared these sentences with unambiguous sentences disambiguated by the complementizer *that* following the main verb. When the verb occurred more often with a direct object than a sentence complement (e.g., *forgot*, as measured by sentence completions), the temporarily ambiguous sentences were harder to process than unambiguous controls. However, no difference was observed when the verb occurred more frequently with a sentence complement than a direct object (e.g., *hoped*). Hence, they concluded that verb subcategorization frequencies have an immediate effect on sentence processing. Recent evidence by Hare, McRae, and Elman (2003) indicates that the semantics of the verb affects subcategorization preferences. For example, *find* occurs with a direct object when it means *locate*, whereas it occurs with a sentence complement when it means *realize*. Hare et al. showed that a context that instantiated one or the other sense of the verb affected processing difficulty at the disambiguation.

Trueswell (1996) manipulated the frequency of the verb as a past participle or past tense in reduced relative clause ambiguities such as (14).

14a. The message recorded by the secretary could not be understood.

14b. The room searched by the police could not be understood.

The verb *recorded* is most frequently used as a past participle, so this supports the reduced relative analysis, whereas *searched* is usually a past tense, consistent with the main clause analysis. Trueswell (1996) observed that the disambiguating *by*-phrase in sentences such as (14a) was harder to process than unambiguous controls, but there was no difference in sentences such as (14b), where the verb was biased toward the past participle analysis. However, this pattern of results depended on the animacy of the first noun phrase; when it was inanimate, difficulty in the ambiguous sentences occurred regardless of the verb bias, suggesting that both animacy and verb bias had to support the reduced relative analysis in order to eliminate any processing difficulty.

A number of other studies also provide evidence that lexical frequency information affects syntactic ambiguity resolution (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; Clifton et al., 1984; Garnsey et al., 1997; Holmes, Stowe, & Cupples, 1989; Mitchell & Holmes, 1985; Snedeker & Trueswell, 2004; Stowe, Tanenhaus, & Carlson, 1991). Importantly, some of these studies suggest that the effect of subcategorization frequency is very early (e.g., Garnsey et al., 1997; Snedeker & Trueswell, 2004). Clearly, this poses difficulties for Garden-Path theory or any other modular model in which the use of frequency information is delayed and supports constraint-based models or other accounts in which frequency information is used immediately during sentence processing.

However, a number of other studies suggest that the processor does not entirely rely on subcategorization frequency information. For instance, Mitchell (1987) had participants read (15) in two segments, indicated by the slash:

- 15a. After the young Londoner had arrived his parents/prepared to celebrate their anniversary.  
15b. After the young Londoner had visited his parents/prepared to celebrate their anniversary.

They spent longer reading the first fragment of (15a) than (15b). This suggests that they initially adopted the transitive analysis and experienced difficulty when they subsequently discovered that this is inconsistent with subcategorization information of the verb. In contrast, they had more difficulty reading the second fragment of (15b) than (15a), suggesting that they had already reanalyzed (15a) but not (15b). Van Gompel and Pickering (2001) found similar results using eye tracking, and therefore countered the claim that Mitchell's effects might have somehow been the results of unusual presentation conditions (Adams, Clifton, & Mitchell, 1998).

Similarly, Pickering, Traxler, and Crocker (2000) observed that readers experienced difficulty with temporarily ambiguous sentence complement clauses even though their verbs were biased toward the sentence complement analysis, and Kennison (2001) showed that temporarily ambiguous sentence complements were harder than unambiguous controls regardless of the verb bias. Finally, McKoon and Ratcliff (2003) failed to find any evidence that frequency information affected the processing of reduced relatives, and argued instead that verb complexity accounts for differences in processing difficulty for reduced relatives (though cf. McRae, Hare, & Tanenhaus, *in press*).

It is probably safest to conclude that lexical frequency information has some influence on syntactic ambiguity resolution, but in many cases does not neutralize difficulty with temporarily ambiguous sentences. This formulation is consistent with constraint-based theories. Lexical frequency information is just one of the many constraints that affect sentence processing, so even when lexical frequency information supports a particular analysis, the alternative analysis may be activated by other sources of information such as a coarse-grained frequency preference. Without knowing all the constraints that affect sentence processing and their weights, it is difficult to test and falsify constraint-based theories. One possibility is to use sentence fragment completion data to estimate the influence that the different constraints together have (e.g., Garnsey et al., 1997; McRae et al., 1998), essentially adopting the view that the sentence processor takes into account extremely fine-grained frequency information related to combinations of different words. However, it is uncertain whether such completion preferences always correctly predict online preferences in syntactic ambiguity resolution (e.g., Binder, Duffy, & Rayner, 2001; Kennison, 2001; Pickering et al., 2000).

The results from studies investigating the use of frequency information can also be explained in modular accounts which assume that the use of frequency information is delayed

(e.g., Ferreira & Henderson, 1990; Frazier, 1987a, 1995). They claim that frequency information cannot neutralize difficulty with temporarily ambiguous sentences, because people initially adopt the structurally simplest analysis, regardless of frequency biases. These models have more difficulty explaining why in some studies, difficulty was neutralized. One possibility is that the second stage of processing, during which frequency information is used, occurs very early and that current methodologies may not always be able to detect the delay in the use of frequency information (e.g., Frazier, 1995). But if this is the case, it makes such accounts extremely hard to test.

Finally, frequency effects represent a long-lasting effect of repeated exposure to syntactic structures. There is some evidence that very recent exposure to as little as one instance of a structure can affect subsequent processing of similar structures. Such effects have been demonstrated in the reading of coordinate structures (Frazier, Munn, & Clifton, 2000; Frazier, Taft, Roeper, Clifton, & Ehrlich, 1984) and in picture-expression matching (Branigan, Pickering, & McLean, 2005). Some studies have also shown behavioral and neuroscientific effects of repeated presentation of one structure (Cuetos et al., 1996; Mehler & Carey, 1967; Noppeney & Price, 2004). Additionally, Trueswell and Kim (1998) and Novick, Kim, and Trueswell (2003) found that comprehension of an ambiguous sentence was affected by the preferred analysis of a subliminally presented verb. These effects may be akin to structural priming effects in language production (Bock, 1986).

#### 4. EFFECTS OF PLAUSIBILITY

The use of semantic plausibility information during sentence processing has been an important test case for whether the sentence processor is modular or interactive. According to modular sentence processing theories such as the Garden-Path theory, plausibility information is ignored during the initial stages of sentence processing, whereas interactive theories such as constraint-based theories claim that plausibility information is used immediately.

Much of the research investigating the use of plausibility information during sentence processing has focused on the reduced relative/main clause ambiguity (Binder et al., 2001; Clifton et al., 2003; Ferreira & Clifton, 1986; McRae et al., 1998; Rayner et al., 1983; Trueswell et al., 1994). A number of studies (Clifton et al., 2003; Ferreira & Clifton, 1986; Trueswell et al., 1994) manipulated animacy, as this often provides a particularly strong plausibility constraint. These studies have compared temporarily ambiguous reduced relatives with either an animate first noun (*defendant*, 16a) or an inanimate noun (*evidence*, 16b) and contrasted them with unreduced relatives that were disambiguated by the words *that was* following the noun.

16a. The defendant examined by the lawyer turned out to be unreliable.

16b. The evidence examined by the lawyer turned out to be unreliable.

Because *examined* can have an animate or inanimate patient, *the defendant examined* (in 16a) is consistent with either the main clause or the reduced relative analysis. Many studies

have shown that the main clause analysis is initially preferred in sentences like (16a) and that reduced relatives with animate heads are harder to process than unambiguous controls (e.g., Binder et al., 2001; Britt, Perfetti, Garrod, & Rayner, 1992; Clifton et al., 2003; Ferreira & Clifton, 1986; McRae et al., 1998; Rayner et al., 1983; Trueswell et al., 1994). The crucial question is whether people process (16b) differently. Interactive models claim that people immediately use animacy information. Because *examined* does not normally have an inanimate agent, they should not adopt the main clause analysis in (16b). Therefore, if the animacy plausibility manipulation is sufficiently strong, (16b) should be no harder than its unambiguous control. In contrast, if the processor initially ignores animacy information, both (16a) and (16b) should be harder to process than their unambiguous controls.

However, the data are somewhat equivocal. In an eye-movement study, Ferreira and Clifton (1986) showed that people experienced more difficulty with reduced relatives than with unambiguous controls even if the head was inanimate, suggesting that people initially ignore plausibility information. However, Trueswell et al. (1994) argued that the plausibility manipulation in this study was not sufficiently strong and with stronger materials, they observed no evidence for processing difficulty when the head of a reduced relative was inanimate. Similar findings were reported by Mak et al. (2002), who investigated the use of animacy information in subject/object relative clause ambiguities in Dutch.

In general, it is difficult to rule out the possibility that the absence of differences in the inanimate conditions is due to a Type II error. Clifton et al. (2003) conducted an experiment similar to Trueswell et al. (1994), but used more items and employed eye-movement measures that may be more sensitive to processing difficulty. They observed that processing difficulty in reduced relatives with inanimate heads was reduced, but not completely eliminated. Similarly, using a very different type of ambiguity, Hoeks, Hendriks, Vonk, Brown, and Hagoort (in press) observed that animacy information did not completely eliminate processing difficulty.

Studies that have investigated different plausibility constraints have produced rather clearer results. Plausibility information unrelated to animacy does not appear to eliminate difficulty with reduced relatives (e.g., Rayner et al., 1983; McRae et al., 1998; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994). For example, McRae et al. (1998) showed that readers experienced difficulty at *arrested by* in (17) relative to unreduced relatives, even though the main clause analysis is implausible.

17. The crook arrested by the detective was guilty of taking bribes.

Similarly, Schriefers et al. (1995) (see also Mecklinger, Schriefers, Steinhauer, & Friederici, 1995) showed that German subject and object relatives were harder to process when they were implausible than when they were plausible (indicating that readers did use plausibility information). But more importantly, the difference in difficulty between subject and object relatives was unaffected by whether plausibility information supported the more difficult object relative analysis or not. Thus, there was no evidence that semantic plausibility affected syntactic ambiguity resolution.



Taken together, the evidence on the use of plausibility information suggests that it does not completely eliminate difficulty with the non-preferred analysis, so the findings from these studies fit fairly comfortably with models which claim that the use of semantic information is delayed during sentence processing. However, they do not necessarily provide evidence against interactive models, because plausibility may be a relatively weak constraint and therefore unable to eliminate garden-path effects.

Other methods may be more suitable for investigating the effects of semantic plausibility on sentence processing. McElree and Griffith (1995) used a speed-accuracy-trade-off method to investigate whether people detected semantic anomalies (18a) more rapidly than subcategorization violations (18b) and syntactic category violations (18c).

- 18a. Some people alarm books.
- 18b. Some people agree books.
- 18c. Some people rarely books.

Participants had to decide if the sentences made sense immediately when they heard a response tone, which occurred at various intervals after the sentence was presented. In this paradigm, participants making very rapid responses tend to perform at chance, while participants making very slow responses tend to perform as well as in an untimed task (i.e., they reach an asymptote). For very slow responses, McElree and Griffith observed similar behavior for (18a–c), indicating that they regarded the violation as equally severe in all conditions. However, participants approached the asymptote more slowly in (18a) than (18b and 18c). Differences in rate indicate differences in processing speed, so it appears that plausibility information is processed slower than subcategorization and syntactic category information.

This conclusion is also consistent with evidence from an ERP experiment by Hagoort (2003). In line with other studies, he observed that semantic anomalies elicited negative-going event-related potentials around 400 ms after the word where the sentence became anomalous (N400), while syntactic anomalies caused by grammatical gender violations resulted in positive event-related potentials at around 600 ms following the anomaly (P600) (e.g., Ainsworth-Darnell, Shulman, & Boland, 1998; Hagoort, Brown, & Groothusen, 1993; Kutas & Hillyard, 1980, 1984; Osterhout & Holcomb, 1993; Osterhout, Holcomb, & Swinney, 1994; Osterhout & Nicol, 1999; cf. Friederici, Hahne, & Mecklinger, 1996; Münte, Heinze, & Mangun, 1993 for evidence for different syntax-related effects). But most interestingly, the size of the N400 was larger when it co-occurred with a gender violation at the same word, whereas the size of the P600 was unaffected by the presence of a semantic violation. Hagoort concluded that syntactic processing affects semantic interpretation, whereas semantic interpretation does not affect syntactic processing. This is consistent with the idea that semantic processing follows syntactic processing and therefore cannot influence it.

In sum, there is fairly substantial evidence that semantic processing is delayed relative to syntactic processing. This is consistent with the predictions of modular accounts of

sentence processing, which claim that informational encapsulation makes it impossible for the syntactic processor to use information from the semantic module. Overall, the data do not fit particularly well with most interactive models. However, results showing that plausibility information does not eliminate garden-path effects may be explained by interactive theories by assuming that plausibility information is too weak to be detected. In fact, it is very difficult to rule out interactive (or constraint-based) theories as a class as a whole, because they encompass a wide variety of different models, some of which may be modified to account for the delayed effects of semantic information. Furthermore, current models are often too underspecified to derive exact predictions about the extent to which semantic information should affect syntactic processing. Hence, rather than contrasting modular and interactive models, it may be more fruitful to test specific instantiations of modular and interactive models that make very specific predictions about how semantic information is used during sentence processing.

## 5. PROSODY

A further important factor that affects sentence processing is prosody (e.g., Carlson, Clifton, & Frazier, 2001; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Pynte & Prieur, 1996; Steinhauer, Alter, & Friederici, 1999). For a complete overview of the literature on prosody, see Speer and Blodgett (this volume). Most research indicates that prosodic information is used very rapidly during sentence processing (e.g., Kjelgaard & Speer, 1999; Snedeker & Trueswell, 2003). However, evidence suggests that speakers do not always reliably provide prosodic cues (e.g., Albritton, McKoon, & Ratcliff, 1996; Snedeker & Trueswell, 2003), so the use of prosodic information may be more limited for naturally produced sentences than for sentences that are carefully constructed to test the use of prosodic information

## 6. INTEGRATION WITH CONTEXT

If the processor immediately draws on structural preferences, frequency of different analyses, or the prosody of the sentence, it bases its initial decision on information internal to the sentence itself. Such accounts make sense if the processor first has to analyze aspects of the sentence before it can relate that analysis to the wider context. But there is another possibility: Initial decisions may involve an immediate integration of the current sentence with the broader linguistic or non-linguistic context.

Mostly such proposals have focused on the linguistic or *discourse* context and have paid particular attention to its referential properties. Put simply, a context introduces one or more entities, and a target sentence is more or less felicitous depending on how it refers to those entities. In cases of ambiguity, the processor is likely to favor an analysis that refers to entities in appropriate ways (or at least, does not refer to them in inappropriate ways). In what came to be known as *referential theory*, Crain and Steedman (1985) and Altmann and Steedman (1988) formalized this account in terms of the *principle of parsimony*, which states that the processor will adopt the analysis that requires postulating the fewest unsupported pre-suppositions. For example, a definite noun phrase such as *the wife* presupposes a wife, and is

therefore supported if the linguistic or non-linguistic context includes a wife, but is unsupported otherwise. Crain and Steedman considered the more complex example in (19):

19a. The psychologist told the wife that he was having trouble with her husband.

19b. The psychologist told the wife that he was having trouble with to leave her husband.

These sentences are temporarily ambiguous, because *that he was having trouble with* can be the start of a complement clause of *told*, as in (19a), or a relative clause modifying *the wife*, as in (19b). Sentence (19a) includes the simple noun phrase *the wife*, and therefore presupposes one wife; but (19b) includes the complex noun phrase *the wife that he was having trouble with*, and therefore presupposes at least two wives (one of whom he is having trouble with, and at least one other).

Garden-Path theory claims that readers should prefer (19a) over (19b) because the complement clause analysis is less syntactically complex than the relative clause analysis (i.e., by minimal attachment). Crain and Steedman (1985) made the same prediction when the sentences are presented in isolation, because (19a) requires postulating one wife, whereas (19b) requires postulating more than one wife. They also made the same prediction in a context mentioning one wife, because (19a) is felicitous but (19b) is not. But if the context mentioned more than one wife, the relative clause analysis would not require postulating any unsupported presuppositions. In contrast, the complement clause analysis becomes less preferred, because it is unclear which wife is referred to. In sum, referential theory predicts that (19b) should be easier to process than (19a) in the absence of a context or a context mentioning one wife, but this pattern should be reversed when the context mentions two wives. Crain and Steedman provided initial experimental evidence in favor of this account.

Many studies have tested the predictions of the referential theory and contrasted them with those of the Garden-Path theory, which claims that discourse information is ignored during initial processing. Some have shown evidence for immediate effects of discourse context (e.g., Altmann & Steedman, 1988; Altmann, Garnham, & Dennis, 1992; Altmann, Garnham, & Henstra, 1994; Van Berkum, Brown, & Hagoort, 1999), while others failed to find evidence for immediate referential effects (e.g., Mitchell, Corley, & Garnham, 1992; Murray & Liversedge, 1994; Rayner, Garrod, & Perfetti, 1992). Generally, it appears that referential contexts can override syntactic ambiguity resolution preferences (that exist in the absence of a context) when the preference for one of the syntactic analyses is relatively weak, but fails to do this when there is a strong bias for one of the analyses (Britt, 1994; Britt et al., 1992; Spivey & Tanenhaus, 1998).

Such findings fit well with constraint-based theories, which claim that discourse context provides just one of the many constraints that are used during syntactic ambiguity resolution. They claim that contextual information is used immediately, but does completely neutralize preferences that occur in the absence of a context if other constraints provide strongly biasing information (e.g., MacDonald et al., 1994; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995). This is because contextual information is only

one of the many sources of information that affect sentence processing. If other sources of information (e.g., frequency) strongly support the analysis that is inconsistent with the context, contextual information may not be sufficiently strong to override this preference.

Context may not only be provided by the preceding linguistic context, but also by a visual context. Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) and Spivey, Tanenhaus, Eberhard, and Sedivy (2002) argued that readers may not always immediately use a linguistic discourse context because they need to retain the linguistic context in memory and it may often be difficult to retrieve this contextual information from memory. In contrast, the effects of a visual context may be stronger, because the objects in the scene are present while the ambiguous sentence unfolds. Indeed, Tanenhaus et al. (1995), Chambers, Tanenhaus, and Magnuson (2004), and Spivey et al. (2002) found very strong effects of context on participants' eye movements to objects in an array when they followed auditory instructions containing syntactic ambiguities (though cf. Snedeker & Trueswell, 2004). For more details of the "visual world" paradigm, see Tanenhaus and Trueswell (this volume).

So far, we have considered how context affects the processing of temporarily ambiguous sentences that contain either an unmodified or a modified noun phrase. However, more recent studies have asked whether context plays any other role in parsing. For example, Hoeks, Vonk, and Schriefers (2002) had people read (Dutch translations of) sentences such as (23) in isolation or following the context (22):

22. When they met the fashion designer at the party, the model and the photographer were very enthusiastic.
23. The model embraced the designer and the photographer opened smilingly a bottle of champagne.

Sentence (23) was difficult in isolation (versus sentences disambiguated by a comma following *designer*). People appeared to treat *the designer and the photographer* as the object of *embraced*, and therefore adopted the incorrect NP-coordination analysis. Hoeks et al. argued that the NP-coordination analysis involves the accommodation of a single topic (the subject *the model*) about which new information is predicated. In contrast, the correct S-coordination analysis for (23) requires the accommodation of two topics, namely the subject of the first clause *the model* and the subject of the second clause *the photographer*, and is therefore argued to be more complex.

But this difficulty disappeared when (23) followed (22), suggesting that people immediately adopted the correct S-coordination analysis. Hoeks et al. (2002) argued that (22) introduced two topics and therefore removed the difficulty with S-coordination. They proposed a principle of minimal topic structure to explain such findings, which can be seen as a generalization of the principle of parsimony. The results of course provide evidence against Garden Path theory.

Results by Liversedge, Pickering, Branigan, and Van Gompel (1998) and Altmann, van Nieuwland, Garnham, and Henstra (1998) suggest that direct and indirect questions also affect

the processing of subsequent sentences in important ways. Liversedge et al. (1998) found that locative sentences such as (24a) took longer to process than agentive sentences such as (24b):

24a. The shrubs were planted by the apprentice that morning.

24b. The shrubs were planted by the greenhouse that morning.

But when they were preceded by an indirect question such as (25) that supported an expectation for a locative, the difference in reading times disappeared:

25. The gardener wondered where to plant the shrubs.

Hence, Liversedge et al. argued that thematic role expectations induced by the context affect syntactic ambiguity resolution (see also Liversedge, Pickering, Clayes, & Branigan, 2003).

Similarly, Altmann et al. (1998) showed that the discourse context can create specific expectations about temporal phrases. They compared sentences such as (26a), where the adverbial *next week* is attached to the first VP containing *implement*, with sentences such as (26b), where the adverbial *next week* is attached to the first VP containing *proposed*. In a neutral context, (26a) was harder to process than (26b).

26a. She'll implement the plan she proposed to the committee next week, they hope.

26b. She'll implement the plan she proposed to the committee last week, they hope.

However, this pattern was reversed when the sentences were preceded by an indirect question such as (27), which creates an expectation for a temporal phrase modifying the first VP.

27. The committee members wonder when Fiona will implement the plan she proposed.

A similar pattern of results was observed in an experiment that used direct questions. Thus, the Liversedge et al. (1998) and Altmann et al. (1998) studies suggest that indirect questions can create expectations that affect the processing of the subsequent sentence.

An issue of particular interest to functional linguistic theories (e.g., Givón, 1984, 1990; Lambrecht, 1994) is how discourse properties interact with different word orders, especially in languages that have a flexible word order. These theories generally assume that different word orders represent ways of expressing different information structures, such as whether a particular noun phrase provides information that is given or new in the discourse. They predict that non-canonical structures should be hard to process out of context, because they require a specific context that licenses them. But in the appropriate context, such non-canonical structures may be no harder to process than canonical structures.

This prediction was tested in a study by Kaiser and Trueswell (2004), who contrasted canonical subject-verb-object (SVO) word order in Finnish with non-canonical object-verb-subject (OVS) word order. As shown by Hyönä and Hujanen (1997), OVS order is harder to process than SVO order outside a discourse context. Kaiser and Trueswell (2004) suggested that this difficulty may be eliminated when OVS sentences are presented in an appropriate context. They argued that in OVS sentences, the object represents new information that has not yet been introduced in the discourse, while the subject represents given information. In contrast, the subject and object may be either given or new in the SVO sentences. In a reading experiment, Kaiser and Trueswell showed that the OVS sentences were facilitated when the object was introduced in the preceding context, but this context did not completely neutralize the difficulty with OVS sentences. Obviously, one possibility is that this difficulty occurs because the OVS structure is hard to adopt despite the supporting contextual information. However, Kaiser and Trueswell (2004) suggested an alternative explanation. They argued that in the OVS order, people anticipate that the postverbal noun is a new discourse entity, and that processing difficulty for OVS sentences in a supporting context occurs because the anticipation incurs a processing cost. In contrast, they do not anticipate any discourse properties of the postverbal noun in the SVO order, because it may be either new or given. They showed evidence for their hypothesis in a visual-world eye-movement experiment, which showed that before they heard the postverbal noun, participants looked more often at a new referent when they heard an OVS sentence than when they heard an SVO sentence. This suggests that the difficulty in the OVS sentences when presented in an appropriate context may have been due to the anticipation of a new discourse entity.

The view that contextual information has a rapid effect on the processing of word order ambiguities is also supported by a study by Knoeferle, Crocker, Scheepers, and Pickering (2005), who investigated SVO/OVS ambiguities in German in a visual-world eye-movement study. In their experiments, listeners saw pictures of actions that were either consistent or inconsistent with the SVO or OVS word order. Knoeferle et al. found that this visual disambiguating information rapidly influenced whether listeners anticipated a subject or an object following the verb. It therefore had an early influence on whether people adopted the SVO or OVS structure.

In addition, there is a clear evidence that people use both linguistic and visual information about referential context during sentence processing. Furthermore, for some ambiguities, linguistic and visual context appear to be able to override syntactic preferences that exist in the absence of a context. These effects emerge at the earliest stages of processing that can be measured with current methodologies. However, for other ambiguities, usually those where the preference in the absence of the context is stronger, context does not completely eliminate preferences that exist for ambiguities out of context. This is consistent with constraint-based models, which assume that context has an immediate effect on sentence processing, but that other sources of information may sometimes be strong enough to override these effects.

## 7. WHY DOES PROCESSING DIFFICULTY OCCUR?

A different issue that distinguishes two-stage and constraint-based theories is whether processing difficulty is due to reanalysis or competition. Two-stage theories usually assume that processing difficulty during syntactic ambiguity resolution is due to the detection of a misanalysis and subsequent reanalysis of the initially adopted structure (e.g., Ferreira & Henderson, 1991; Frazier & Rayner, 1982; Pritchett, 1992; Rayner et al., 1983). In contrast, most constraint-based theories assume that processing difficulty occurs when two syntactic analyses receive approximately equal support from the various sources of information and compete (e.g., MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Tabor & Tanenhaus, 1999). Most studies showing evidence for processing difficulty during syntactic ambiguity resolution are consistent with either assumption (Van Gompel, Pickering, & Traxler, 2001). For example, competition may occur at *arrested* in (28) because plausibility information supports the main clause analysis, whereas frequency information or perhaps structural preferences support the main clause analysis (e.g., McRae et al., 1998).

28. The burglar arrested by the cop was in court.

Alternatively, difficulty may occur because the processor initially adopts the main clause analysis, but semantic information is inconsistent with this, so the processor is forced to reanalyze.

In order to discriminate between reanalysis and competition as the mechanism for explaining processing difficulty, Traxler et al. (1998), Van Gompel et al. (2001), and Van Gompel, Pickering, Pearson, and Liversedge (2005) compared the processing of globally ambiguous sentences with that of disambiguated sentences. Using various types of ambiguities, they showed that globally ambiguous sentences (e.g., (29a)) were easier to process than sentences that were disambiguated toward either analysis (e.g., 29b and 29c).

29a. I read that the bodyguard of the governor retiring after the troubles is very rich.

29b. I read that the governor of the province retiring after the troubles is very rich.

29c. I read that the province of the governor retiring after the troubles is very rich.

Constraint-based theories claim that semantic plausibility should have an immediate influence on the extent to which syntactic analyses compete, so competition in globally ambiguous sentences (29a) should be stronger than in disambiguated sentences such as (29b and 29c), where plausibility information supports only a single analysis. To explain why globally ambiguous sentences are easier than disambiguated sentences, competition models would have to assume that plausibility information is delayed and therefore cannot affect competition processes. However, this explanation seems clearly incompatible with current constraint-based theories, which assume that all sources of information are used immediately during syntactic ambiguity resolution. Furthermore, Van Gompel et al. (2005) observed that globally ambiguous sentences were no harder to process than syntactically unambiguous sentences such as (30).

30. I read quite recently that the governor retiring after the troubles is very rich.

This finding is hard to explain for competition models, regardless of whether they assume that the use of plausibility information is delayed. They predict no competition when only a single analysis is syntactically possible, so (30) should be harder easier to process than (29a).

Traxler et al. (1998) and Van Gompel et al. (2001, 2005) explained their results using a probabilistic reanalysis model. They argued that for ambiguities such as (29a), where both analyses are approximately equally preferred, the processor initially adopts each analysis about half the time. When this analysis is inconsistent with later disambiguating information (e.g., semantic information), reanalysis occurs. Hence, readers have to reanalyze about half the time in sentences that are disambiguated toward one analysis, and about half the time in sentences that are disambiguated toward the other analysis, so both disambiguated sentences are relatively hard to process. In contrast, the initial analysis is always plausible in globally ambiguous and unambiguous sentences, so no reanalysis occurs.

## 8. REANALYSIS

So far, we have focused on the traditional question of the initial stages of parsing, by focusing on the question of how people make initial decisions. Such research has tended to use garden-path sentence in which the initially favored analysis turns out to be incorrect, and assumes that people then at least try to reanalyze. For a long time, there was surprisingly little research into how reanalysis takes place. However, there are a number of strands of research into reanalysis, and this section attempts to bring them together.

Given the evidence that reanalysis tends to be difficult, people might only reanalyze when there is no grammatically possible continuation. Indeed, such an assumption was made in the early days of the “Garden-Path” model, with J. D. Fodor and Frazier (1980) proposing that the analysis is “not to be changed in response to subsequent words unless there is no other way of proceeding.” (p. 427). They called this the *Revision-as-Last-Resort (RALR)* principle. It follows from a more general principle whereby the processor makes as few changes to the current representation as possible, because this minimizes processing cost, and accords with the principle of Minimal Attachment, which minimizes processing cost during initial processing.

If “no other way of proceeding” is taken to mean no grammatical way of proceeding, then RALR is almost certainly incorrect. Many studies of initial processing use plausibility to disambiguate, and assume that people abandon a grammatical but sufficiently implausible analysis. Indeed, Rayner et al. (1983) found difficulty with (3), repeated as (31):

31. The spy saw the cop with a revolver but the cop didn’t see him.

They claimed that people initially adopted the high (VP-attached) analysis, but reanalyzed when they realized that this meant that the spy used the revolver to see the cop. This



analysis is grammatical but implausible, and the implication is that Garden-Path theory assumes that plausibility can lead to reanalysis. In accord with this, people have more difficulty abandoning a plausible analysis than an implausible analysis, presumably because their degree of commitment to an analysis is affected by its plausibility (Pickering & Traxler, 1998). Interactive accounts of course assume that plausibility can affect all stages in the process of choosing between analyses (e.g., MacDonald et al., 1994; Trueswell et al., 1994).

A more controversial question is whether the processor ever abandons grammatical and plausible (and reasonably common) analyses. Sturt, Pickering, Scheepers, and Crocker (2001) and Schneider and Phillips (2001) report experiments that suggest that it does not tend to do so. Sturt et al. used sentences like (33):

33a. The troops who found the enemy spy had shot himself were later mentioned in the press report.

33b. The troops who found the enemy spy had shot themselves and were later mentioned in the press report.

Because *found* strongly prefers a noun phrase object, all theories assume that people initially adopt that analysis. When they encounter *had shot*, they can treat *the troops who found the enemy spy* as its subject. This is compatible with revision being a last resort, but it requires the processor to adopt an analysis that is disfavored, primarily because any kind of recency (or locality) principle would favor low attachment, because a verb (*found*) occurs between the high and low attachment sites (*troops* and *spy*, respectively). Alternatively, they can reanalyze and treat *the enemy spy* as the subject of *had shot*. Participants had more difficulty with (33a) than (33b), indicating that they did not reanalyze, even though other principles clearly favor high attachment and reanalysis is comparatively easy for this ambiguity (as discussed above). These findings, together with those of Schneider and Phillips, suggest that people may not abandon plausible analyses.

## 9. HOW DIFFICULT IS REANALYSIS?

Most parsing research makes use of “difficult” ambiguous sentences that cause subjectively clear garden-path effects, such as the classic reduced relatives (Bever, 1970). This may be because such sentences make a good didactic point, because their difficulty makes researchers notice them in the first place, or because they tend to lead to statistically significant results. In fact, some ambiguous sentences cause much less processing difficulty than others, and one important reason for this appears to be that some kinds of reanalysis are more straightforward than others. Why might this be?

Rather unusually for psycholinguistics, many of the claims regarding difficulty of reanalysis have been based on subjective judgments. In the long run, this is clearly unsatisfactory, but it constitutes a reasonable place to start, and experimental work has made some attempt to test these intuitions. Specifically, the theoretical enterprise has been

concerned with the difficulty of particular kinds of structural change, with the goal of determining why converting initial representations into final representations can be more or less difficult. Contrast (34a) and (34b) below (from Sturt, Pickering, & Crocker, 1999):

34a. The Australian woman saw the famous doctor had been drinking quite a lot.

34b. Before the woman visited the famous doctor had been drinking quite a lot.

In both the cases, *the famous doctor* is initially treated as the object of the preceding verb (*saw/visited*), and is eventually reanalyzed to be the subject of the following verb (*drinking*). Although these operations might appear similar, they hide an important difference. In (34a), *the famous doctor had been drinking a lot* is a complement clause licensed by *saw*, and so *the famous doctor* is in some sense still subordinate to *saw*. But in (34b), *visited* is intransitive, and forms part of the subordinate clause itself, whereas *the famous doctor* forms part of the main clause. It is therefore not subordinate to *visited*.

Most theories of reanalysis difficulty assume that some set of sentences including (34b) cause more difficulty, whereas another set including (34a) cause less difficulty. The first important account was due to Pritchett (1988, 1992), who claimed that reanalysis is hard when it requires a constituent to be moved out of its thematic domain (corresponding roughly to the arguments of a verb or other argument assigner). In other accounts, reanalysis is difficult if it involves abandoning the structural relations that describe tree structure (Gorrell, 1995; Sturt & Crocker, 1996), with the underlying intuition being that the processor prefers to preserve representations where possible. This is of course very similar to Fodor and Frazier's principle of minimal revisions, discussed above, and with the experimental data of Sturt et al. (2001) and Schneider and Phillips (2001). Whatever the precise account that explains the representations used in analysis and reanalysis, there are good reasons to assume that the processor makes strident attempts to avoid difficult reanalysis.

In fact, some theories go so far as to assume that sentences like (34a) do not cause difficulty. However, this is a case where intuitions are certainly wrong, because many well-controlled experimental studies have shown difficulty with (34a) (Garnsey et al., 1997; Rayner & Frazier, 1987; Pickering & Traxler, 1998; Pickering et al., 2000; Sturt et al., 1999; Trueswell et al., 1993). Indeed, Sturt et al. (1999) contrasted (34a) and (34b) above, in experiments in which the relative plausibility and verb biases were carefully controlled. In self-paced reading, participants experienced more difficulty with *had been drinking* in (34b) than (34a), though (34a) did cause clear processing difficulty as well.

We might expect that the effects of verb bias on reanalysis difficulty would have been systematically investigated, and indeed there is some evidence that degree of verb bias affects processing difficulty (Garnsey et al., 1997; Mitchell & Holmes, 1985; Trueswell et al., 1993). However, it is extremely hard to determine whether it affects choice of initial analysis, as constraint-based theories tend to assume, or reanalysis difficulty, as Garden-Path theory assumes. Indeed, research specifically concerned with reanalysis has focused on cases where the misanalysis is not in reasonable doubt. The clearest case of

this concerns the characteristics of the “ambiguous” region. From the earliest days of parsing research, a common intuition is that the longer a misanalysis is retained, the harder it should be to abandon in favor of the correct analysis (Frazier & Rayner, 1982; Warner & Glass, 1987). Using speeded grammaticality judgment, Ferreira and Henderson (1991) claimed that such effects were largely due to the position of the head noun, with participants judging (35b) ungrammatical more often than (35a):

- 35a. After the boy scratched the big and hairy dog yawned loudly.  
 35b. After the boy scratched the dog that is big and hairy yawned loudly.

Participants may strongly commit to an analysis when they encounter the relevant head constituents, and therefore commit to the initial analysis for longer when the head is early than when it is late. However, it is also possible that reanalysis difficulty relates in part to the grammatical complexity of the ambiguous region, with the presence of the embedded clause in (35b) contributing to reanalysis difficulty. Although comparable effects have not always been found in reading-time studies (Ferreira & Henderson, 1993; Sturt et al., 1999), Tabor and Hutchins (2004) did replicate the pattern in self-paced reading. Whereas the precise characteristics of the ambiguous region that affect reanalysis difficulty may still be in question, there is little doubt that syntactic and semantic factors are important.

## 10. DEALING WITH AMBIGUITY IN REANALYSIS

Most research on reanalysis has simply assumed that people know what to do when they realize they have misanalyzed: They must attempt to construct the correct analysis. But sometimes it is not this straightforward, because the processor may still be faced with more than one alternative. What does the processor do then? Does it simply use whatever strategies it would have used if it had been faced with the same ambiguity during initial processing? Although this might seem most straightforward, proponents of modular accounts often argue that syntax-based principles are used because such information becomes available before other sources of information. Alternatively, reanalysis might be able to draw on all relevant information, and would therefore operate according to different principles from initial analysis. Additionally, it might pay attention to the relative difficulty of converting from the initial analysis to each new analysis.

Unfortunately, there has been little work on these issues, perhaps because the relevant sentences and their controls are quite complex and hard to construct. Sturt, Scheepers, and Pickering (2002) showed that participants experienced difficulty with (36a), relative to disambiguated controls:

- 36a. The photographers found the countess who heard the choirboy had really enjoyed herself at the concert in the town hall.  
 36b. The photographers found the countess who heard the choirboy had really enjoyed himself at the concert in the town hall.

Before encountering *had*, both *the countess* and *the choirboy* are presumably treated as the objects of the preceding verbs. But at *had*, it becomes clear that either *found* takes a complement clause (*found the countess ... had*) or *heard* does (*heard the choirboy had*). Two experiments showed difficulty following the reflexive pronoun in (36a) but not (36b), which implied that participants had assumed that *heard* took the complement clause. This finding suggests that reanalysis follows a recency-based principle like late closure, with the new words *had really enjoyed* being attached into recently processed material (*heard the choirboy*). This provides some reason to believe that reanalysis may follow comparable principles to initial analysis.

## 11. RETENTION OF ABANDONED ANALYSES

Traditional theories assume, implicitly or explicitly, that abandoned analyses are simply discarded. However, Christianson, Hollingworth, Halliwell, and Ferreira (2001) presented a major challenge to this view. They presented participants with transitive/intransitive ambiguities such as (37).

37. While the man hunted the deer that was brown and graceful ran into the woods.

When subsequently asked *Did the man hunt the deer?*, they tended to answer *yes*. This might have been due to failure to adopt the correct analysis, but participants also tended to answer *yes* to *Did the deer run into the woods?*, which suggests that they did reanalyze. It is possible that participants did reanalyze and then simply inferred that the man hunted the deer. Indeed, they were more likely to respond “yes” after (37) than after a condition that ended *paced in the zoo* (in which case it is unlikely that someone hunted the deer). But participants were more likely to give the incorrect answer after (37) than when the clauses were reversed, which does not change meaning but does prevent the misanalysis. Second, incorrect answers were more common with the difficult garden path in (37) than the easier garden path employing the semantically equivalent phrase *the brown and graceful deer* (cf. Ferreira & Henderson, 1991). Perhaps most importantly, Christianson et al. contrasted the conditions in (38).

38a. While Anna dressed the baby that was small and cute spit up on the bed.

38b. While Anna dressed, the baby that was small and cute spit up on the bed.

Here, the control sentence (38b) differs from (38a) only in the presence of the comma after *dressed*. Additionally, intransitive interpretations of *dressed* and other “reflexive absolute transitive” verbs are inconsistent with transitive interpretations: *Anna dressed* must mean that Anna dressed herself. Hence participants could not be adopting the intransitive analysis of (38a) and inferring that Anna must be dressing the baby. However, participants still answered *yes* to *Did Anna dress the baby?* more often after (38a) than (38b).

The study employed an explicit measure of comprehension, which is potentially subject to strategic effects or biases. Christianson et al.’s (2001) account assumes that participants

produced a “final” representation of the target sentence before they answered the question, but it may be that the question affects the processing of the target sentence. However, there is reading-time evidence for the retention of analyses during learning (Kaschak & Glenberg, 2004), and some suggestion that abandoned analyses can serve as primes for subsequent acts of sentence production (Van Gompel, Pickering, Pearson, & Jacob, in press). These results suggest that some trace of the initial analysis can be retained alongside the correct analysis, but there are different possible explanations of this. Christianson et al. assume that people fail to reanalyze fully, and may even construct an ungrammatical representation after reanalysis that combines elements of the two analyses. But it is also possible that the two analyses simply remain activated in parallel. In any case, it would be unwise to assume that abandoned analyses can have no further role in processing.

## 12. STRUCTURAL COMPLEXITY AND MEMORY LOAD

Starting with the first sentence processing theories, it has often been assumed that working memory plays an important role in sentence processing and that people’s working memory limitations affect the ease with which sentences are processed (e.g., Caplan & Waters, 1999; Chomsky & Miller, 1963; Kimball, 1973; Gibson, 1991, 1998; Just & Carpenter, 1992; Lewis, 1996; MacDonald & Christiansen, 2002; Miller & Chomsky, 1963; Stabler, 1994; Wanner & Maratsos, 1978; Yngve, 1960). In contrast to most other sentence processing research, much of this work has investigated largely unambiguous sentences.

One strand of research has focused on whether all linguistic processes share the same pool of working memory resources (e.g., Just & Carpenter, 1992) or whether working memory resources dedicated to sentence processing are different from those used for other, more conscious verbal tasks (e.g., Waters & Caplan, 1996; Caplan & Waters, 1999). According to the *shared resources account*, individual differences in working memory capacity as assessed by the reading span test (e.g., Daneman & Carpenter, 1980) should affect sentence processing. In contrast, the *dedicated resources account* claims that the working memory resources used for sentence processing and the reading span test are different, so this test should not predict sentence processing effects. Caplan and Waters (this volume) describe the literature testing these contrasting accounts in detail. Although a number of studies seem to suggest that sentence processing is influenced by people’s working memory span as determined by the span test (e.g., Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992), the interpretation of these studies is controversial (Waters & Caplan, 1996), and many of the effects have proven difficult to replicate (e.g., Caplan & Waters, 1999; Traxler, Williams, Blozis, & Morris, 2005).

Another strand of research has developed models of how the structural complexity of sentences is determined by people’s working memory capacity. Since the beginning of modern psycholinguistics, researchers have attempted to determine why some nested or center-embedded sentences cause processing breakdown (e.g., Chomsky & Miller, 1963; Yngve, 1960), and why some largely unambiguous sentences such as object relatives

tend to be harder to process than others such as subject relatives (e.g., Wanner & Maratsos, 1978). There has now been a resurgence of interest in this area, largely as a result of Gibson (1998), which provided an account of processing complexity that at the same time sought to explain some issues in ambiguity resolution. He proposed the syntactic prediction locality theory (SPLT), which claims that two factors contribute to sentence complexity, storage costs, and integration costs. Both draw on the same pool of working memory resources (cf. Just & Carpenter, 1992).

Storage costs occur when there is a dependency between two syntactic elements in a sentence and the first element has to be stored in memory before it can be integrated with the later element. Integration costs occur when this integration occurs and a syntactic prediction is satisfied. For example, the SPLT predicts that storage costs occur between a *wh*-constituent (a filler) and its trace position, and that integration costs occur at the trace position. The SPLT claims that both storage and integration costs are influenced by locality, with locality being defined in terms of the number of new discourse referents that is being processed: Both storage and integration costs increase the more new discourse referents that have been processed since the prediction of a syntactic dependency is made at the first linguistic element.

Gibson (1998) argued that syntactic ambiguity preferences in long-distance dependencies can be explained as locality preferences that occur because the processor attempts to minimize storage costs. The processor favors positing a gap in sentences such as (9) above as early as possible, because this reduces storage costs at the following words. Thus, the SPLT provides a memory cost motivation for strategies predicting early gap filling such as the active-filler strategy (e.g., Frazier & Flores d'Arcais, 1989; De Vincenzi, 1991). Furthermore, the locality component of the SPLT also accounts for recency effects in attachment ambiguities such as (8) above: It predicts that integration costs are larger when there is a dependency between two distant syntactic elements than between two local elements, thus providing an independent motivation for recency preferences.

However, the SPLT's main contribution is in explaining processing cost in (largely) unambiguous sentences containing syntactic dependencies. Experiments have shown evidence for both storage and integration costs, as predicted by the SPLT. For example, Chen, Gibson, and Wolf (2005) investigated whether storage costs occurred in dependencies between an NP (e.g., *the knowledge* in (41a)) and a subsequent verb (e.g., *came*).

- 41a. The detective suspected that the knowledge that the guard protected the jewels came from an insider.
- 41b. The detective suspected that the thief knew that the guard protected the jewels and so he reported immediately to the museum curator.

They observed that reading times for the intervening region were longer than in sentences such as (41b), where *knew* does not predict a dependency with a later verb. Further evidence for storage costs comes from ERP studies, which show a sustained negativity in

the ERP signal between a *wh*-constituent and its subsequent trace position (e.g., Fiebach, Schlesewsky, & Friederici, 2002; Phillips, Kazanina, & Abada, 2005). It is assumed that this sustained negativity reflects memory costs involved in maintaining *wh*-dependencies.

ERP studies have also provided evidence for integration costs. Kaan, Harris, Gibson, and Holcomb (2000) tested *wh*-dependencies and observed a P600 effect at the verb where the *wh*-constituent had to be integrated relative to a condition where there was no *wh*-dependency. Similar P600 effects have been observed in other studies (e.g., Fiebach et al., 2002; Phillips et al., 2005).

Much of the research on storage and integration costs has focussed on the processing of subject and object relative clauses (Gibson, 1998; Grodner & Gibson, 2005; Hsiao & Gibson, 2003; Warren & Gibson, 2002). Gibson (1998) argued that object relative clauses such as in (39) above are harder to process than subject relative clauses because both storage and integration costs are higher in object relatives. In (39a), the relative pronoun *who* has to be stored in memory across a new referent (*the senator*) before it can be integrated with its trace following the verb *attacked*, whereas the *wh*-dependency does not cross any new referents in (39b). This should result in additional storage costs during the region *the senator* and additional integration costs at *attacked* in the object relative clause compared to the subject relative clause.

An important claim of the SPLT is that storage and integration costs are influenced by the number of new discourse referents that a syntactic dependency crosses. Warren and Gibson (2002) tested this in a series of complexity-rating experiments that varied the type of referring expression between the relative pronoun and the relative clause verb in object relatives. Complexity of object relatives was lowest for indexical pronouns such as *I* and *you*, which refer to discourse entities that are highly accessible (Ariel, 1990). In contrast, complexity was high for indefinite and definite NPs, which tend to refer to discourse entities that are relatively inaccessible. Hence, Warren and Gibson (2002) concluded that sentence complexity is higher when a dependency crosses a referring expression that requires more processing effort because it refers to a less accessible discourse entity.

This account contrasts with that of Gordon and colleagues (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, & Levine, 2002), who claimed that the effect of different types of referring expressions in subject and object relative clauses is due to interference. When the head NP (e.g., *the reporter* in (39b)) and the embedded NP (e.g., *the senator*) are of the same type (e.g., both are common nouns or proper names), this results in interference when the NPs have to be retained in memory during processing. Furthermore, interference is larger when the syntactic structure is hard to process, as in object relatives. This correctly accounts for Warren and Gibson's (2002) finding that object relatives are rated as more complex in sentences such as (39), where both NPs are common nouns than in sentences where the embedded NP is an indexical pronoun. It also explains why difficulty with object relatives is reduced when one NP is a proper noun and the other a common noun, as shown by Gordon et al. (2001, 2004). The SPLT has difficulty in explaining some of the results of the experiments conducted by Gordon and colleagues.

In particular, Gordon et al. (2001) observed that both (42a) and (42b) were facilitated relative to conditions, where both NPs were either common nouns or proper names.

42a. It was the barber that Bill saw in the parking lot.

42b. It was John that the lawyer saw in the parking lot.

The SPLT explains why (42a) is easier to process than object relatives with two definite NPs, because proper names are more accessible than definite NPs. However, it does not explain why (42a) is easier to process than an object relative with two proper names, as in both sentences, the NP that crosses the dependency is the same (*Bill*). It also does not explain why (42b) is easier to process than a sentence with two definite NPs or two proper nouns.

In a subsequent series of experiments, Gordon et al. (2004) showed that the effects of different types of referring expressions cannot be due to the frequency with which particular referring expressions are used as either the subject or object in relative clauses. Their corpus study showed that indefinite (e.g., *a barber*) and generic NPs (e.g., *barbers*) occur much more often as objects than subjects in relative clauses, whereas there was little difference for definite NPs. But despite this, processing difficulty in subject and object relatives was unaffected by whether the embedded NP was indefinite, generic, or definite.

However, neither the SPLT nor the interference-based account is likely to provide a complete explanation of subject and object relative clause processing. Traxler, Morris, and Seely (2002) showed that both semantic plausibility and animacy affect relative clause processing, although they do not affect initial processing. They investigated sentences where only one of the NPs was plausible as a subject and only one was plausible as the object of the relative clause verb, and contrasted them with sentences similar to (39), where both NPs are equally plausible as either the subject or object. Eye-movement measures reflecting early processing showed that object relative clauses were harder to read than subject relative clauses, but this difference was unaffected by plausibility. However, plausibility affected processing in later measures (see also Traxler et al., 2005). In another experiment, which manipulated the animacy of the NPs, they found that object relatives were harder to read than subject relatives when the subject was animate and the object inanimate, but difficulty with the object relatives was reduced when the animacy of the grammatical roles was reversed (cf. Traxler et al., 2005). Hence, it appears that semantic information affects relative clause processing, though its use may be somewhat delayed.

### 13. DO PEOPLE CONSTRUCT UNGRAMMATICAL REPRESENTATIONS?

The great majority of parsing research has assumed that people only construct grammatical representations that are consistent with the rules of grammar. Most research has simply assumed that people only construct grammatical representations, and has attempted to determine which grammatical analysis is adopted or favored at different



points in processing. The possibility that people construct ungrammatical representations would require psycholinguists to re-think their models and accounts of data, for example, considering the possibility that apparently ungrammatical analyses can interfere with grammatical ones during initial selection or reanalysis.

However, there are a few, perhaps unrelated, pieces of evidence suggesting that people may compute ungrammatical representations. First, Gibson and Thomas (1999) found that people actually preferred certain ungrammatical over grammatical sentences involving center embedding. Second, Christianson et al.'s (2001) investigation of reanalysis suggested that people do not fully abandon analyses that are shown to be incorrect. In one sense, this means that ungrammatical analyses may be retained, but of course it is also possible to argue that what is retained is a trace of the analysis before it was shown to be impossible (cf. Kaschak & Glenberg, 2004). However, Christianson et al. considered an alternative account in which people construct an ungrammatical representation, for example, one where *the deer that was brown and graceful* in (43) serves both as the subject of *ran* (as it "should") and as the object of *hunted* (see J. D. Fodor & Inoue, 1998).

Tabor, Galantucci, and Richardson (2004) found that people sometimes consider locally possible analyses even if the global syntactic context rules them out. They contrasted reduced relatives like (43a and 43b) with unambiguous controls (including *who was* or *that was*):

- 43a. The bandit worried about the prisoner transported the whole way by the capricious guards.  
 43b. The bandit worried about the gold transported the whole way by the capricious guards.

The verb *transported* must form part of a reduced relative in both sentences. If the parser only constructs grammatical analyses, it should not consider treating *transported* as a main clause verb. However, readers experienced more difficulty with *transported* in (43a) than (43b), because the ungrammatical main clause analysis was more plausible in (43a) (prisoner transporting something) than (43b) (gold transporting something). It appears as though the plausible ungrammatical analysis interfered with the selection of the correct analysis. The processor appears to construct analyses bottom-up, with *transported* being interpreted with respect to the previous noun phrase, even though such an analysis is globally ungrammatical.

Both Christianson et al. (2001) and Tabor et al. (2004) suggest that the processor considers ungrammatical analyses, but in both cases, these analyses may be "by-products" of the application of procedures consistent with grammar that are used during syntactic ambiguity resolution. However, other studies provide some evidence that the processor may consider ungrammatical analyses in unambiguous sentences. From the 1960s onwards, it has been known that passives can be harder to process than actives (e.g., Slobin, 1966; Forster & Olbrei, 1973; Herriot, 1969) and that object relatives and clefts can be harder than subject relatives and clefts (see Townsend & Bever, 2001). Sometimes these

effects seem to be restricted to cases where the sentences are “reversible” (e.g., *the woman was visited by the man*), but sometimes they also occur for non-reversible sentences (e.g., *the cheese was eaten by the mouse*); see Ferreira (2003).

In either case, these results are compatible with an account in which the processor sometimes computes ungrammatical analyses, either before computing the correct analysis or in parallel with it. Ferreira (2003) and Townsend and Bever (2001) suggest that, as well as parsing using the grammar, the processor uses simple heuristics, in particular the NVN strategy, whereby the parser interprets the first noun phrase as the agent and the second noun phrase as the patient, at least in English. The processor should be particularly prone to adopting this strategy when it is supported by plausibility information. In sentences like *the dog was bitten by the man*, both the NVN strategy and plausibility information support the incorrect active interpretation, so this sentence should sometimes be misinterpreted as meaning the dog bit the man. Ferreira (2003) provided experimental evidence for this claim, using a task in which participants were asked to identify the “do-er” and the “acted-on” entities. An important issue is to determine whether these results generalize to more implicit tasks such as those monitoring reading time.

Finally, some very interesting evidence from event-related potentials may also suggest that the processor considers ungrammatical analyses. Kim and Osterhout (2005) had participants read sentences like *the hearty meal was devouring the kids*. If readers simply constructed grammatically appropriate analyses, they should have treated *the hearty meal* as the agent of *devouring*, and then realized its semantic anomaly. However, *devouring* elicited a P600 effect, which is associated with syntactic rather than semantic anomaly. This suggests that readers interpreted *the hearty meal* as the patient of *devouring*, then reanalyzed. In contrast, an implausible sentence like *the dusty tabletops were devouring thoroughly* elicited an N400 effect, suggesting that *the dusty tabletops* was interpreted as the implausible agent of *devouring*. In other words, the syntactic misanalysis only occurs when an ungrammatical analysis is strongly plausible. In another study, Van Herten, Kolk, and Chwilla (2005) had participants read *the fox that at the poacher hunted* (in Dutch, meaning the fox that hunted the poacher), and found a P600 effect rather than the expected N400 (cf. Kolk, Chwilla, Van Herten, & Oor, 2003; Hoeks, Stowe, & Doedens, 2004 for similar findings). They suggested the P600 effect occurs because the structural analysis conflicts with a simultaneous analysis that assigns the most plausible interpretation to the combination of two noun phrases and the verb (in accord with another heuristic assumed by Bever, 1970).

These results may be compatible with Tabor et al. (2004), which also suggested that readers considered a highly plausible analysis but ungrammatical analysis. However, Kim and Osterhout’s (2005) results contrast with Ferreira (2003), because the plausible ungrammatical analysis is not consistent with the NVN strategy. One possibility is that verbs automatically activate their argument slots (e.g., Ferretti, Gagné, & McRae, 2003), and that the processor initially uses local semantic information to associate potential arguments with those slots. Such an account does not require the construction of complete analyses based on ungrammatical strategies, such as NVN. In summary, these

results suggest that readers may sometimes compute ungrammatical analyses, but it is too early to be certain whether they require wholesale revision of accounts of parsing.

Although such work suggests that processor may consider ungrammatical analyses, there are also some reasons to believe that the processing may be extremely well-tuned to the grammar, so that the processor precisely considers those analyses that are grammatical. In general, sentence processing research makes its predictions on the basis of the exact grammatical properties of a sentence fragment, and if the parser were regularly considering ungrammatical analyses, it would seem unlikely that so many predictions would be upheld (e.g., difficulty occurring exactly when theory claims that a garden path should occur). A good example of exact grammar-processor correspondence comes from the work on the processing of sentences involving potential island constraint violations, where most evidence suggests that the processor considers potentially grammatical analyses but does not consider analyses that are certainly ungrammatical (McElree & Griffith, 1998; Phillips, 2005; Stowe, 1986; Traxler & Pickering, 1996). It is difficult to reconcile these results with the suggestion that the processor routinely misanalyses relatively simple and frequent constructions such as passives.

#### 14. CONCLUSIONS

As this review shows, the majority of sentence processing research has continued to address relatively traditional topics, such as the initial factors affecting processing, re-analysis, and structural complexity. It now seems reasonable to conclude that completely modular accounts of initial processing such as traditional Garden-Path theory are not correct, and that the processor draws on a range of sources of information during initial processing. However, there are good reasons to believe that structural information does play a role during sentence processing and cannot be reduced to a set of weakly interacting constraints. Whereas research has shown that discourse context, plausibility, and frequency play important roles during sentence processing, they often do not entirely override basic preferences for particular types of structure.

One structural factor about which there is striking agreement among researchers is recency: People prefer to attach a new phrase to a more recent than a less recent attachment site. In the literature on ambiguity resolution, this means that they prefer more local to less local dependencies. This is explained by a syntax-driven principle like right association (Kimball, 1973) or late closure (Frazier, 1987a), or in terms of the decay of activation within connectionist or hybrid networks (e.g., Stevenson, 1994; Vosse & Kempen, 2000). In the literature on processing complexity, a recency (or locality) preference occurs because shorter dependencies involve less memory cost than longer ones (all other things being equal) (Gibson, 1998). Of course, there are some notable exceptions to recency as a principle explaining sentence processing, which raise the issue of exactly how recency should be defined. For example, the preferred analysis for both *the spy saw the cop with the binoculars* and *the daughter of the colonel who was on the balcony* tends to be a high attachment. However, both cases involve two possible attachment sites within a single

clause, suggesting that recency may only operate when the attachment sites are in different clauses (or perhaps different thematic domains, Frazier & Clifton, 1996), as in *John said Mary died yesterday*. Most accounts assume that the recency preference is a consequence of minimizing computational costs (e.g., Frazier, 1987a; Gibson & Pearlmuter, 1998), though Mitchell et al. (1995) argued that it may be the result of a preference for the most frequent structure. However, even if the latter turns out to be the case, recency would still be a primitive principle underlying language processing.

Our review suggests that research will continue to try to resolve traditional questions about the various constraints that affect initial processing. However, the field is also beginning to address other issues that are both deeper and broader. On the one hand, research on the architecture of the processor (e.g., competition vs. race) and on the apparent use of ungrammatical representations during initial analysis and reanalysis suggests ways to determine why the processor is organized in the way that it appears to be. On the other, we are beginning to see the use of more naturalistic methods of enquiry, the investigation of everyday language with its “imperfections,” and an attempt to situate parsing research in the “visual world”, so that parsing research no longer merely investigates how people parse isolated well-crafted sentences while reading or listening to idealized speech.

One recent new avenue of research has been the investigation of naturalistic language. Until recently, the great majority of psycholinguistic research involved participants producing or comprehending language in isolation from any interlocutor. In other words, it has been concerned almost exclusively with monologue (Clark, 1996; Pickering & Garrod, 2004). This limitation has affected parsing research at least as much as other fields, and has led to an almost exclusive focus on the comprehension of carefully constructed language, most notably written text. However, there are a few areas in which parsing research has begun to address the processing of less idealized language. One issue is the use of visual contexts in language comprehension (e.g., Altmann & Kamide, 1999; Knoeferle et al., 2005; Snedeker & Trueswell, 2003; Tanenhaus et al., 1995; see Tanenhaus & Trueswell, this volume). However, the contexts used so far have been limited to a few relevant entities (typically about four), and will have to be expanded considerably before claims about naturalistic language can easily be made. One step in this direction was provided by a study of task-oriented dialog (Brown-Schmidt, Campana, & Tanenhaus, 2005). Another example of naturalistic research in language comprehension is provided by Hanna and Tanenhaus (2004), in the context of a cookery task involving two interlocutors. Although these studies did not specifically investigate sentence processing, it may prove possible to use similar tasks to ask questions about parsing when language is used to refer to visually present entities.

Another respect in which naturalistic language differs from carefully constructed language is the ubiquity of disfluency, such as repetitions, self-interruptions, self-correction, and filled pauses like *um* and *uh*. Such disfluencies are extremely common, occurring perhaps six times per 100 words (e.g., Fox Tree, 1995), and appear to be used by comprehenders (Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Brennan & Williams, 1995). Bailey and Ferreira (2003) showed that disfluencies also affect syntactic ambiguity

resolution processes. They tested temporarily ambiguous sentences such as (44), where *the waiter* can initially be analyzed as part of a conjoined noun phrase or the subject of a conjoined clause.

44. Sandra bumped into the busboy and the waiter told her to be careful.

When a disfluency (*uh uh*) preceded *waiter*, participants were more likely to judge these sentences grammatical than when it preceded *busboy*. Listeners therefore may have used disfluencies to guide their choice of analysis or to assist in recovery from misanalysis. Interestingly, in this type of ambiguity, similar effects occurred with interruptions due to environmental sounds, but not when the interruptions were replaced by adjectives.

Finally, Lau and Ferreira (2005) presented listeners with sentences involving self-correction with a filled pause, such as (45):

45. The little girl chosen, uh, selected for the role celebrated with her parents and friends.

In this case, the reparandum (*chosen*) is consistent with the correct (reduced-relative) analysis of the sentence. The sentence was judged grammatical more often than a comparable sentence in which the reparandum was ambiguous (*picked*). These results suggest that the listener retains a representation of the ambiguous word and that it is not immediately “wiped clean” from the representation. However, it is not clear whether listeners compute an analysis using the reparandum, in a manner perhaps akin to the way that abandoned analyses appear to be retained (Christianson et al., 2001), or whether the reparandum primes the participle analysis of the ambiguous replacement verb (*selected*). After this, the reparandum may be entirely discarded (though of course it need not be). This interpretation would suggest that a reparandum behaves similarly to a subliminal “fast” prime (Trueswell & Kim, 1998).

These examples suggest that parsing research may begin to focus more on naturalistic language, dialog, and the integration with complex non-linguistic contexts. One effect of these trends is that there may be a closer link between comprehension and production, in particular between parsing and syntactic encoding during production. Dialog in particular involves tightly coupled production and comprehension, which suggests that people may straightforwardly access information that is common to both processes. In fact, researchers in production tend to assume that some comprehension takes place during self-monitoring (e.g., Levelt, 1989), but researchers in comprehension have rarely considered the possibility that the processor may draw on production mechanisms. One interesting case in which production may be implicated is when the comprehender predicts upcoming structure. In isolated sentences, there is a good evidence that individual words can be predicted (Van Berkum et al., 2005; DeLong, Urbach, & Kutas, 2005). In the same way, comprehenders may also predict grammatical structure, and perhaps use the production system to generate those predictions. The study of dialog, production/comprehension interactions, and predictive processing may prove to be linked in a way that extends the field of parsing research into new domains.

**REFERENCES**

- Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, 18, 129–144.
- Adams, B. C., Clifton, C., & Mitchell, D. C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin & Review*, 5, 265–270.
- Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language*, 38, 112–130.
- Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 714–735.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, 31, 685–712.
- Altmann, G. T. M., Garnham, A., & Henstra, J. A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology Learning, Memory, and Cognition*, 20, 209–216.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J. A. (1998). Late closure in context. *Journal of Memory and Language*, 38, 459–484.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, 15, 578–582.
- Bader, M., & Meng, M. (1999). Subject–object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, 28, 121–143.
- Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, 49, 183–200.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In: B. MacWhinney, & E. Bates (Eds), *The crosslinguistic study of sentence processing*. Cambridge, MA: Cambridge University Press.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In: J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–352). New York, NY: Wiley.

- Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, *44*, 297–324.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.
- Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, *45*, 391–411.
- Boland, J. E., & Boehm-Jernigan, H. (1998). Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language*, *39*, 684–719.
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*, 774–806.
- Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 468–481.
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398.
- Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, *33*, 251–283.
- Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory and Language*, *31*, 293–314.
- Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2005). Real-time reference resolution by nave participants during a task-based unscripted conversation. In J. C. Trueswell & M. K. Tanenhaus (Eds), *World-situated language processing: Bridging the language as product and language as action traditions*. Cambridge, MA: MIT Press.
- Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology*, *49A*, 664–695.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–126.
- Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*, 58–81.
- Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech*, *36*, 353–372.
- Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, *27*, 826–833.

- Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 687–696.
- Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, *52*, 144–169.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In: R. D. Luce, R. R. Bush, & E. Galanter (Eds), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *23*, 696–708.
- Clifton, C., Speer, S., & Abney, S.P. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, *30*, 251–271.
- Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317–334.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In: D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 320–358). Cambridge, England: Cambridge University Press.
- Crocker, M. W. (1995). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht: Kluwer.
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, *30*, 73–105.
- Cuetos, F., Mitchell, D. C., & Corley, M. M. B. (1996). Parsing in different languages. In: M. Carreiras, J. E. Garcia-Albea, & N. Sebastian-Galles (Eds), *Language processing in Spanish* (pp. 145–187). Mahwah, NJ: Erlbaum.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- DeLong, K. A., Urbach T. P., & Kutas M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.



- Desmet, T., Brysbaert, M., & De Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *Quarterly Journal of Experimental Psychology*, 55A, 879–896.
- Desmet, T., & Gibson, E. (2003). Disambiguation preferences and corpus frequencies in noun phrase conjunction. *Journal of Memory and Language*, 49, 353–374.
- Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes*, 21, 453–485.
- De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht, The Netherlands: Kluwer.
- De Vincenzi, M., & Job, R. (1995). An investigation of late closure: The role of syntax, thematic structure, and pragmatics in initial interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1303–1321.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164–203.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 555–568.
- Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30, 725–745.
- Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology*, 47, 247–275.
- Ferretti, T. R., Gagné, C. L., & McRae, K. (2003). Thematic role focusing by participle inflections: Evidence from conceptual combination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 118–127.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47, 250–272.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: Cambridge University Press.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9, 427–473.

- Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition*, 8, 417–459.
- Fodor, J. D., & Inoue, A. (1998). Attach anyway. In: J. D. Fodor, & F. Ferreira (Eds), *Reanalysis in sentence processing* (pp. 101–141). Dordrecht, The Netherlands: Kluwer.
- Ford, M., Bresnan, J. W., & Kaplan, R. M. (1982). A competence based theory of syntactic closure. In: J. W. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.
- Forster, K. I., & Olbrei, I. (1973). Semantic heuristics and syntactic analysis. *Cognition*, 2, 319–347.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34, 709–738.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. dissertation, Indiana University Linguistics Club, University of Connecticut.
- Frazier, L. (1987a). Sentence processing: A tutorial review. In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.
- Frazier, L. (1987b). Syntactic Processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, 5, 519–559.
- Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In: D. A. Balota, G. B. Flores d Arcais, & K. Rayner (Eds), *Comprehension processes in reading* (pp. 303–330). Hillsdale, NJ: Erlbaum.
- Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, 24, 437–468.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93–126.
- Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331–344.
- Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, 29, 343–370.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.
- Frazier, L., Taft, L., Roeper, T., Clifton, C., & Ehrlich, K. (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory & Cognition*, 12, 421–430.

- Frenck-Mestre, C., & Pynte, J. (2000). Resolving syntactic ambiguity: Cross-linguistic differences? In: M. D. Vincenzi, & R. Job (Eds), *Cross-linguistic perspectives on language processing*. Dordrecht, The Netherlands: Kluwer.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1219–1248.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished Ph.D. thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes*, 8, 147–161.
- Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, 59, 23–59.
- Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, 2, 262–268.
- Gibson, E., Pearlmutter, N. J., & Torrens, V. (1999). Recency and lexical preferences in Spanish. *Memory & Cognition*, 27, 603–611.
- Gibson, E., & Schutze, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language*, 40, 263–279.
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14, 225–248.
- Gilboy, E., Sopena, J. M., Clifton, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, 54, 131–167.
- Givón, T. (1984). *Syntax: A functional-typological approach* (Vol. 1). Amsterdam, The Netherlands: John Benjamins.
- Givón, T. (1990). *Syntax: A functional-typological approach* (Vol. 2). Amsterdam, The Netherlands: John Benjamins.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1411–1423.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97–114.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*, 425–430.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge: Cambridge university press.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science, 29*, 261–290.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*, 883–899.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*, 105–115.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language, 48*, 281–303.

Hemforth, B. (1993). *Kognitives Parsing: Repräsentationen und Verarbeitung grammatischen Wissens*. Sanct Augustin: Infix Verlag.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In: M. W. Crocker, M. J. Pickering, & C. Clifton, Jr. (Eds), *Architectures and mechanisms for language processing*. Cambridge, UK: Cambridge university press.

Herriot, P. (1969). The comprehension of active and passive sentences as a function of pragmatic expectations. *Journal of Verbal Learning and Verbal Behavior, 8*, 166–169.

Hoeks, J., Hendriks, P., Vonk, W., Brown, C., & Hagoort, P. (in press). Processing NP- versus S-coordination ambiguity: Thematic information does not completely eliminate processing difficulty. *Quarterly Journal of Experimental Psychology A*.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19*, 59–73.

Hoeks, J. C. J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language, 46*, 99–119.

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28*, 668–689.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition, 90*, 3–27.

Hyönä, J., & Hujanen, H. (1997). Effects of case marking and word order on sentence parsing in Finnish: An eye fixation analysis. *Quarterly Journal of Experimental Psychology A, 50*, 841–858.

- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kaan, E. (2001). Effects of NP type on the resolution of word-order ambiguities. *Journal of Psycholinguistic Research*, 30, 529–547.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159–201.
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94, 113–147.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133, 450–467.
- Kayne, R. (1984). *Connectedness and binary branching*. Dordrecht, The Netherlands: Foris.
- Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review*, 8, 132–138.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–225.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15–47.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.
- Kjelgaard, M. M., & Speer, S. S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153–194.
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95, 95–127.
- Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1–36.
- Koster, J. (1975). Dutch as an SOV language. *Linguistic Analysis*, 15, 111–136.

- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word dependency and semantic association. *Nature*, *307*, 161–163.
- Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge, UK: Cambridge University Press.
- Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, *20*, 633–666.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, CA: MIT Press.
- Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*, 93–115.
- Liversedge, S. P., Pickering, M. J., Branigan, H. P., & Van Gompel, R. P. G. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 461–475.
- Liversedge, S. P., Pickering, M. J., Clayes, E., & Branigan, H. P. (2003). Thematic processing of adjuncts: Evidence from an eye-tracking experiment. *Psychonomic Bulletin & Review*, *10*, 667–675.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*, 50–68.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.
- Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226–228.
- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, *45A*, 73–87.
- McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 134–157.

- McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 432–460.
- McKoon, G., & Ratcliff, R. (2003). Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review*, *110*, 490–525.
- McRae, K., Hare, M., & Tanenhaus, M. K. (2005). Meaning through syntax is insufficient to explain comprehension with reduced relative clauses: Comment on McKoon and Ratcliff (2003). *Psychological Review*, *112*, 1022–1031.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283–312.
- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition*, *23*, 477–494.
- Mehler, J., & Carey, P. (1967). Role of surface and base structure in the perception of sentences. *Journal of Verbal Learning and Verbal Behavior*, *6*, 335–338.
- Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In: R. D. Luce, R. R. Bush, & E. Galanter (Eds), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Wiley.
- Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In: M. Coltheart (Ed.), *Attention and performance XII: The psychology of language* (pp. 601–618). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mitchell, D. C., & Brysbaert, M. (1998). Challenges to recent theories of crosslinguistic variations in parsing: Evidence from Dutch. In: D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective (Syntax and Semantics 31)* (pp. 313–335). San Diego: Academic Press.
- Mitchell, D. C., Corley, M. M., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 69–88.
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*, 469–488.
- Mitchell, D. C., & Holmes, V. M. (1985). The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language*, *24*, 542–559.
- Münste, T. F., Heinze, H. J., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, *5*, 335–344.
- Murray, W. S., & Liversedge, S. P. (1994). Referential context effects on syntactic processing. In: C. Clifton, Jr., L. Frazier, & K. Rayner (Eds), *Perspectives on sentence processing* (pp. 359–388). Hillsdale, NJ: Erlbaum.

- Noppeney, U., & Price, C. J. (2004). An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience*, *16*, 702–713.
- Novick, J. M., Kim, A., & Trueswell, J. C. (2003). Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, *32*, 57–75.
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, *8*, 413–437.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 786–803.
- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*, 283–317.
- Phillips, C. (2005). *The real time status of island constraints*. Unpublished manuscript.
- Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407–428.
- Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*, 229–259.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–225.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 940–961.
- Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 447–475.
- Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Stanford, CA: CSLI and University of Chicago Press.
- Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language*, *64*, 539–576.
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, *11*, 165–191.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.



Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology*, 39A, 657–673.

Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, 45, 109–139.

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language*, 45, 308–336.

Schriefers, H., Friederici, A. D., & Kuhn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34, 499–520.

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, 40, 409–431.

Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 5, 219–227.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49, 238–299.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1521–1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447–481.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55, 227–267.

Stabler, E. P. (1994). The finite connectivity of linguistic structures. In: C. Clifton, Jr., L. Frazier, & K. Rayner (Eds), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.

Steedman, M. J. (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191–196.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, 23, 295–322.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227–245.

- Stowe, L. A., Tanenhaus, M. K., & Carlson, G. N. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, 34, 319–340.
- Sturt, P., & Crocker, M. W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11, 449–494.
- Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40, 136–150.
- Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort? *Journal of Memory and Language*, 45, 283–307.
- Sturt, P., Scheepers, C., & Pickering, M. (2002). Syntactic ambiguity resolution after initial misanalysis: The role of recency. *Journal of Memory and Language*, 46, 371–390.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.
- Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 431–450.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–271.
- Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, 23, 491–515.
- Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In: C. Umiltà, & M. Moscovitch (Eds), *Attention and performance 15: Conscious and nonconscious information processing*. (pp. 589–615). Cambridge, MA: MIT Press.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, 27, 597–632.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454–475.

- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558–592.
- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*, 204–224.
- Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566–585.
- Trueswell, J. C., & Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language*, *39*, 102–123.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 528–553.
- Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 683–692.
- Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467.
- Van Gompel, R. P. G., & Pickering, M. J. (2001). Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998). *Psychonomic Bulletin & Review*, *8*, 851–857.
- Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (in press). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*.
- Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284–307.
- Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*, 225–258.
- Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*, 241–255.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In: M. Halle, J. Bresnan, & G. A. Miller (Eds), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, 26, 714–738.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79–112.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104, 444–466.

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early-closure attachment on first-pass reading times in French. *Quarterly Journal of Experimental Psychology*, 50A, 421–438.

This page intentionally left blank

## Chapter 13

### Prosody

*Shari Speer and Allison Blodgett*

This chapter addresses the role of prosody in language processing. Prosody has been called “the organizational structure of speech” (Beckman, 1996). We use the term “prosody” to refer to stress, rhythm, and intonation in spoken sentences. The study of prosody in processing necessarily relies upon the theories and methods developed to investigate spoken language in multiple disciplines: phonetics, phonology, speech production and perception, and psycholinguistics. Prosodic structure is formally described in linguistic theories of autosegmental phonology and intonation, and has measurable acoustic–phonetic correlates, including variation in fundamental frequency, spectral information, amplitude, and the relative duration of sound and silence across and between utterances.

The importance of prosody to understanding the language comprehension system stems from the primacy of spoken language. Infants acquire language from input that is almost entirely auditory, and indeed show a preference for the prosody of their native language as early as 3 days of age, suggesting pre-natal sensitivity (Mehler et al., 1988). The majority of adult language input is also spoken, and in fact many of the world’s languages and dialects have no written form. Thus spoken language, structured by prosodic form, is the “base” on which the mental representations and processes that compose linguistic processing are built.

As the study of language processing has shifted from considering the recognition of isolated words and sentences of text to considering the speaker/hearer’s behavior in “everyday” conversation in broader discourse contexts, an understanding of the fundamental role of prosodic structure has taken on increasing importance. Prosodic structure informs language processing at multiple levels of the spoken signal (and has been argued to inform text processing as well, cf. Fodor, 2002).

Take for example the utterance sets in (1)–(3):

1. What’s that ahead in the road?/ What’s that, a *HEAD* in the *ROAD*? (example attributed to Ken Church)

2. Triceratops oil/Try, Sarah, topsoil/Trice, air atop soil (example due to Janet Pierrehumbert)
3. The professor said her student had on socks that did not match./The professor, said her student, had on socks that did not match.

The phoneme strings in each of these utterance sets are identical. Thus, when spoken in isolation, a listener must rely on prosody alone to recover the speaker's intended meaning. The speaker produces prosodic information to indicate the identity of utterance components that correspond to multiple units of linguistic analysis, including syllables (*try* vs. *trice*), words (*ahead* vs. *HEAD*; *atop soil* vs. *topsoil*), syntactic constituents ([ahead in the road]<sub>PP</sub> vs. [[a head] [in the road]<sub>PP</sub>]<sub>NP</sub>; [The professor]<sub>NP</sub> [said]<sub>VP</sub> vs. [The professor]<sub>NP</sub> <sub>S</sub>[said]<sub>VP</sub>), and the pragmatic status of items in the discourse (nonemphatic *road* vs. emphatic *ROAD*). The broad range of utterance types for which a meaning distinction is dependent on prosodic information (items that function like minimal pairs in phonology) is well established (for example sentence pairs, see Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Wales & Toner, 1979).

The goal of this chapter is to review selected literature on prosodic processing in adult speakers and listeners, with a focus on prosodic phrasing and sentence-level prominence. The structure of the chapter is as follows: First, we give a brief history of some early attempts to understand the impact of prosodic structure on comprehension, highlighting some of the basic issues that continue to motivate research in the area and presenting some of the methodological challenges inherent in the study of prosodic effects. Next, we review briefly phonetic measures and phonological annotation methods commonly used to describe the prosody of production data and spoken experimental materials. We discuss strategies for the construction of experimental materials and argue that the precise specification of prosodic form is critical to the understanding and replication of prosodic effects in processing. We then present current research findings on prosodic phrasing and its relationship to syntactic processing, and prosodic prominence as it affects information structure, and interacts with phrasing. For reasons of space, we limit our discussion primarily to research conducted in English, although we note that a thorough and general understanding of the nature of prosodic processing cannot rely on work in a single language. Finally, we review some recent sentence-processing models that incorporate prosodic structure. This chapter does not provide extensive review of work on the processing of lexical level prosodic information, such as English lexical stress and Japanese lexical accent (see Cutler, Dahan, & VanDonselaar, 1997, for an overview of these issues).

## 1. A BRIEF HISTORY

The impact of sound-based phrasing and prominence on comprehension has long been noted by language researchers. Take for example one of the earliest controversies in psycholinguistic research, that surrounding the seminal "click migration studies," which established the importance of syntactic constituency to the process of sentence comprehension (Fodor & Bever, 1965; Garrett, Bever, & Fodor, 1966; Wingfield & Klein, 1971).

The experiments used spoken sentence materials, presented with bursts of noise or “clicks” at locations before, after, or coincident with syntactic constituent boundaries. Participants listened to the sentences and then wrote down the locations where they remembered hearing the clicks. An analysis of their errors showed that they had heard “migrations,” with more clicks falsely reported at syntactically determined boundaries. However, because the sentence materials were spoken, skeptics offered the alternative explanation that clicks migrated to the audible phrase breaks that speakers tend to pronounce between clauses. A follow-up study (Garrett et al., 1966) used cross-spliced materials to substantially remove prosodic phrasing and hold sentence pronunciation constant across syntactic conditions, thus demonstrating the syntactic effect independent of spoken phrasing. This design, of course, obscures any effect of prosodic structure. In contrast, a separate study manipulated both factors (Wingfield & Klein, 1971). Sentences were pronounced with phrasing to indicate syntactic meaning and cross-spliced so that syntactic and prosodic boundaries either conflicted or coincided. Listeners performed the click location task and also wrote down the sentence they remembered hearing. The pattern of “migration” errors showed that clicks were falsely reported at both prosodic and syntactic phrase boundary locations. Interestingly, there were many errors in the versions of the sentences that participants wrote after listening, and the majority of these altered the syntactic content of the utterance to make it consistent with the prosodic phrasing that was heard.

Around the same time, researchers in phonetics were cataloguing the types of syntactic ambiguity that were amenable to resolution by prosodic phrasing, especially the location of pauses. Typically, speakers read ambiguous sentences such as *The old men and women were the first into the lifeboats* (from Lehiste, 1973), sometimes with instructions to indicate whether both *men* and *women* were old, and other times without such instructions to disambiguate. Listeners were tested to see if they could accurately distinguish the meaning intended by the speaker. Results showed that prosody could indeed be used to determine the final meaning assigned to many syntactic and semantic ambiguities, but that others (e.g., *Flying planes can be dangerous*) remained impervious to disambiguation (cf. Lehiste, 1973; Wales & Toner, 1979; Cooper & Paccia-Cooper, 1980).

Although these early studies showed clear effects of prosody on comprehension, the focus of psycholinguistic research at the time was on establishing syntactic effects rather than on the aspects of pronunciation that might give rise to their recognition from the spoken signal. Prosody was generally considered a “nuisance” variable and a drawback to conducting sentence processing research with spoken materials. As such, it was controlled by using cross-spliced “monotone” or “neutral prosody” materials, or simply by dispensing with speech altogether in favor of text materials. We note that, in hindsight, these solutions must be of limited effectiveness. “Monotone” sentence pronunciations must still contain a prosodic organizational structure, albeit one with a reduction in pitch range and/or a potentially ungrammatical prosodic form (e.g., list intonation used for sentence production), or one that introduces unknown syntactic bias (e.g., syntactic attachment of a constituent can be signaled by either the presence *or* the absence of a prosodic boundary).



The study of prosodic effects on comprehension began in earnest in the early 1980s and 1990s. Many reaction time studies showed processing advantages for spoken sentence stimuli that contained prosody consistent with the correct syntactic parse, as compared to prosody that was inconsistent (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Nagel, Shapiro, & Nawy, 1994; Nagel, Shapiro, Tuller, & Nawy, 1996; Pynte & Prieur, 1996; Slowiaczek, 1981; Warren, Grabe, & Nolan, 1995). However, other researchers using similar procedures and syntactic materials produced only null or weak effects (e.g., Watt & Murray, 1996; Nicol & Pickering, 1993; for a comprehensive review, see Warren, 1997). The controversy arising from this work established basic questions that motivate current research on prosodic processing. Specifically, do speakers reliably produce prosodic phrasing and prominence to indicate the syntactic, semantic, and pragmatic content of an utterance? Do they do so only when they are aware of ambiguity and instructed to disambiguate? Do listeners reliably make use of prosodic information during comprehension? Are systematic correspondences such as those between prosodic and syntactic constituency used generally in comprehension, or is their usefulness limited to circumstances where other types of disambiguating context is unavailable? If prosody is important in spoken language comprehension, what aspects of prosodic structure are critical to the processing system, how are they recognized, and what is the time course of their integration with lexical, syntactic, and discourse aspects of comprehension?

## 2. METHODOLOGICAL CHALLENGES

Looking across early studies with conflicting findings, we argue that a primary reason for the lack of consistency was the lack of a standard to allow comparison of the manipulated prosodic structures. Methods used to manipulate prosodic structure varied widely across studies. Some used “untrained” or “naïve” speakers, who produced a felicitous prosodic contour by speaking, while holding in mind one of two meanings for a syntactic ambiguity, or by reading the sentence in a disambiguating context (including preceding semantic information and text marked with punctuation, boldface, or underlining). Others used speakers trained as actors or radio announcers, who were given instructions to disambiguate. Still others used speakers trained in phonetics and phonology who were instructed to instantiate particular prosodies, or speech synthesizers that were set to produce a particular set of durations and tones. Once prosodies were produced, some researchers did not describe the sound characteristics at all, while others gave a brief impressionistic description. Some provided phonetic measurements of duration and fundamental frequency, some for the actual materials used in a study, but others for a separate set of similar materials. Few provided phonological transcription or annotation for the specific materials used to demonstrate prosodic effects on comprehension. The combination of differences in speech style and differences in prosodic description led to an unsurprising confusion about the type of sound-based information that is important for the resolution of syntactic ambiguity.

As many researchers have noted (e.g., Selkirk, 1984; Pierrehumbert, 1980; Price et al., 1991; Warren et al., 1995; Beckman & Pierrehumbert, 1986), the correspondence between

prosody and other linguistic structures is complex. A sentence with a particular syntactic structure may be pronounced grammatically with a variety of prosodies. For example, the sentence in (4) (from Selkirk, 2000) can be grammatically pronounced with any of the three phrasings shown (reading [4a] requires focus on the verb or pronoun). In addition, a particular prosodic phrasing may be used to pronounce a variety of syntactic forms grammatically. The examples in (5) and (6) show a pair of sentences that can carry the same tune, because they have the same pattern of weak and strong syllables, despite their different syntactic structures.

4. a. She loaned / her rollerblades / to Robin.  
    b. She loaned her rollerblades / to Robin.  
    c. She loaned her rollerblades to Robin.
5. The color printer / was loud and slow.
6. On Tuesday morning / we saw the show.

A similar complexity concerns the mapping from the phonetic string to the prosodic representation, where a single pronunciation may be ambiguous between two prosodic structures, and a single prosodic structure may have more than one phonetic implementation (Shattuck-Hufnagel & Turk, 1996; Beckman, 1996). This state of affairs, where there is no single component at one level of linguistic analysis that corresponds uniquely to a single component at another level of linguistic analysis, has been called the “lack of invariance” problem. It is typical of the speech signal and is a long-standing topic of research on speech perception (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Miller, 1990). The effect of this complexity on the study of prosodic disambiguation of syntax is to produce potentially conflicting results. While the grammatical prosody used for a particular sentence type in one study may have been syntactically disambiguating, a different, but still grammatical, prosody in another study may not have been.

To resolve this problem, researchers called for explicit specification of the acoustic and phonological forms present in experimental stimuli. Such specifics are critical for the replication and extension of experimental findings (Warren, 1997; Kjelgaard & Speer, 1999). An additional concern is that the prosody of experimental materials be evaluated for naturalness, or the extent to which it represents a “typical” sentence pronunciation (Allbritton, McKoon, & Ratcliff, 1996). The majority of current research on prosody and language processing employs acoustic phonetic description as well as auditory annotation of experimental materials. Papers included in the following sections are generally those that have provided such specifications for manipulated prosodic structures. Many also specify the naturalness or typicality of the utterances used. Methods for such specifications include pretesting auditory stimuli to establish comparable acceptability of pronunciations across conditions (e.g., Kjelgaard & Speer, 1999), constructing materials based on prosodies collected from naïve speakers in production studies (Warren et al., 1995; Ito & Speer, 2006), using materials collected in such production studies as stimuli for experiments (Schafer, Speer, & Warren, 2003), or conducting simultaneous production/comprehension experiments with naïve speakers and listeners (Snedeker & Trueswell, 2003; Kraljic & Brennan, 2005).

### 3. DESCRIBING PROSODIC STRUCTURE

The following section summarizes current measures frequently used to describe prosodic form, with a focus on the types of description used in this chapter. The prosodic structure inherent in spoken utterances is typically described in two ways: (1) Utterances are submitted to phonetic analyses of duration and fundamental frequency (and sometimes intensity and more detailed analysis of spectra). (2) Utterances are described using an annotation system to mark intonational events. Figure 1 shows a Praat<sup>1</sup> display with an utterance labeled with both types of information.

#### 3.1. Duration and Fundamental Frequency Measures

In English, prosodic phrase boundaries are characterized by segmental lengthening and pitch movement (e.g., fall, rise, or fall–rise) on the phrase final word, and often by a silent duration following it. Sentence-level prosodic prominence is characterized by localized pitch excursions, either high (peaks) or low (valleys). Figure 1 displays the amplitude-by-time waveform for an utterance (from Kjelgaard & Speer, 1999), time-aligned to its fundamental frequency ( $F_0$ ) contour and duration values for its two major prosodic phrases and the intervening silence. Vertical gray lines in the  $F_0$  window correspond to word boundaries and pitch peaks in the signal (pitch events are described in more detail

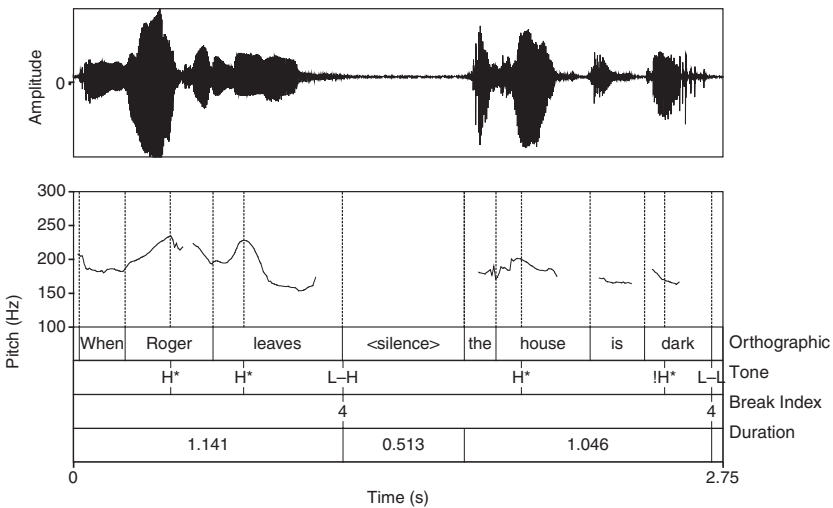


Figure 1. A ToBI-annotated utterance with amplitude waveform, pitch trace, and duration values. Utterance from Kjelgaard & Speer, 1999.

<sup>1</sup> Praat is a freeware package created for the phonetic analysis of speech stimuli (Boersma & Weenink, 2006; www.praat.org).

in the next section). In the specification of stimuli, marking the location of word boundaries in order to measure the duration of critical words is usually accomplished via a combination of listening and visual inspection of the waveform (sometimes accompanied by a spectral image). This is necessary, as continuous speech often contains no silence and/or little reduction of amplitude between words, due to coarticulation of word-final segments with the initial segment of the following word. Fundamental frequency values are most often extracted automatically by algorithm. Prominence is compared across stimuli using the most extreme  $F_0$  value in the region of a peak or trough, with fall-rise patterns requiring two values. A comparison of the speech waveform and the  $F_0$  trace in Figure 1 shows “gaps” in the available  $F_0$  information due to the presence of unvoiced segments (e.g., the fricative /s/ in “house,” the final /k/ of “dark”). For this reason, stimuli are often constructed to contain predominately voiced (and sonorant) segments in regions critical to a prosodic manipulation.

### 3.2. Annotation Systems

A variety of annotation systems are in current use. A common system used for English is ToBI, (Tone and Brake Indices; Silverman et al., 1992; Beckman & Ayers, 1994), which we will use in the description of the English experiments discussed in this review. ToBI systems have been (and continue to be) developed for other languages, including but not limited to Cantonese (Wong, Chan, & Beckman, 2005), German (Grice, Bauman, & Benzmueller, 2005), Korean (Beckman & Jun, 1996), and Japanese (Venditti, 1995; Campbell & Venditti, 1995). Similar systems have also been developed for Dutch (Gussenhoven, Reitveld, Kerkhoff, & Terken, 2003) and for cross-dialect comparisons within a language, e.g., IViE (Intonational Variation in English; Grabe, Post, & Nolan, 2005), which is used in the transcription of international dialects of English. The systems have in common that they are based on autosegmental and metrical theories of phonology. Briefly, these theories represent the rhythm and timing of a sentence as a pattern of weak and strong beats (unstressed and stressed syllables in English), hierarchically arranged according to a tree or grid structure to form constituents, e.g., feet, prosodic words, phonological phrases, and intonation phrases.

Autosegmental theories of intonation represent the tune of a sentence as a series of tonal events, and annotations specify the locations of these events as identified in a particular utterance (this is shown in the “Tone” tier in Figure 1). Some tones are aligned at the right edge of a phrase. In English, these are the boundary tone, written as L% or H%, and the phrase accent, written as L- or H-. ToBI specifies the correspondence between tones and phrasal constituents for two levels of phrasing. The intonation phrase (IP) is delimited by a single boundary tone at its right edge. The intermediate phrase (ip) is delimited by a phrase accent, which is realized over the material between the final pitch accent in a phrase and its final boundary tone. Other tones are directly associated with the lexically stressed syllable of a prominent word, and time aligned with a peak or valley in the  $F_0$  trace for annotation. These are pitch accents, written as T\*, and may be composed of one or two tones. The inventory of pitch accents for English, as characterized in the

ToBI system, includes H\*, L\*, L+H\*, L\*+H, and H+!H\*.<sup>2</sup> ToBI conventions specify that every utterance has at least one IP, each IP has at least one ip, and each ip has at least one pitch accent. ToBI annotation also includes five levels of break index to mark the level of disjuncture between the words and phrases within an utterance (shown in the “Break Index” tier in Figure 1). These are 0, indicating a merger of words, such as “wanna” for “want to;” 1, for most word-level breaks; 2, for a very strong word-level break or weak intermediate phrase level break, often with a mismatch between the durational and tonal indicators for break size; 3, for a juncture at an intermediate phrase boundary; and 4, for a juncture at an intonation phrase boundary.

### 3.3. Recognizing Prosodic Form

Before prosodic structure can influence the processing of other aspects of an utterance, prosody itself must be recognized and represented in the processing system. Just as other levels of linguistic representation can be ambiguously specified by the input (e.g., lexical items and syntactic structure), an utterance may be ambiguous between two prosodic structures. This phenomenon is discussed in the literature on phonetic transcription (Shattuck-Hufnagel & Turk, 1996; See Beckman, 1996, for an extensive discussion), and we draw attention to it here, citing cases that have particular relevance for the construction of experimental stimuli.

In English, particular combinations of timing and tone can result in ambiguity between intermediate and intonation phrase levels when phrasal boundaries are not sentence-final. For example, a high, level phrase-final pitch contour could indicate either a H- phrase accent and thus an ip boundary, or it could indicate a H-L% sequence, and thus an IP boundary. Similarly, a low or falling final contour could indicate either an L- phrase accent or a L-L% phrase accent-boundary tone combination. The difference between intermediate and intonation phrases of these types can often be distinguished by relying on other aspects of the utterance that optionally co-occur at the boundary location. For example, the L-L% sequence is more likely than L- to show a drop in  $F_0$  to the bottom of the speaker’s pitch range, greater lengthening of the phrase final word, and following silence. In contrast, L- and H- are easily distinguished from the bi-tonal L-H% (“continuation rise”) and the rising H-H% (“question”) contours.

Ambiguity can also arise in the recognition of the presence/absence of pitch accents, and the recognition of pitch accent type. A presence/absence ambiguity of the H\* accent can arise in English declarative sentences, which often exhibit a sequence of two or more H\* accents followed by a L-L%.<sup>3</sup> The portion of an utterance that occurs between the two

<sup>2</sup> The symbol ! before a H tone indicates “downstep,” a process that lowers the pitch height of an accent with reference to the one immediately preceding it.

<sup>3</sup> This sequence is often called a “hat pattern,” as the pitch contour is shaped like the crown of a fedora. That is, it rises to an initial H\* peak, dips slightly on unaccented syllables that occur between accents, rises again to a comparable peak for the second H\*, and then falls.

H accents is generally pronounced with a relatively high pitch. As the number of words between the accented syllables increase, lexically strong syllables remain high pitched, and thus there can be an ambiguity as to whether they, too, carry a H\* accent. Similarly, L\* accent placement may be ambiguous on phrase-final words when a phrase ends with a L-L% boundary, as the final low pitch of the utterance reaches the bottom of the speaker's pitch range, compressing the region available for the execution of a valley to indicate accent. An ambiguity of pitch accent type that is much discussed in recent literature is that between the L+H\* and H\* accents. While L+H\* accents can be distinguished from H\* by a low initial region, steeper rise to the H tone, and later alignment in the strong syllable, differences in segmental content, speech rate, and overall pitch range can obscure these features from the signal. Researchers in phonetics differ in whether they support a distinction between the two accents (cf. Bartels & Kingston, 1994; Krahmer & Swerts, 2001; Ladd & Schepman, 2003).

As noted above, the correspondence between prosodic structure and other levels of linguistic representation such as syntactic and discourse structure is a many-to-many mapping. Ambiguities contained in the speech signal that conveys prosodic structure can substantially increase the complexity of this correspondence. On the one hand, such complexity has disadvantages for psycholinguistic experimentation: grammatical stimuli can take many forms, leading to ambiguous stimuli and potentially to conflicting results across studies. On the other hand, the complex correspondence between prosody and other structures can also work to the advantage of the experimenter: Ambiguity has provided a window of opportunity for researchers in other areas of psycholinguistics for generations, allowing them to specify the mechanisms that underlie multiple levels of language processing. If we carefully specify the phonology and phonetics of our experimental materials, we can make use of the ambiguities to begin to develop a principled account of when and where particular prosodic entities influence the comprehension process.

#### 4. PROSODIC PHRASING AND SYNTACTIC PROCESSING

Current work on prosody and language processing shows a consistent, early influence of a prosodic representation on the production and comprehension of syntactic meaning in sentences and discourse. Three broad conclusions are suggested by the recent work on prosodic phrasing and syntactic processing summarized in this section.

1. Prosody has an immediate rather than delayed effect on a range of syntactic parsing decisions, suggesting that prosody can determine the incremental construction of a syntactic representation on the fly during sentence comprehension. These effects have implications for the architecture of the language processing mechanism, suggesting that prosodic structure is primary to the process of recovering sentence structure.
2. The production of prosodic phrasing seems to be driven by grammatical constraints on the processing system rather than by situational demands for speakers to use prosodic form to disambiguate syntax. While this conclusion remains somewhat

controversial, the preponderance of current evidence suggests that speakers produce a prosody consistent with the meaning they intend, regardless of the situationally varying needs of the listener. In turn, listeners make use of the prosodic structure provided by speakers to determine basic syntactic constituency relations.

3. Prosodic processing effects are attributable to the use of a global prosodic representation rather than to local markers in a lexico-syntactic representation. That is, the effect of a particular local prosodic boundary is dependent on its position in the overall prosodic structure of an utterance. Depending on prosodic context, similar effects can be found using different levels of local prosodic phrasing – e.g., in English, syntactic attachment can be disambiguated by either an intermediate or an intonation phrase, depending on the presence and order of other boundaries in a particular prosodic structure.

## 5. IMMEDIATE INTEGRATION OF PROSODIC AND SYNTACTIC STRUCTURES

Initial evidence for the immediate use of prosodic phrasing in the recovery of syntactic structure came from work on temporary syntactic ambiguity in early/late closure syntactic ambiguities (e.g., *When Roger leaves the house it's/is dark*; Kjelgaard & Speer, 1999). Such sentences are temporarily ambiguous at the attachment of the noun phrase *the house*, but the ambiguity is resolved immediately by the following words *is* or *it*. For early closure syntax, the structurally ambiguous noun phrase *the house* serves as the subject of the main clause, while for late closure syntax, *the house* serves as a direct object in the subordinate clause. Studies of silent reading have repeatedly shown a processing disadvantage (called a “garden path effect”) for early closure versions of such sentences as compared to their late closure counterparts, with this difference attributed either to lexical factors (cf. Tanenhaus & Carlson, 1989; MacDonald, Perlmutter, & Seidenberg, 1994) or to an initial misanalysis of the syntactic constituent structure (cf. Frazier & Clifton, 1996; Ferreira & Henderson, 1990).

Table 1 shows an example stimulus set from the study. Three prosodic conditions – cooperating, baseline, and conflicting – were created for the two syntactic structures. Materials were produced by a trained phonetician. In conditions with cooperating prosody, a full intonation phrase boundary (L-L%) was aligned with the end of the syntactic clause boundary for the subordinate clause (i.e., with *leaves* in the case of early closure syntax and with *the house* in the case of late closure syntax). In conditions with conflicting prosody, a full intonation phrase boundary (L-L%) was aligned with the location of the subordinate clause boundary from the alternative closure condition. In the baseline prosody conditions, the speaker produced an ambiguous prosody – a low phrase accent (L-) was used to create a tonal plateau and similar word durations across the syntactically ambiguous region (e.g., *leaves the house*). The sequence is ambiguous because, in the absence of segmental lengthening on either the verb or the noun, the location of the intermediate phrase boundary is undetermined. Materials were ToBI-annotated and presented with supporting mean  $F_0$  values and word durations, to specify the prosodic and acoustic characteristics of the auditory stimuli.

Table 1  
ToBI transcriptions for each prosodic condition for early and late closure syntactic ambiguities.

Prosody:	Cooperating			H*L-L%			H*L-L%
	baseline		L+H*	L-			H*L-L%
	conflicting					H*L-L%	H*L-L%
Syntax:	Early closure	When	Roger	leaves	the	house	is dark.
Prosody:	Cooperating					H*L-L%	H*L-L%
	baseline		L+H*			L-	H*L-L%
	conflicting			H*L-L%			H*L-L%
Syntax:	Late closure	When	Roger	leaves	the	house	it's dark.

Note: From Kjelgaard and Speer (1999), Experiments 1–3.

Kjelgaard and Speer’s (1999) baseline prosodies capitalized upon prosodic ambiguity, allowing comparison of prosodic contours that were felicitous for either syntactic structure to those that biased the listener toward a particular interpretation. This comparison directly addresses the interaction of prosodic and syntactic structures during sentence comprehension. The inclusion of a baseline contrasts with the design of the majority of studies of prosodic phrasing, where prosodic manipulations resemble the unambiguous prosody of the cooperating and conflicting conditions (e.g., Blodgett, 2004; Carlson, Clifton, & Frazier, 2001; Kang & Speer, 2004; Schafer, 1997; Slowiaczek, 1981; Watson & Gibson, 2004). Indeed few researchers seem to acknowledge the fact that speakers can produce prosodic contours that map to multiple and sometimes competing syntactic structures (cf. Carlson, 2001). By using a three-way contrast of cooperating, baseline, and conflicting prosodies, Kjelgaard and Speer were able to investigate the extent to which cooperating and conflicting prosodies respectively facilitate or interfere with processes of language comprehension.

In separate experiments, Kjelgaard and Speer (1999) tested the effects of phrase-final IP with substantial following silence and the effects of phrase-final IP with no following silent duration (H- in cooperating and conflicting conditions). Each set of materials was used in an end-of-sentence judgment experiment as well as in a cross-modal naming task. Participants in judgment task experiments listened to full sentences, pressing a button when they had comprehended the sentence. Cross-modal naming participants heard an auditory fragment from a particular prosodic condition, as in (7)

- H\*
7. [[When Roger leaves the house <sub>L</sub>]<sub>L%</sub>]

At the offset of the auditory stimulus, a visual target, congruent with either early closure syntax (e.g., *is*) or late closure (*it’s*) was presented. Participants named the visual



target as quickly as possible and then provided an ending to the sentence fragment that the auditory and visual stimuli set up. Thus, the sentence completion part of the task forced participants to integrate the visual target and auditory stimulus. Data from the combined experiments allowed evaluation of both the immediate and the longer-term effects of prosody on syntactic parsing.

Results in all experiments showed that the prosody of an utterance can have early and immediate effects on language comprehension. More specifically, the prosodic phrasing of an utterance can facilitate or interfere with early syntactic processing. Facilitation was demonstrated because naming times to early closure targets following early closure cooperating prosody were significantly shorter than naming times to early closure targets following early closure baseline prosody. In addition, there was no evidence of any garden path effect in the early closure cooperating condition. Naming times for these utterances were not significantly different from naming times in the late closure cooperating condition. Interference was demonstrated because naming times following conflicting prosody were significantly longer than naming times following baseline prosody regardless of whether the visual target required early closure syntax or late closure syntax. While similar facilitation and interference effects were observed in speeded end-of-sentence judgment tasks, it is the cross-modal naming time data – collected at the point of syntactic ambiguity resolution – that demonstrate the immediacy with which prosody influences syntactic processing.

As conducted, Kjelgaard and Speer's (1999) results were also consistent with a modified and somewhat weakened conclusion: namely, that prosodic boundary location determines the initial structure of an early/late closure ambiguity, but only when the lexical content of the auditory fragments provides equal support for the two closure types. Kjelgaard and Speer's own sentence completion data (from pretesting) demonstrated that participants were as likely to provide early closure completions to written versions of the auditory fragments as they were to provide late closure completions. However, the finding that phrasing immediately influenced which of two syntactic structures were built – when the fragments as a group equally supported the structural alternatives – was quite similar to previous eye-tracking results (Garnsey, Pearlmutter, Myers, & Lotocky; 1997; but cf. Pickering, Traxler, & Crocker, 2001). They found that plausibility immediately influenced the resolution of a syntactic ambiguity during reading, but only when the verbs were as likely to occur with one structure as the other. Thus, if prosodic phrasing and plausibility have similar effects on syntactic processing, it could be the case that prosodic phrase boundary location would have little effect in determining the initial closure structure for fragments that were strongly biased toward a particular closure type.

Recent results from Blodgett (2004) provide evidence against such a conclusion. Instead, they support a stronger claim that intonation phrase boundaries have an immediate influence on syntactic processing regardless of lexically based syntactic biases. As shown in Table 2, Blodgett's auditory stimuli consisted of three sets: a transitive-bias set containing verbs like *loads* that occur more often with a direct object than without, an intransitive-bias set containing verbs like *moves* that occur more often without a direct

Table 2

ToBI transcriptions for the two intonation phrase boundary conditions and three verb-bias sets in Blodgett (2004), Experiment 1.

Closure bias of verb set	Example auditory fragments with late and early intonation phrase boundary prosodies					
			H*			H*] <sub>L-H%</sub>
			H*	]	<sub>L-H%</sub>	H*
Late (transitive)	Whenever	the	lady	loads	the	car
Equi	Whenever	the	lady	checks	the	room
Early (intransitive)	Whenever	the	lady	moves	the	door

object than with, and an equi-bias set (i.e., the original set from Kjelgaard & Speer, 1999) containing both verb types and few heavily biased verbs.

Blodgett's (2004) experiment demonstrated that even when auditory fragments contained verbs such as *loads*, naming times to the early closure target *is* were significantly shorter than naming times to the late closure target *it's* following auditory fragments that contain an intonation phrase boundary immediately following the verb.

Interestingly, Blodgett (2004) did not replicate Kjelgaard and Speer's (1999) finding that *intermediate* phrase boundary location can also determine the initial syntactic structure for the ambiguity in Table 2. This is perhaps because Blodgett's ip auditory fragments – even those containing verbs that occur more often without a direct object than with – were biased toward a late closure structure and interpretation, while Kjelgaard and Speer's were not. Of course, others have reported effects of intermediate phrasing on sentence comprehension (e.g., Carlson et al., 2001; Schafer, 1997). For example, Schafer (1997) demonstrated that the relative locations of intermediate phrase boundaries influenced the interpretation of sentences with globally ambiguous prepositional phrase attachments. The intermediate phrasing in (8) resulted in a higher proportion of verb phrase modification responses (61.5%) than did the intermediate phrasing in (9) (44.3%).

8. [[The bus driver angered the rider <sub>L-</sub>] [with a mean look. <sub>L-</sub>] <sub>L%</sub>]

9. [[The bus driver angered <sub>L-</sub>] [the rider with a mean look. <sub>L-</sub>] <sub>L%</sub>]

Immediate effects of prosodic phrasing on syntactic parsing have also been shown for prepositional phrase attachment ambiguities such as *Tap the frog with the flower*. Snedeker and Trueswell (2003, Experiment 3) collected eye-movement data from listeners in a toy-moving task. Spoken stimuli were produced by naïve speakers, who produced both high- and low-attached PP sentences as instructions. On each trial, speakers were shown toy displays consistent with both syntactic forms, such as a flower instrument, a plain frog, a frog holding a little flower, a distracter animal, and a distracter instrument. The experimenter acted out the desired toy manipulation, either using the

flower to touch the plain frog, or using her hand to touch the frog with the small flower. Speakers then saw a relevant text sentence, which was removed and then produced from memory. Results from analyses of duration and ToBI annotation showed clear use of prosodic phrasing to indicate syntactic attachment. Speakers' intended high-attachment utterances were more likely to be pronounced with a boundary after the direct object (e.g., "*frog*") than those intended as low attachment.

Listeners' behavior was consistent with the speakers' intended instructions. That is, when speakers placed the strongest boundary after *frog*, listeners were most likely to use the flower as an instrument to tap the plain frog, and when speakers placed the strongest boundary after *tap*, listeners tapped the frog holding the flower. Most interestingly, eye-movement data showed that the speakers' prosody affected interpretation prior to the onset of the ambiguous PP, suggesting that prosodic phrasing not only influences initial parsing but can also be used to predict upcoming spoken material. More specifically, when speakers intended the high-attached instrumental meaning, and no prosodic boundary occurred prior to the direct object noun, listeners looked to both the plain and flower-holding frogs during their response to the spoken noun phrase (NP). In contrast, when speakers intended the low-attached modifier meaning, placing a boundary before the noun, listeners looked mostly at the frog holding a flower. This suggests they had used prosody to eliminate the plain frog as the potential referent of the NP, even before hearing the modifier.

Several recent studies of English show early effects of prosodic phrasing for syntactic structures not previously discussed here. Schepman and Rodway (2000) demonstrated with cross-modal naming that prosodic boundaries influence the attachment of structurally ambiguous relative clauses following coordinated NPs. While the researchers claim that some of their auditory stimuli contained only intermediate phrase boundaries at the critical juncture, the length of the mean pause duration (289 ms) that they report in their acoustic measurements raise doubt that their stimuli contained anything but intonation phrase boundaries. Watson and Gibson (2005) provided only an incomplete prosodic description of auditory stimuli that they created through splicing together parts of different utterances. However, their study does represent some of the only work to demonstrate early effects of intonation phrase boundaries on the processing of syntactically unambiguous prepositional phrase attachments using an online task, namely, lexical decision.<sup>4</sup>

Other recent studies in English, including Kang and Speer's (2004) demonstration that intonation boundary location influences the interpretation of globally ambiguous participle constructions (e.g., *Aaron followed a poor guy drinking his soda*) and Watson and Gibson's (2004) demonstration that intonation boundaries influence the ease of comprehension for head-dependency relationships in unambiguous sentences (e.g., *An artist*

---

<sup>4</sup> See Boland (1993) for evidence supporting the claim that cross-modal naming is sensitive to syntactic processing, while lexical decision is sensitive to syntactic and semantic processing.

*arranged a donation of the paintings of the landscape to the museum*), have relied on off-line end-of-sentence judgment tasks that can say little regarding the time course with which prosodic representations influence language comprehension and little regarding the execution of processing mechanisms incrementally.

## 6. PRODUCTION CONDITIONS AND PROSODIC PHRASING

Evidence from language comprehension studies represents only half of the language processing picture. Although it is important to demonstrate that listeners use prosodic structure to process spoken language, it is just as important to demonstrate that speakers actually produce in everyday speech the types of prosodic structures that are manipulated in laboratory investigations. Although an extensive literature examines the production of prosodic form, for the purposes of this chapter we limit ourselves to the discussion of production techniques used to create stimuli for studies of spoken language comprehension.

The debate in the literature on this topic was opened by a paper by Allbritton et al. (1996). They examined naïve speakers' productions of a range of ambiguities previously shown to be resolvable by prosodic information (taken from Price et al., 1991). They showed that while speakers who were instructed to disambiguate prosodically did so, those who were first presented with short disambiguating paragraph contexts, and no instruction to disambiguate, did not. On this basis, they claimed that comprehension studies using "laboratory speech" stimuli – carefully selected utterances involving a prosodic feature that contrasts minimally in the compared conditions – do not represent typical language processing. They suggested that speakers' awareness of ambiguity, stemming from instructions to disambiguate, or from the contextual presence of ambiguity, is the primary factor that influences the saliency of prosodic contrasts in their productions. However, other researchers suggested that because Allbritton et al.'s studies used text stimuli, they did not address directly the issue of whether naïve speakers and listeners make use of prosodic distinctions in "typical" spoken language processing. Readers and talkers have different pragmatic goals and processing demands, resulting in well-documented differences between spontaneous and read speech (e.g., Ayers, 1994; Blaauw, 1994). Although many studies of prosodic processing continue to use "laboratory speech," several recent studies have developed novel techniques to elicit speech tokens from participants. These techniques vary in their ability to elicit spontaneous speech and multiple tokens of the same type of utterance from the same speaker.

Schafer et al. (2003; see also Schafer, Speer, Warren, & White, 2000; Speer, Warren, & Schafer, submitted; Warren, Schafer, Speer, & White, 2000) elicited quasi-spontaneous speech via a cooperative board game that they designed for pairs of participants to play. One participant – the driver – knew the origin of various game pieces, as well as their ultimate goal locations. The driver instructed the other participant – the slider – as to which pieces to move. The slider knew the locations of hazards (ravenous goats) and bonuses (cookies), and he or she used that information to ask questions, provide confirmations,

and decide on what direction to move any given piece. Throughout the game, participants used a fixed set of sentence frames and object names to build their conversation and to navigate the board. These included utterances for acknowledging correct moves and correcting missteps. Because participants worked to accumulate points, they used the utterances to achieve a joint communicative goal. At the same time, because participants were limited to a particular set of utterances, the researchers obtained multiple repetitions of the same sentence type from the same speaker. The utterances in (10) form a sample of dialog from an actual game.

10. Driver: I want to change the position of the square with the triangle.  
 Slider: Which triangle do you want to change the position of the square?  
 Driver: The red one. When that moves the square it should land in a good spot.  
 Slider: Good choice. When that moves the square will encounter a cookie.

There are three constructions of interest in this particular dialog. First, the driver's initial utterance contains a globally ambiguous prepositional phrase attachment ambiguity. The PP *with the triangle* can either modify *the square* and indicate a house-shaped game piece or it can modify *to change* and indicate that a triangle is needed to move a square. Second, the driver and slider's second utterances each contain a temporary early/late closure ambiguity (i.e., *When that moves the square ...*). Third, the driver and slider's initial utterances each contain a *want to* sequence. In the slider's utterance, a *wh*-gap occurs between *want* and *to*. However, in a comparable utterance not represented in this dialog (i.e., *Which triangle do you want to change the position of this time?*), a gap does not occur.

In addition, Schafer et al. (2000) designed their game so that the degree of referential ambiguity varied. That is, some of the time the arrangement of game pieces disambiguated utterances that were syntactically ambiguous. At other times, the arrangement of game pieces did not disambiguate utterances. Ultimately, Schafer et al. conducted prosodic and acoustic analyses of all driver and slider utterances in the varying referential contexts. Results showed that drivers and sliders consistently used prosodic boundaries to mark syntactic boundaries. For example, speakers produced an intonation phrase boundary at the end of the subordinate clause in 53 out of 55 productions of an utterance with late closure syntax (e.g., *When that moves the square it should land in a good spot*) and also in 43 out of 53 productions of an utterance with early closure syntax (e.g., *When that moves the square will encounter a cookie.*). Even the remaining 10 early closure utterances contained an intermediate phrase boundary in the same location. In addition, speakers produced these prosodic phrasings regardless of whether a game situation disambiguated syntactically ambiguous utterances.

Kraljic and Brennan (2005) also had participants work in pairs, producing spontaneous speech in a toy-moving task. Speakers were given no text, and instead used a card with graphic images to create a spontaneous instruction (e.g., *Put the dog in the basket on the star*). Toy arrays were used to create situations that were ambiguous (e.g., objects included a dog in a basket, another basket on a star, and another star) and unambiguous

(e.g., only a dog and a basket on a star). Participants produced a variety of sentence forms, but the majority of utterances took the syntactically ambiguous form of interest to the researchers (V NP PP PP). Analysis of duration showed that speakers produced prosodic boundaries that disambiguated the utterances regardless of whether the situational context was ambiguous. In addition, Kraljic and Brennan (2005) measured how aware speakers were of the situational ambiguity. As it turns out, many speakers had difficulty assessing accurately whether situations were ambiguous or not. They noted that if speakers fail in their assessment of situational ambiguity, then it is not surprising that it does not seem to influence the prosody of speakers' utterances.

In contrast to Schafer et al. (2000) and Kraljic and Brennan (2005), however, the results of Snedeker and Trueswell (2003, Experiments 1 & 2) suggest that speakers are more likely to produce disambiguating prosody when a visual scene does not disambiguate a syntactically ambiguous utterance. Recall that in their production/comprehension task, discussed above, an experimenter performed one of the possible actions for an ambiguous sentence like *Tap the frog with the flower*. At the same time, the speaker saw a card containing the text of the sentence. The card was then removed, and the action repeated. The speaker then gave a verbal instruction to another listener, with the goal of getting the listener to perform the same action as the experimenter.

Across experiments, productions were compared in ambiguous and unambiguous contexts. In the ambiguous case, speakers and listeners had access to a frog, a flower, and a frog that held a flower. Speakers used prosody to signal their syntactic intentions: when they meant for listeners to use a flower to tap a frog, they tended to produce a larger prosodic break before the prepositional phrase than before *the frog*, and when they meant for listeners to use their hand to tap the frog that had the flower, they tended to produce a larger prosodic break before *the frog* than before the prepositional phrase. In this condition, listeners were able to repeat the appropriate action approximately 70% of the time.

In unambiguous displays, however, speakers had access to objects that were consistent with only a single syntactic structure. That is, when listeners were supposed to use a flower to tap a frog, objects in the speakers' display included a flower and only a single frog. When listeners were supposed to use their hand to tap a frog holding a flower, objects in the speakers' display include two frogs, one of which had a flower, but no solitary flower object. Listener displays still contained objects that allowed both actions. In this condition, listeners failed to repeat the appropriate action at better than chance levels, and although speakers produced some prosodic disambiguation, it was far weaker than that which had occurred in the condition with ambiguous displays. The authors concluded that situational context influenced the production of prosodic disambiguation, and that listeners make use of prosodic phrasal cues when they are (however unreliably) available.

Taken together, these studies suggest that awareness of ambiguity or rather awareness of a lack of ambiguity could influence the prosody that speakers produce. When speakers' attention was drawn to the ambiguity in the Snedeker and Trueswell (2003) study,

they were more likely to disambiguate, and they disambiguated more strongly than when there was no apparent ambiguity. However, the Kraljic and Brennan (2005) study suggests that speakers often may not develop such awareness automatically in conversational contexts. Their results, along with those of Schafer et al. (2000), suggest that production of prosodic phrasing may be driven by grammatical constraints on the processing system rather than by situational context. Several outstanding issues remain in this controversy. Differences among the studies include the presence/absence of text sentences, the relative complexity and length of the prepositional phrase attachment sentences used, and the extent to which participants engaged in a conversation, as well as differences in the precise nature of instructions and task demands. However, it appears that speakers most frequently produce a prosody consistent with the meaning they intend, and that disambiguation is not restricted to situations where they are aware of their listener's needs. In turn, listeners make use of the prosodic structure provided by speakers to determine basic syntactic constituency relations.

A typical criticism of the traditional reading task is that speakers end up producing utterances in the absence of any conversational or communicative intent. Different researchers have attempted to rectify that problem, while preserving the relative ease with which materials can be constructed, recorded, and analyzed. For example, Fernandez and Bradley (2004) employ what they call the "New York Post to New York Times" technique. Participants read simple sentences and combine them to produce utterances that contain a complex sentence. Although the task is still non-conversational, participants have a greater responsibility for constructing the target sentences. Watson and Gibson (2004) use what they call a "two-participant" reading technique in which the speaker is aware that there is a listener who will have to answer a comprehension question based on the speaker's utterance. This task is somewhat similar to the Snedeker and Trueswell (2003) design in the sense that the task is still non-conversational, but the speaker is aware that a listener must perform some task on the basis of the speaker's utterance.

The importance of studies that demonstrate effects of prosody on spontaneous speech production and comprehension should not be underestimated, but these studies and studies with carefully controlled lab speech are perhaps best viewed as complementing each other. Whereas spontaneous speech studies provide insight into the nature of speech production and comprehension processes "in the wild" – or at least in the zoo – lab speech studies often provide the item power that comes from multiple repetitions of the same utterance type.

## **7. THE INFLUENCE OF A PROSODIC REPRESENTATION ON SYNTACTIC PARSING**

Several recent studies indicate that prosodic phrasal effects stem from utterance-level prosodic representations. End-of-sentence judgment tasks have been particularly useful for demonstrating that it is the global prosodic representation – not just the presence of a local break between the words of a sentence – that plays a pivotal role in determining

structures and interpretations (e.g., Carlson et al., 2001; Clifton, Carlson, & Frazier, 2002; Frazier, Clifton, & Carlson, 2004; Schafer, 1997; Schafer et al., 2000, 2003; Speer et al., 2003).

In a series of experiments, Carlson et al. (2001) manipulated the relative size of prosodic boundaries in globally ambiguous utterances like those in (11)–(14). In these utterances, the temporal phrase *after John visited* could indicate when Susie learned something or when Bill telephoned.

- |   |  |                     |
|---|--|---------------------|
| 11. [[Susie learned <sub>L-</sub> <sub>H%</sub> ] | [[that Bill telephoned <sub>L-</sub> <sub>H%</sub> ] | after John visited. |
| 12. [[Susie learned <sub>L-</sub> <sub>H%</sub> ] | [[that Bill telephoned <sub>L-</sub>                 | after John visited. |
| 13. [[Susie learned <sub>L-</sub> <sub>H%</sub> ] | [[that Bill telephoned                               | after John visited. |
| 14. [[Susie learned <sub>H-</sub>                 | [that Bill telephoned <sub>L-</sub>                  | after John visited. |

When there was no difference in the size of the prosodic boundaries preceding the temporal phrase, as in (11) and (14), listeners were more likely to say that the temporal phrase indicated when Susie learned something (21% and 25% of judgments, respectively). This was in comparison to the utterances in which a larger prosodic boundary preceded a smaller boundary, as in (12) and (13). In these conditions, listeners judged the temporal phrase as modifying when Susie learned something only 14% and 15% of the time, respectively. Thus, it was not merely the presence of the local prosodic boundary just prior to the point of syntactic ambiguity that influenced how the temporal phrase was parsed and interpreted. Rather, it was the combination of prosodic boundaries that influenced comprehension.

Similar results come from studies using quasi-spontaneous speech obtained in a conversational setting. Recall that Schafer et al. (2000, 2003) analyzed productions of temporarily ambiguous utterances like those in (15) and (16) (repeated from example (10)), which were obtained from participants engaged in a cooperative board game.

15. When that moves the square it should land in a good spot.
16. When that moves the square will encounter a cookie.

Phonetic and ToBI analyses showed that speakers consistently produced a stronger prosodic boundary after *moves* relative to the boundary produced after *the square* in utterances in which *the square* functioned as the subject of the main clause (i.e., in (16)). These productions were then used in a forced-choice task. A new group of listeners used the relative break size to correctly identify how speakers had completed truncated versions of the utterances in (15) and (16). This finding is consistent with the idea that listeners are actively recovering and making use of a global prosodic representation.

Blodgett (2004) demonstrated that even Kjelgaard and Speer's (1999) finding that the location of an intonation phrase boundary can determine the initial structure for the



temporary syntactic ambiguity in (17) is dependent in part on the global prosodic representation. More specifically, it depends in part on the presence of the incomplete intonation phrase that ends the auditory fragment in (17).<sup>5</sup>

- H\*                      H\*
17. [[Whenever the lady loads <sub>L</sub>.] <sub>H%</sub>] [[the car
- H\*
18. [[Whenever the lady loads <sub>L</sub>.] <sub>H%</sub>]

Recall that naming times to the early closure target *is* were significantly shorter than naming times to the late closure target *it's* following auditory fragments like the one in (17). These fragments contained an intonation phrase boundary immediately following the verb, and the effect held even when the verbs in the fragments occurred more often with direct objects than without. However, when participants were asked to name noun phrases that could serve as plausible direct objects following the auditory fragment in (18), they were just as likely to provide a main clause sentence completion as they were to provide a direct object. Thus, it is not the case that an intonation phrase boundary at that location *by itself* resolved the syntactic ambiguity in favor of an early closure structure. If that had been the case, sentence completions in response to fragments like (18) should have overwhelmingly contained early closure structures. Thinking ahead to the section on models of prosody and processing, approaches that appeal to local cues instead of global prosodic representations are likely to provide inadequate explanations of prosody's effect in comprehension.

## 8. PROMINENCE IN PROCESSING AND ITS INTERACTION WITH PHRASING

As reviewed above, studies of the role of prosodic phrasing on syntactic interpretation during language comprehension investigate the influence of just one component of prosodic structure. When studies involve lab speech, experimenters frequently control the type and location of accentual information in order to limit their investigation to the role of prosodic phrasing. However, many recent studies in English have manipulated pitch accent information by itself, or simultaneously manipulated pitch accent and phrasing information. Two broad conclusions are suggested from the work on accent and phrasing discussed in this section.

1. Pitch accent placement and type reliably affect the focus and information structure of an utterance, and some contrastive accents may even be used predictively in discourse.
2. Pitch accent placement, once thought to influence only focus or pragmatic relations, can also affect the syntactic interpretation of a sentence.

<sup>5</sup> Contrary to Watson and Gibson (2004), an auditory stimulus that is truncated mid-prosodic phrase does not necessarily serve as a cue to a prosodic boundary.

## 9. PITCH ACCENT PLACEMENT

Pitch accent placement has been shown to affect sentence comprehension by conveying the focus and information structure of an utterance. For example, Birch and Clifton (1995) asked participants to rate question–answer dialogs either for the appropriateness of the intonation or for the appropriateness of the meaning. A sample set of materials is shown in Table 3.

Because *new* questions did not mention math in the question, the mention of math in the answer represented new information. Because *given* questions did mention math, the repeated mention of math in the answer represented given information. The two question types were crossed with three accentual patterns in the answers: V NP, NP, and V. In the first two patterns, the word *math* was accented. In the V NP pattern, the preceding verb was accented as well. In the V pattern, only the verb was accented.

The results provided evidence that new argument noun phrases should be accented. Listeners rated the intonation of the V NP and NP answers as more appropriate than the intonation of the V answers following *new* questions. Their “makes sense” judgments showed the same pattern. In addition, reaction times for the “makes sense” judgments were significantly shorter when the NP was accented (218 ms) than when it was not (255 ms). The results also provided evidence that given argument noun phrases should not be accented. Participants rated the intonation of the V answers as more appropriate than the intonation of the V NP and NP answers following *given* questions. Again, “makes sense” judgments showed a similar pattern. However, the results further demonstrated that accenting the verb when the given argument noun phrase was accented, as in the V NP condition, significantly increased acceptability in intonation and meaning and decreased “makes sense” judgment times relative to the NP condition. Thus, it is not simply the case that given or new status corresponds directly to the presence and absence of accents. This finding is consistent with more recent work in language production that indicates that a simple parallel association between deaccentuation/accentuation and the given/new status of a word cannot account for accent distribution (Bard & Aylett, 1999).

Table 3  
Sample question – answer stimuli from Birch and Clifton (1995).

Question types		Answer types	
			L+H* L- L+H* L-L%
		V NP	Yes, she TEACHES MATH.
New	Isn't Kerry pretty smart?		L+H* L-L%
Given	Isn't Kerry good at math?	NP	Yes, she teaches MATH.
			L+H* L-L%
		V	Yes, she TEACHES math

Recent work (Dahan, Tanenhaus, & Chambers, 2002) suggests that English pitch accent information is used immediately as listeners single out the object referent of a word in a spoken sentence. Participants in an eye-tracking experiment viewed object displays including a triangle, a candle, and a candy. They listened to pre-recorded sequences like those in (19), first hearing an instruction to move either the candle or the candy (19a), and then hearing one of two following instructions, one with an accented target word (CANDLE, 19b), and the other with an unaccented target (candle, 19c) (CAPITAL letters indicate words carrying a contrastive L+H\* accent).

19. a. Put the candle/candy below the triangle.  
 b. NOW put the CANDLE above the square  
 c. Now put the candle ABOVE THE SQUARE.

Eye movements to candle and candy objects were measured during the initial syllable of “candle” in the second utterance. Results showed that when listeners had heard the word “candle” in the first utterance, they initially looked more often at the candy during the accented “CAN” syllable in the second utterance. When they had heard “candy” in the first utterance, they initially looked more often at the candle during the accented CAN syllable in the second utterance. Effects were reversed when the deaccented “can” syllable was presented in the initial utterance of the sequence. Interestingly, listeners were able to visually isolate the referent even before the onset of its name, suggesting that in discourse context, the overall intonational structure of an utterance can be used to predict the upcoming object name. However, this very early use of intonation may have been due to the L+H\* accent on the initial “NOW” in the critical sequence (... candy ... NOW ... CANDle), rather than to the presence of an appropriate contrastive accent on the target word.

Ito and Speer (submitted) used eye movements to investigate the effect of pitch accent placement as listeners recognize and evaluate prosodic information and use it to guide visual search. Participants followed pre-recorded auditory instructions to decorate holiday trees using ornaments organized by object category on a grid (e.g., balls, angels, eggs, drums, bells, onions, candies, and stockings in a variety of colors). The majority of instructions were given with intonation patterns modeled on a preceding English production study using the same task. These patterns were meant to appropriately convey the information status of words. Occasionally, however, participants heard instructions with infelicitous intonation patterns. Consistent with Dahan et al. (2002) results showed that felicitous intonational prominence facilitated fixations to the target ornament as compared to infelicitous intonational prominence. In particular, when a prominent pitch accent was assigned to a word that conveyed contrastive discourse information, looks to the target occurred sooner and more often. For example, a L+H\* accent on “BLUE” in instructions such as “*First, hang the orange ball .... Next, hang the BLUE ball,*” produced shorter latencies to initially fixate the cell of balls on the ornament grid than a comparison condition where the L+H\* accent was produced infelicitously on the noun (e.g., “*Next, hang the blue BALL*”). This indicates that the intonation pattern of a modifier may help listeners predict the upcoming noun. While the

Dahan et al. results demonstrated that appropriate and inappropriate contrastive accents affected looks to the object associated with the target word, the Ito and Speer results showed that such accents can also influence the discourse status of an as-yet unmentioned object. Both findings implicate pitch accents in the structuring of discourse information relations during conversation.

## 10. INTERACTIONS OF PITCH ACCENT AND PHRASING

Pitch accent location has recently been shown to influence syntactic processing as well as the processing of information and focus structure. Schafer et al. (2000) demonstrated that pitch accent location could influence the resolution of globally ambiguous sentences like the one in (20). On the one hand, the phrase *who's cold* can function as a relative clause and indicate that the pretty little girl is cold. On the other hand, it can function as an embedded question and describe the question that was asked.

20. I asked the pretty little girl *who's cold*.

In a series of experiments, a single H\* pitch accent was placed on either word in the phrase *who's cold*. Results showed that the proportion of embedded clause judgments increased when the H\* occurred on *who's* relative to when it occurred on *cold*. The researchers attributed the finding to effects of focus. That is, because *wh-* questions are typically focused, and because high pitch accents commonly indicate focus in English, the occurrence of a high pitch accent on the *wh-*word in (20) contributes to an embedded question structure and interpretation.

Pitch accent location had a similar influence on the structure and interpretation of ambiguous relative clause attachments like the one in (21). In these constructions, the relative clause can modify either *the propeller* (N1) or *the plane* (N2).

21. The sun sparkled on the propeller of the plane that the mechanic was so carefully examining.

Schafer, Carter, Clifton, and Frazier (1996) demonstrated that the proportion of N2 attachment judgments increased when a H\* – the only pitch accent in the *N1 of N2* sequence – occurred on N2. Maynell (1999) replicated this effect using contrastive pitch accents. Participants judged a L+H\* accent on N1 as indicating an N1 attachment 83% of the time, and they judged a L+H\* accent on N2 as indicating an N1 attachment only 71% of the time. Maynell further demonstrated that prosodic phrasing also played a role in comprehension. When utterances contained a prosodic boundary just prior to the relative clause, participants overwhelmingly judged utterances as indicating N1 attachments. Yet, when utterances did not contain a prosodic boundary in the same location, they were more likely to judge utterances as indicating N2 attachments than N1 attachments. Thus, listeners used both pitch accent and prosodic phrasing information to determine the structure and meaning of the utterance.

Just as previous studies demonstrated the importance of the global prosodic representation when it comes to effects of prosodic phrasing, Carlson (2001) demonstrated the importance of the global representation when it comes to effects of pitch accents. Her experiments investigated in part the effects of *parallelism* between pitch accents, where parallelism was defined in terms of pitch accent type, location, and peak height. Carlson demonstrated that this parallelism influences the resolution of globally ambiguous sentences like those shown in Table 4.<sup>6</sup> If the utterance is interpreted as a coordinate structure, Bob has insulted the guests and Sam, and Josh has visited the office and Sarah. However, if the utterance is interpreted as a gapping structure, Bob and Sam have each insulted the guests, and Josh and Sarah have each visited the office, albeit at different points in time.

In Carlson’s (2001) two cooperating prosodies, the pitch accents for the ambiguous phrase (i.e., *and Sam during the dance*) were parallel to the pitch accents in the beginning conjunct. For example, the L\* at *Sam* and the L+H\* at *dance* correspond to *Bob* and *dinner* when there should be a gap, and to *guests* and *dinner* when there should be no gap. In a sentence comprehension task, participants judged utterances with cooperating prosody for the gap syntax to be gap structures significantly more often (44%) than utterances with cooperating prosody for no gap syntax (28%). Carlson also included a baseline prosody that did not contain pitch accents that were parallel in terms of type and location. This baseline prosody was intended to serve as a prosody that was felicitous for either a gap structure or a no gap structure, similarly to the baseline prosody in Kjelgaard and Speer (1999). Consistent with this interpretation, participants judged utterances with baseline prosody to be gap structures at a proportion that is numerically intermediate (38%) between the two cooperating conditions and also similar to findings from Carlson’s experiment with written versions of her sentences. That is, the baseline prosody seemed to allow both interpretations without biasing listeners toward either one.<sup>7</sup>

Table 4  
Summary of the three prosodic conditions from Carlson (2001).

Cooperating prosody for gap syntax:				
L* H-		L+H* L-H%	L* H-H%	L+H* L-L%
Cooperating prosody for no gap syntax:				
H*	L* H-	L+H* L-H%	L* H-	L+H* L-L%
Bob	insulted the guests	during dinner	and Sam	during the dance
Baseline prosody:				
H* L-	L+H* L-	L+H* L-H%	H* !H-L%	L+H* L-L%
Josh	visited the office	during the vacation	and Sarah	during the week

<sup>6</sup> Carlson (2001) was also investigating parallelism in terms of verb object plausibility and animacy. This design is part of the reason behind the different sentences in the cooperating and baseline prosodies.

<sup>7</sup> Baseline prosodies represent a major challenge for developing stimuli. Carlson points out that the prosodic phrasing in the baseline might not support both structures equally, and she provides no appropriateness norms to justify the baseline attribute.

## 11. MODELS OF PROSODY IN PROCESSING

So, how does prosody influence processes of sentence comprehension? For starters, researchers are wise to point out that prosodic structure is one component in the overall well formedness of an utterance (e.g., Jun & Schafer, 1993; Kjelgaard & Speer, 1999; Schafer et al., 1996). While prosody can often disambiguate syntactically ambiguous utterances, it does not necessarily follow that its purpose or function is to resolve ambiguity. Indeed, Kraljic, and Brennan (2005) demonstrated that speakers are frequently unaware of whether an utterance is ambiguous in a particular situation. Presumably, correspondences among the various linguistic representations – such as prosody, syntax, semantics, discourse – are grammatically constrained. Thus, any processing account that treats prosodic representations simply as tools to ambiguity resolution or as representations that only show their influence during ambiguity resolution will likely fall short.

Here, we compare two processing accounts – Schafer (1997) and Blodgett (2004) – that start at the phonological processor. Following Speer, Kjelgaard, and Dobroth (1996), the phonological processor builds an abstract prosodic representation. This representation helps preserve lexical information in memory, but is not itself dependent on lexical information. The prosodic representation serves as input to the syntactic and semantic processors. Because processing at this level – as at all levels – is incremental, the prosodic representation is constantly updated and available to influence processing at other levels.

In both of these prosody-first accounts, intonation phrase boundaries trigger wrap-up of any outstanding processing. According to Schafer (1997) *interpretive domain hypothesis*, intonation phrase boundaries trigger wrap-up of any semantic or pragmatic processing that happens to be incomplete at the time an intonation boundary occurs. Schafer demonstrated experimentally that when clauses containing lexically ambiguous nouns (e.g., *glasses* in *Although the glasses were ugly ...*) were disambiguated toward their less frequent meaning (*they held a lot of juice*), comprehension times were significantly longer when an intonation phrase boundary separated the clauses than when an intermediate phrase boundary separated them. When the lexically ambiguous nouns were disambiguated toward their more frequent meaning (*Stacey wore them anyway*), comprehension times were not affected by the type of prosodic boundary. Blodgett (2004) found similar effects of intonation boundaries for the semantic resolution of verbs that participate in thematic role assignment ambiguities. In addition, Blodgett (2004) provided evidence that intonation phrase boundaries trigger wrap-up of any outstanding syntactic processing as well.

According to Schafer (1997) and Blodgett (2004), intonation phrase boundaries influence ambiguity resolution processes via a processing mechanism that forces integration and interpretation of various linguistic structures. In addition, intonation phrase boundaries trigger this integration mechanism regardless of whether there are ambiguities to be resolved. In this way, ambiguity resolution is simply a by-product of these wrap-up mechanisms. Watson and Gibson (2005) take a somewhat different approach with their anti-attachment hypothesis. They propose the existence of a parsing strategy

in which the parser prefers not to attach incoming material to lexical content that immediately precedes an intonation phrase boundary. They argue that the strategy reflects in part constraints on the likelihood of speakers producing intonation phrase boundary during production. To their credit, the link between processes of production and comprehension is one that deserves greater attention in work on prosody and processing. However, the anti-attachment hypothesis essentially treats intonation phrase boundaries as a local cue to syntactic attachment decisions. Furthermore, in related work, Watson and Gibson (2003) conflate intonation phrase boundaries and intermediate phrase boundaries in analyses, thus making it hard to know whether the anti-attachment hypothesis applies to both types of prosodic phrase boundaries in English, and if so, whether it applies equally, and if not, why not.

In contrast to Watson and Gibson (2005), Schafer (1997) and Blodgett (2004) discriminate between effects of intonation phrase boundaries and intermediate phrase boundaries. According to Schafer (1997), intermediate phrase boundaries influence the availability of syntactic attachment sites via the prosodic visibility hypothesis. That is, attaching an incoming syntactic node into an existing structure is easier when the attachment site is within the same intermediate phrase as compared to when it is not. In addition, successive intermediate phrases gradiently reduce the visibility between an incoming syntactic node and its potential attachment sites. As a result, it is possible for disambiguating syntactic information to become available before an attachment decision can be made in the context of an otherwise syntax-first garden path model (e.g., Frazier, 1990; Frazier & Clifton, 1996). In turn, this reduces the number of garden path experiences. For example, given the prosodic phrasing for the early/late closure ambiguity in (22), the intermediate phrase boundary could delay attachment of *the house* long enough for a main verb to appear and disambiguate the utterance.

22. [[When Roger leaves <sub>H<sub>c</sub></sub>]            [the house

Blodgett (2004) takes a different approach by considering the goodness-of-fit between the prosodic representation (specifically, the prosodic phrasing contained therein) and one or more syntactic representations in the context of a constraint-based lexicalist model (e.g., Boland, 1997; Trueswell, Tanenhaus, & Garnsey, 1994). The more closely intermediate phrases align with the constituent structure of syntactic alternatives, the more weight the prosodic representation adds to those syntactic alternatives.

Blodgett's (2004) goodness-of-fit approach can account for the prosodic visibility data, and it also might help solve problems that arise when attachment preferences are explained in terms of what Watson and Gibson (2005) call "the domain hypothesis." According to this hypothesis, processing difficulty increases if an intonation phrase boundary separates two dependent heads. Watson and Gibson are right when they say that the types of approaches in Kjelgaard and Speer (1999) and Schafer (1997) cannot explain why the intonation phrase boundary in (23) biases a listener toward a structure and interpretation in which the cop used the telescope to see the spy. This is because the two attachment sites share the same prosodic visibility.

## 23. [[The cop saw the spy]] [[with the telescope]]

However, because the anti-attachment hypothesis appeals only to local break information, Watson and Gibson (2005) are likely to encounter a similar problem in trying to explain differences in attachment preferences for utterances like those in (24) and (25) (from Schafer, 1997). Although both utterances contain a prosodic phrase boundary before the ambiguous prepositional phrase, (24) resulted in a higher proportion of verb phrase attachment structures and interpretations (61.5%) than (25) did (52.6%). If only the break that is local to *the rider* matters, anti-attachment makes no prediction. If the breaks that are local to both lexical heads matter (i.e., *angered* and *the rider*), anti-attachment again makes no prediction.

24. [[The bus driver angered the rider<sub>L</sub>] [with a mean look<sub>L</sub>] L%]25. [[The bus driver<sub>L</sub>] angered<sub>L</sub>] the rider<sub>L</sub>] [with a mean look<sub>L</sub>] L%]

Watson and Gibson's (2005) anti-attachment hypothesis can also be compared with Carlson et al.'s (2001) informative break hypothesis. According to the latter account, the parser finds prosodic boundaries of the same size to be uninformative during ambiguity resolution. In contrast, a [smaller break/larger break] pair or a [larger break/smaller break] pair could influence attachment decisions. For example, the unequal boundaries in (26) suggest that the ambiguous temporal phrase *after John visited* modifies the telephoning event.

26. [[Susie learned<sub>L</sub>] H%]                      [[that Bill telephoned<sub>L</sub>] after John visited.

While the informative break hypothesis provides at least a descriptive account of the effects of phrasing on attachment decisions, it too has several weaknesses. For starters, Schafer's (1997) prosodic visibility hypothesis (and Blodgett's (2004) goodness-of-fit approach) can each account for the same results without having to suggest that the syntactic parsing is only sensitive to prosodic phrasing information when it is "informative." And while appealing to pairs of boundaries does allow the global prosodic representation to influence processing, as it is currently devised, the informative break hypothesis makes no predictions when an initial prosodic boundary is encountered during incremental processing. For example, the informative break hypothesis does not predict the resolution of the classic early/late closure ambiguity in (27).

27. [[When Roger leaves<sub>L</sub>] H%]    [[the house

## 12. CONCLUSIONS

Research on prosody over the past 20 years has broadened our understanding of language comprehension in ways that would not have been possible based solely on the study of written language. Indeed, the research on prosody and processing raises doubts that effects observed in reading necessarily correspond to similar effects in auditory comprehension. The same can be said of the underlying cognitive processes that researchers infer from



those effects. We anticipate that one possible shift for the field will come with the advent of improved methodologies for tapping prosodic structures during silent reading.

The research has shown that elements of prosody including phrasing and prominence have a very early influence on the recognition of a variety of linguistic mental representations. Such research raises questions for processing accounts that treat the syntactic representation as the driving force behind comprehension. We have suggested that precise specification of the prosodic aspects of experimental stimuli combined with the exploitation of prosodic ambiguity can do much to advance our investigation of spoken language processing.

More generally, processing accounts have a long way to go in terms of explaining how the multiple components in prosodic representation work together to influence various other levels during sentence comprehension. There are also interesting questions regarding to what extent segmental, lexical, syntactic, semantic, and discourse factors may influence the perception of prosody, and how that might come about. Although we have focused on studies that investigate English, successful processing accounts will be those that explain similarities and differences in processing across languages that have similar and different prosodic systems.

## REFERENCES

- Allbritton, D., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 714–735.
- Ayers, G. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics*, 44, 1–49.
- Bard, E., & Aylett, M. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, August, 1753–1756.
- Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In: P. Bosch, & R. van der Sandt (Eds), *Focus and natural language processing*. *IBM Working Papers in Logic and Linguistics*, 6, 1–10, Heidelberg:IBM.
- Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17–67.
- Beckman, M., & Ayers, G. (1994). Guidelines for ToBI labelling. Unpublished MS. Ohio State University. Version 3. March 1997. Downloadable ms. [[http://ling.ohio-state.edu/Phonetics/etobi\\_homepage.html](http://ling.ohio-state.edu/Phonetics/etobi_homepage.html)]. For information on obtaining by ftp, send e-mail to [tobi@ling.ohio-state.edu](mailto:tobi@ling.ohio-state.edu) and visit [<http://ling.ohio-state.edu/~tobi/>]
- Beckman, M., & Jun, S. -A. (1996). *K-ToBI (Korean ToBI)* labelling convention version 2. Manuscript. Ohio State University and UCLA. Manuscript is available at [<http://www.linguistics.ucla.edu/people/jun/sunah.htm>.]

- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- Birch, S., & Clifton, Jr., C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, 38 (4), 365–391.
- Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, 14, 359–375.
- Blodgett, A. (2004). *The interaction of prosodic phrasing, verb bias, and plausibility during spoken sentence comprehension*. Unpublished doctoral dissertation, Ohio State University.
- Boersma, P. & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4.25)[Computer program]. Retrieved from <http://www.praat.org/>
- Boland, J. E. (1993). The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects. *Journal of Psycholinguistic Research*, 22, 133–152.
- Boland, J. E. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*, 12, 423–484.
- Campbell, N., & Venditti, J. (1995). J-ToBI: An intonational labelling system for Japanese. *Proceedings of the Autumn Meeting of the Acoustical Society of Japan (ASJ)*, 1, 317–318.
- Carlson, K. (2001). The effects of parallelism and prosody in the processing of gapping structures. *Language and Speech*, 44 (1), 1–26.
- Carlson, K., Clifton, Jr., C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45, 58–81.
- Clifton, Jr., C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45 (2), 87–114.
- Cooper, W., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Cutler, A., Dahan, D., & Donselaar, W. V. (1997). Prosody in comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292–314.
- Fernandez, E., & Bradley, D. (2004). *Acoustic correlates of the phrasing patterns in English and Spanish sentences containing the RC attachment construction*. Paper presented at the 5th Annual SUNY/CUNY/NYU Linguistics Mini-Conference, May, Stony Brook, NY.
- Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16 (4), 555–568.

Fodor, J. (2002). Prosodic disambiguation in silent reading. *NELS*, 32, 113–132.

Fodor, J., & Bever, T. (1965). The psycholinguistic reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, 4, 414–420.

Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In: D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds), *Comprehension processes in reading* (pp. 303–330). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, Jr., C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., Clifton, Jr., C., & Carlson, K. (2004). Don't break, or do: Prosodic boundary preferences. *Lingua*, 114, 3–27.

Garrett, M., Bever, T., & Fodor, J. (1966). The active use of grammar in speech perception. *Perception & Psychophysics*, 1, 30–32.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.

Grabe, E., Post, B., & Nolan, F. (2005). Modelling intonational variation in English. The IViE system. In: *Proceedings of Prosody 2000*, Krakow, Poland, 2–5 October.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In: Sun-Ah Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Gussenhoven, C., Reitveld, T., Kerkhoff, J., & Terken, J. (2003). Transcription of Dutch intonation. Online courseware at [<http://todi.let.kun.nl/ToDI/home.htm>]

Jun & Schafer (1993) on p. 527. This should be: Jun (2003). Jun, S-A (2003). Prosodic Phrasing and Attachment Preferences. *Journal of Psycholinguistic Research* 32(2) pp. 219–249.

Kang, S., & Speer, S. R. (2004). Prosodic disambiguation of participle constructions in English. In: B. Bel, & I. Marlien (Eds), *Speech prosody 2004 proceedings*, Nara, Japan, March.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153–194.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, 34, 391–405.

Kraljic, T., & Brennan, S. E. (2005). Using prosody and optional words to disambiguate utterances: For the speaker or for the addressee? *Cognitive Psychology*, 50, 194–231.

Ladd, R., & Schepman, A. (2003). “Sagging transitions” between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, 31, 81–112.

- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7(2), 107–122.
- Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of ambiguity resolution. *Psychological Review*, 101, 676–703.
- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology*, 45A (1), 73–87.
- Maynell, L. (1999). *Effect of pitch accent placement on resolving relative clause ambiguity in English. Poster presented at the 12th Annual CUNY conference on human sentence processing*, New York, March.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143–178.
- Miller, J. (1990). Speech Perception. In: Osherson, D. & Lasnik, H. (Eds), *Language: An invitation to cognitive science* (pp. 69–93). Cambridge, MA: MIT Press.
- Nagel, H. N., Shapiro, L. P., & Nawy, R. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research*, 23, 473–485.
- Nagel, H. N., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research*, 25 (2), 319–344.
- Nicol, J. L., & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research*, 22 (2), 207–237.
- Pickering, M. J., Traxler, M. J., & Crocker, M.W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, 43, 447–475.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation. MIT.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90, 2956–2970.
- Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11, 165–192.
- Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. Unpublished doctoral dissertation, University of Massachusetts.
- Schafer, A., Carlson, K., Clifton, Jr., C., & Frazier, L. (2000). Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and Speech*, 43 (1), 75–105.

Schafer, A., Carter, J., Clifton, Jr., C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, 11, 135–163.

Schafer, A. J., Speer, S. R., & Warren, P. (2003). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In: M. Tanenhaus, & J. Trueswell (Eds), *World situated language use: Psycholinguistic, linguistic, and computational perspectives on bridging the product and action tradition*. Cambridge, MA: MIT Press.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29 (2), 169–182.

Schepman, A., & Rodway, P. (2000). Prosody and parsing in coordination structures. *The Quarterly Journal of Experimental Psychology*, 53A (2), 377–396.

Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing. In: G. Bruce, & M. Horne (Eds), *Prosody, theory and experiment: Studies presented to Gosta Bruce*. Dordrecht: Kluwer.

Selkirk, L. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25 (2), 193–247.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of the 1992 International conference on spoken language processing*, 2, 867–870.

Slowiaczek, M. L. (1981). *Prosodic units as language processing units*. Unpublished doctoral dissertation, University of Massachusetts.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25, 249–271.

Tanenhaus, M., & Carlson, G. (1989). Lexical structure and language comprehension. In: W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 529–561), Cambridge, MA: MIT Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.

Venditti, J. (1995) Japanese ToBI Labeling Guidelines. Manuscript with examples, Ohio State University. For information on obtaining by ftp, send e-mail to venditti@ling.ohio-state.edu.

- Wales, R., & Toner, H. (1979). Intonation and ambiguity. In: W. C. Cooper, & E. C. T. Walker (Eds), *Sentence processing*. Hillsdale, NJ: Erlbaum.
- Warren, P. (1997). Prosody and language processing. In: S. Garrod, & M. Pickering (Eds), *Language processing* (pp. 155–188). Hove: Psychology Press.
- Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology, and parsing in closure ambiguities. *Language and Cognitive Processes*, 10, 457–486.
- Warren, P., Schafer, A. J., Speer, S. R., & White, S. D. (2000). Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations. *UCLA Working Papers in Phonetics*, 99, 5–33.
- Watson, D., & Gibson, E. (2003). *International phrasing and constituency in language production and comprehension*. Paper presented at GLOW Workshop III: Boundaries in International Phonology, Lund, Norway, April.
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 713–755.
- Watson, D., & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, Special issue, 59, 279–300.
- Watt, S., & Murray, W. (1996). Prosodic form and parsing commitments. *Journal of Psycholinguistic Research*, 25(2), 291–318.
- Wingfield, A., & Klein, J. (1971). Syntactic structure and acoustic pattern in speech perception. *Perception & Psychophysics*, 9, 23–25.
- Wong, W. Y. P., Chan, M. K. M., & Beckman, M. E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. To appear in S. -A. Jun (Ed.), *Prosodic models and transcription: Towards prosodic typology*. Oxford University Press.

This page intentionally left blank

## Chapter 14

### The Syntax–Semantics Interface: On-Line Composition of Sentence Meaning

*Liina Pylkkänen and Brian McElree*

Natural language often challenges us to derive salient differences in meanings from superficially similar expressions. For example, consider the resultative, depictive, and small clause constructions in (1)–(3).

1. The stylist combed the hair straight. (Resultative)
2. The stylist combed the hair wet. (Depictive)
3. The stylist considered the hair straight/wet. (Small clause)

Unlike prenominal adjectives, which mostly serve as noun modifiers in English (c.f., *The stylist combed/considered the straight/wet hair*), postnominal adjectives can relate to the structure within which they are embedded in a variety of ways. In (1), the resultative construction, the adjective *straight* relates causally to the action described by the verb (*viz.*, the hair became straight as a result of the stylist combing the hair). The superficially similar depictive in (2), on the other hand, involves no causal relation but rather asserts that the direct object was in the state described by the postnominal adjective during the event described by the verb. Finally, the small clause construction in (3) likewise involves no causal relation, but unlike the depictive, it does not assert that the hair was straight or wet during the act of considering, or ever as a matter of fact.

How are we able to compute these types of distinct interpretations? Part of the answer is that our language comprehension system includes compositional operations that are able to assemble meaning from our knowledge of words and how our grammar codes the types of relations implicit in (1)–(3). However, although psycholinguists have made remarkable progress over the past few decades in understanding the mechanisms involved in identifying the meaning of words and those involved with incrementally assigning grammatical structure, very little attention has been devoted to semantic composition. Consequently, we understand comparatively less about this component of the language processing system. This is somewhat surprising since compositional processes sit at the interface between lexical and syntactic processing on the one hand and higher-order discourse processing on the other. A complete theory of comprehension requires an



understanding of how comprehenders rapidly assemble an interpretation as lexical and syntactic constraints become available incrementally in a left to right analysis of the input.

In this chapter, we draw from current perspectives in semantics to outline a taxonomy of compositional operations that can serve as a framework for psycholinguistic and neurolinguistic investigations of semantic interpretation. We highlight research that finds natural expression within this framework, and, perhaps as importantly, given the paucity of relevant psycholinguistic research on composition, we point to areas that need further investigation.

Interpreting an expression typically requires integrating it into an evolving discourse model. Frequently, this requires resolving ambiguities at conceptually distinct levels, fixing reference, and drawing inferences to align local and global aspects of the discourse. To fully accomplish this task, it is uncontroversial that, in addition to lexical and syntactic constraints, comprehenders must draw upon pragmatic knowledge. The importance of high-level constraints has been illustrated by findings that comprehenders sometimes adopt a pragmatically plausible interpretation even if it is incongruent with lexical and syntactic constraints. For example, Wason and Reich (1979) and Natsopoulos (1985) documented “verbal illusions” in which comprehenders interpret sentences such as *No head injury is too trivial to be ignored* in a manner paraphrased by “No matter how trivial it might appear, a head injury should be treated.” This interpretation cannot be derived from any veridical application of local compositional operations (for an overview, see Sanford & Sturt, 2002). Similarly, Ferreira and colleagues (Ferreira, 2003; Ferreira, Ferraro, & Bailey, 2002) have shown that comprehenders often fail to accurately interpret surprisingly simple and common sentences, apparently opting for shallow forms of processing that are “good enough” for some purposes.

For some researchers, these findings challenge the existence of a compositional mechanism. For example, Ferreira et al. (2002, p. 11) state “[the] assumption of compositionality seems eminently plausible, but results in the literature on the psychology of language call it into question.” Shallow and incorrect interpretations indicate that language stimuli, like stimuli in other domains, can be processed to different depths depending on task demands and subjective criteria, but they provide virtually no grounds on which to motivate extreme approaches that would eschew altogether the notion of a compositional mechanism. For, without such a mechanism, it is difficult to imagine how comprehenders would be able to understand novel utterances or exploit the lexical properties and syntactic constraints necessary to derive interpretations in cases such as (1)–(3).

Others have used a failure to find unique brain (ERP and fMRI) responses to violations of semantic and pragmatic knowledge to mount an attack on a distinct level of semantic integration. Hagoort, Hald, Bastiaansen, and Petersson (2004) contrasted sentences that were either false for Dutch speakers given their culture-specific knowledge, e.g., *The Dutch trains are white...*, or false based on more general knowledge, e.g., *The Dutch trains are sour...* They suggested that the latter was anomalous for “semantic internal reasons” because “the core meaning of sour is related to taste and food...[and] semantic

features related to taste and food do not apply to trains” (p. 438). ERP responses to the two anomalies patterned differently in the spectral domain, but both anomalies elicited similar N400 effects. This led Hagoort et al. to argue that there is no principled distinction between semantic and world-knowledge integration.

However, it is entirely unclear whether compositional processes would draw a distinction between general and culture-specific knowledge by including, for example, feature checks on properties such as “taste” and “food.” Indeed, to do so would preclude constructing various “figurative” interpretations that routinely require comprehenders to compose interpretations with just these types of feature mismatches (c.f., *Some lawyers are sharks*, Glucksberg, 2001). Even if compositional processes respect such a distinction, comprehenders might have used pragmatic knowledge in this task to attempt to construct a plausible construal of *The Dutch trains are sour*, which would engender the same N400 component observed with *The Dutch trains are white*. Hence, we see no reason to draw strong conclusions concerning the nature of semantic processing from this type of null result.

We believe that a complete understanding of language comprehension requires a detailed understanding of the set of composition operations that comprehenders have at their disposal, and how these operations interface with the rest of the language processing system. This will require psycholinguistic and neurolinguistic research to isolate the different operations involved in semantic composition, to identify which are costly for the language processor to compute, and why some might be more costly than others. However, a necessary first step is to clearly identify candidate operations.

## 1. COMPOSITIONALITY

The principle of compositionality, usually attributed to Frege (1892), can be stated as follows:

The meaning of an expression is a function of the meanings of its parts and the way they are syntactically combined.

Given the systematicity and productivity of language, some version of compositionality is assumed by all linguistic theories of semantic interpretation. However, a major point of controversy is how strictly compositionality applies. In some theories, there is a complete homomorphism between syntax and semantics. Under this strong interpretation of compositionality (henceforth *strong compositionality*), the meanings of sentences are *fully* determined by the meanings of their constituents and by the syntactic way the constituents are combined (e.g., Fodor & Lepore, 2002; Montague, 1970). Alternative theories allow for semantic rules that do not correspond to any syntactic process (henceforth *weak compositionality*). In these theories each syntactic step still corresponds to a semantic step, but there is an additional inventory of purely semantic rules that may serve to change the meaning of a constituent to “fit” that of another (e.g., Barker, 2002; Hendriks, 1988; Jacobson, 1999; Partee & Rooth, 1983).

Whether natural language obeys strong or weak compositionality is an unresolved question about the syntax–semantics mapping. In what follows, we first sketch three rules of compositional interpretation that are taken to be basic, strongly compositional, and generally uncontroversial (Heim & Kratzer, 1998; Jackendoff, 2002). These rules specify the basic means of composing meanings from the products of lexical and syntactic analysis, and we believe they represent the types of operations that should be incorporated into psycholinguistic models of interpretation, either explicitly or functionally. We then review well-known phenomena that challenge strong compositionality and outline the main approaches to these phenomena in linguistic theory. Throughout, we discuss extant psycholinguistic results that bear on these issues, situating them in the outlined framework.

## 2. BASIC RULES OF COMPOSITIONAL INTERPRETATION

### 2.1. Semantic Types and Functional Application

The task of a theory of semantic interpretation is to characterize how elements in a syntactic string semantically relate to one another. Clearly, this depends on how we conceptualize the meanings of the elementary building blocks, the *terminal nodes* of a syntactic tree. Informally, an unchallenged view of lexical meaning is that the meanings of words have “holes” in them, which need to be filled by other words. For example, *destroy* has no valid interpretation unless it occurs in the context of a noun phrase (NP) that can be interpreted as the entity undergoing destruction, c.f. (4). Since *destroy* is not complete without a destroyee, the meaning of *destroy* can be taken to contain a variable standing for the destroyed entity.

4. The boy destroyed the sand castle.
5. \*The boy destroyed.

Whether the meanings of transitive verbs also contain variables for the subject argument is controversial, but for ease of exposition, let us assume that they do. Hence, the meaning of *destroy* would be a formula such as  $y \text{ destroy } x$ . Since *destroy* cannot by itself be used to assert that someone destroyed something, a standard way to characterize its meaning is as a rule or *function*. In other words, *destroy* is a function that requires two individuals as its input and produces as an output a description of a destruction of something by someone. This is formally expressed in the lambda calculus in (6). The input, or the *arguments* of the function, is prefixed with lambdas, and the value, or the output of the function, follows the lambda terms.

6. **destroy**:  $\lambda x.\lambda y. \text{destroy}(y, x)$

The function in (6) specifies that *destroy* requires two individuals as its arguments in order to yield a sentence that can be evaluated as either true or false. Given this, its *semantic type* is said to be a function from individuals to truth-values. Individuals and

truth-values are taken to be basic, or so-called “saturated” types; they take no arguments, and therefore they can only function as arguments, not as predicates. The type of individuals is labeled ‘*e*’ and the type of truth-values ‘*t*’. All other semantic types can be derived from these basic types, according to the rule given in (8) (Montague, 1974).

7. Basic types: *e* (individuals) and *t* (truth-values)
8. If  $\sigma$  and  $\tau$  are semantic types, then  $\langle\sigma, \tau\rangle$  is a semantic type.

This system of semantic typing assumes that the only *ontological* categories that natural language cares about are individuals and truth-values. It assumes that all meanings can be expressed in terms of these two categories. For example, the meaning of *destroy* would be described in terms of the derived type  $\langle e(e, t)\rangle$ .

Most semanticists would agree that more ontological categories are required to explain semantic phenomena. In this system, for example, all sentences have denotations that we should be able to evaluate as either true or false in the actual world. However, a hallmark of natural language is its ability to describe situations other than the actually occurring ones, e.g., *I believe it is raining*. The verb “believe” creates an “opaque” or “intensional” context, and this requires enriching the basic ontology with a notion of *possible worlds*. Although intensionality is a core topic in natural language semantics, we will not discuss it here, as it has not figured prominently in psycholinguistics (but see Clark & Haviland, 1974).

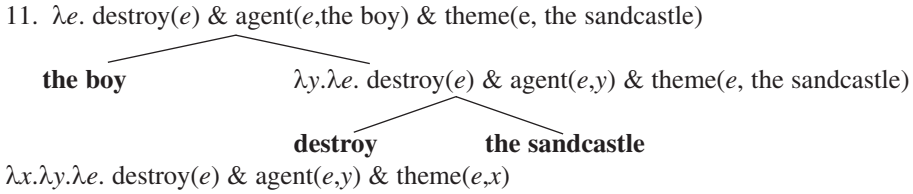
More relevant to extant psycholinguistic work on compositionality is the enrichment of the basic ontology by events (Davidson, 1967; Kratzer 1996, in press; Parsons, 1985, 1990; Rothstein, 1998; Tenny & Pustejovsky, 2000). This work captures the intuition that verbs essentially describe properties of events. Formally, this can be realized by adding an event variable to the verb’s argument structure. In (9), the revised entry for *destroy*, ‘*e*’ is a variable ranging over events. The two individuals that are realized in the syntax as the subject and object of the verb are participants in the event *e*, and they relate to the event via the thematic relations ‘agent’ and ‘theme’. In this analysis, the semantic type of *destroy* would be  $\langle e\langle s, t\rangle\rangle$ , where ‘*s*’ is the semantic type for events.

9. **destroy**:  $\lambda x.\lambda y.\lambda e.$  destroy(*e*) & agent(*e*, *y*) & theme(*e*, *x*)

The event argument has become an indispensable tool in investigations of verbal semantics and argument structure. Therefore, in the discussions to follow, we will assume that the basic ontology of natural language includes events. In this framework, the types of some common expressions would then be as in (10):

- |                                       |   |
|---------------------------------------|---|
| 10. Semantic type                     | Expression  |
| <i>e</i>                              | names ( <i>John</i> ), definite NP ( <i>the cat</i> )   |
| $\langle e, t\rangle$                 | common nouns ( <i>cat</i> ), adjectives ( <i>tall</i> ) |
| $\langle e, s, t\rangle$              | intransitive verbs ( <i>run</i> )                       |
| $\langle e\langle s, t\rangle\rangle$ | transitive verbs ( <i>kick</i> , <i>destroy</i> )       |

Equipped with this system, we can return to our example sentence *the boy destroyed the sandcastle*. In order to interpret this sentence, we first apply the function in (9) to the two NPs *the boy* and *the sandcastle*.<sup>1</sup> This saturates the two individual arguments of *destroy*.<sup>2</sup>



The tree in (11) illustrates the basic compositional rule of *Functional Application (FA)*, so called because it refers to the application of a function to its arguments. Every time we apply a function to its syntactic sister, the denotation of the sister replaces the variable in the function, and we erase the corresponding lambda term to indicate saturation of the argument. The formal definition of Functional Application is given in (12).

12. Functional Application :
- If  $\alpha$  is a branching node,  $\{\beta, \gamma\}$  is the set of  $\alpha$ 's daughters, and  $[[\beta]]$  is a *function* whose domain<sup>3</sup> contains  $[[\gamma]]$ , then  $[[\gamma]] = [[\beta]] ([[ \alpha ]])$ .  
 (Heim & Kratzer, 1998)

In sum, when the relationship between two syntactic elements is one of *selection*, such that one element denotes a function that selects another element as an argument, the mode of semantic composition is Functional Application. A large body of psycholinguistic research, briefly reviewed below, has investigated the role of argument structures in parsing and interpretation. Within the framework proposed here, this research can be viewed as investigating the real-time deployment of Functional Application. Next, we turn to situations where two elements combine in the absence of a selectional relation.

**2.2. Predicate Modification**

How do we interpret modifiers, such as the adjective *angry* in (13), which presumably does not stand in a predicate–argument relation with either *boy* or *the*?

13. The angry boy destroyed the sandcastle.

---

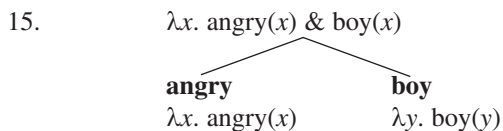
<sup>1</sup> We will ignore the interpretation of tense, aspect, and various other higher inflectional heads.  
<sup>2</sup> The event argument remains unsaturated, and so at some point it must be existentially closed either by another predicate or by a default operation. However, these details are not important for our discussion.  
<sup>3</sup> The “domain” of the function refers to the set of all possible input values to the function.

For (13) to be true, the individual who destroyed the sandcastle must have both the properties of being a boy and the properties of being angry. Thus, a rule of composition that formed the intersection of boyness and angeriness when given these two properties as its input would account for (13). Predicate modification (PM), stated in (14), achieves exactly this, and the tree diagram in (15) illustrates the application of this rule to (13).

14. Predicate Modification:

If  $\alpha$  is a branching node,  $\{\beta, \gamma\}$  is the set of  $\alpha$ 's daughters, and  $[[\beta]]$  and  $[[\gamma]]$  are both of type  $\langle e, t \rangle$ , then  $[[\alpha]] = \lambda x. [[\beta]](x) \ \& \ [[\gamma]](x)$ .

(slightly modified from Heim & Kratzer, 1998)

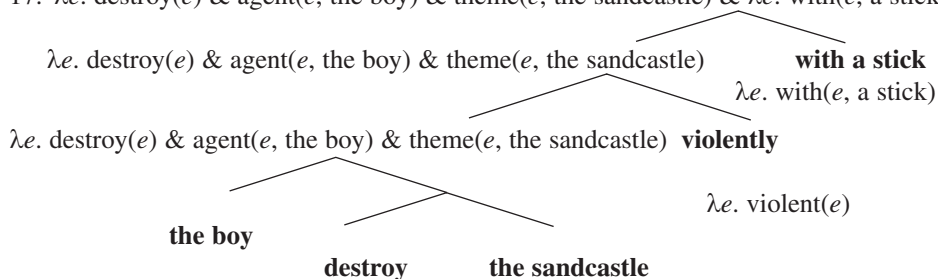


In an event semantics framework, a rule resembling Predicate Modification can capture intersective modification in the verbal domain as well. For example, all entailments in (16) can be accounted for by treating VPs and VP–adverbs as the same type of predicates of events, as shown by the step-by-step derivation in (17).

16. The boy destroyed the sandcastle violently with a stick.

- The boy destroyed the sandcastle.
- The boy destroyed the sandcastle violently.
- The boy destroyed the sandcastle with a stick.

17.  $\lambda e. \text{destroy}(e) \ \& \ \text{agent}(e, \text{the boy}) \ \& \ \text{theme}(e, \text{the sandcastle}) \ \& \ \lambda e. \text{with}(e, \text{a stick})$



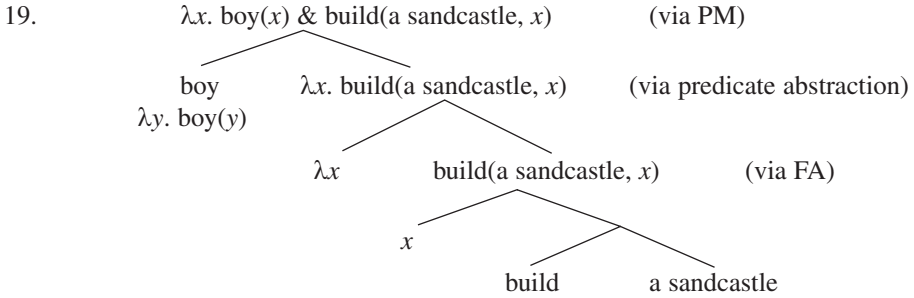
Alongside Functional Application, Predicate Modification is considered a basic composition rule that maintains full transparency of meaning. Of course, not all modification is intersective. For example, an *alleged murderer* is not necessarily a murderer. Thus, Predicate Modification is not an across-the-board interpretive rule for adjuncts. Space does not allow us to fully discuss nonintersective modification (but see Kamp & Partee, 1995; Frazier, 1999; Murphy, 2002).

### 2.3. Predicate Abstraction

Relative clauses have the surface syntax of sentences, but semantically they function as modifiers. For (18) to be true, the individual liked by Mary needs to have both properties of being a boy and of having built a sandcastle.

18. Mary likes the boy who built a sandcastle.

We could derive this meaning if *who built a sandcastle* was converted into a predicate of type  $\langle e, t \rangle$  and then combined with *boy* via Predicate Modification. For this, most theories include a rule called *Predicate (lambda) Abstraction*, which converts a formula into a predicate by binding a free variable within the formula with a lambda-operator. In our example, the relative pronoun *who* is thought to originate in the subject position of the verb, leaving behind a trace, represented as a variable. In its moved position, the relative pronoun is represented and interpreted as a lambda abstractor (for simplicity, we ignore event variables here).



Predicate abstraction is said to be involved in many types of constructions (for a review, see Partee, ter Meulen, & Wall, 1990). Recently it has also been proposed to be *the* variable binding mechanism in movement (Heim & Kratzer, 1998). Strictly speaking, predicate abstraction is not a compositional rule, but rather an operation of merging a covert variable binding operator into the syntactic tree.<sup>4</sup>

### 2.4. Real-Time Processing of Strongly Compositional Expressions

The three operations outlined above are generally viewed as strongly compositional. Although they do not exhaust the inventory of composition rules, they provide a formal account of many of the primary ways that meanings are composed. As such, they provide a detailed set of hypotheses for the investigation of the actual processing operations in semantic interpretation.

<sup>4</sup> In this sense, predicate abstraction is more similar to, say, existential closure, except that the lambda operator changes the type of its complement.

### 2.4.1. Differences among basic operations?

If we take the view of word meaning described above quite literally, and regard lexical items as functions, it would seem intuitive that Functional Application should be perhaps the most basic composition operation that comprehenders have at their disposal. This operation enables comprehenders to exploit lexical properties to compose basic predicate–argument relations. There now exists a large body of psycholinguistic evidence indicating that lexical information is used to rapidly converge on an interpretation. Carlson and Tanenhaus (1988) provided initial demonstrations that the thematic grid of a verb is used to assign thematic roles to arguments immediately after processing the verb and to resolve different types of ambiguities. Over the past decade, researchers have amassed considerable evidence supporting this claim (e.g., Boland, Tanenhaus, & Garnsey, 1990; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Carlson & Tanenhaus, 1988; Trueswell, Tanenhaus, & Kello, 1993).

Neither Predicate Modification nor Predicate Abstraction establish basic predicate–argument relations. If Functional Application is indeed the more basic operation, we might expect that it would be preferred to other operations. Current evidence is suggestive, although less than definitive.

Several studies have found that argument phrases are easier to process than adjunct phrases in otherwise neutral contexts (e.g., Boland, 2005; Boland & Boehm-Jernigan, 1998; Clifton, Speer, & Abney, 1991; Kennison, 1999, 2002; Liversedge, Pickering, Clayes, & Branigan, 2003; Liversedge, Pickering, Branigan, & van Gompel, 1998; Schütze & Gibson, 1999; Speer & Clifton, 1998). For example, Clifton et al. (1991) found that arguments (e.g., *The saleswoman tried to interest the man in a wallet...* or *The man expressed his interest in a wallet...*) were read faster than adjuncts that modified either a preceding verb (e.g., *The man expressed his interest in a hurry...*) or a preceding noun (e.g., *The saleswoman tried to interest the man in his fifties...*). Schütze and Gibson (1999) reported that participants read prepositional phrases (PP) in verb–NP–PP sequences faster when they were arguments of an NP (e.g., *The company lawyers considered employee demands for a raise...*) as opposed to adjuncts of the verb (e.g., *The company lawyers considered employee demands for a month...*). Finally, Kennison (1999) reported that participants read agentive by-phrases in complex event nominals faster when they were arguments (e.g., *The frequent collection of butterflies by the kids...*) rather than adjuncts (e.g., *The numerous collections of butterflies by the kids...*).

To the degree that adjuncts involve Predicate Modification, the privileged processing status for arguments suggests a preference for Functional Application. A notable complication is that most published studies have investigated sentences with temporary structural ambiguities. For example, in both the Clifton et al. (1991) and the Schütze and Gibson (1999) studies, the PPs were ambiguous between an argument and adjunct analysis until the noun was encountered. Hence, the slower processing times for adjuncts may only reflect parsing preferences.

It is difficult to cleanly disentangle syntactic and semantic effects in strongly compositional constructions, because of the tight linkage between syntactic and semantic composition. However, Schütze and Gibson (1999) argue that simple structural parsing



strategies (e.g., *Minimal Attachment* or *Late Closure*, Frazier & Rayner, 1982) cannot explain their findings. They suggest that a better account is one that posits “an argument preference strategy” (e.g., Abney 1989). This strategy essentially grants priority to Functional Application in proposing that comprehenders strive to saturate the arguments of a function rather than pursuing analyses that involve operations such as Predicate Modification. Other findings also partially implicate interpretative factors in the argument advantage. Speer and Clifton (1998) found that the plausibility of a constituent as an argument or as an adjunct dramatically modulated overall reading times. Plausibility did not fully eliminate the argument advantage, but it did interact with argumenthood such that the argument advantage was less with high plausibility. This partly implicates interpretative factors, as plausibility reflects the semantic fit of an item as an argument or as a modifier.

An alternative means of investigating whether Functional Application has a privileged status is to determine whether it is computationally less costly than other operations. Indeed, Functional Application might be preferred because it is less costly. Unfortunately, we do not currently have good direct measures of the relative cost of different basic compositional operations. Studies demonstrating that arguments are easier to process than adjuncts in structurally ambiguous contexts do not necessarily bear on the issue, since the processing cost might reflect an initial bias for arguments. Hence, a challenge for future research is to construct and examine relatively unambiguous structures that vary the mode of composition holding other factors constant (e.g., frequency of occurrence).

#### 2.4.2. *Intrinsic processing issues*

In addition to investigating potential differences between basic composition operations, research is needed to systematically examine how different factors affect each operation. The expectation is that such research will provide the foundation for developing explicit processing models, as analogous research has done in the investigation of other components of language comprehension (e.g., word recognition). Below, we note a few salient issues.

The intrinsic complexity of the compositional operation may be an important determinant in performance. Concerning Functional Application, for example, two recent papers have argued that the complexity of a verb’s event structure affects processing speed. McKoon and Macfarland (2000, 2002) reported longer recognition times for verbs with externally caused events (e.g., *break*) than for verbs with internally caused event (e.g., *grow*), which under their analysis (following Levin & Rappaport Hovav, 1995) have a simpler noncausative event structure. Gennari and Poeppel (2003) report that lexical decision and self-paced reading times were longer for eventive verbs (e.g., *destroy*, see (9)), which describe causally structured events, than for stative verbs (e.g., *love*), which involve a simpler event structure.

Natural language can contain ambiguities at any number of conceptually distinct levels of representation, and the interpretive level is no exception. A major issue for any processing model is how comprehenders resolve ambiguities in the application of an operation. For example, there can be uncertainties about whether an adverb relates to the

event, as in the resultative construction in (1), or to one of individuals in the event, as in object depictive construction in (2) (for preliminary findings, see Frazier & Clifton, 1996; Pykkänen & McElree, 2004). Likewise, in noun modification, there are often uncertainties about whether the modification is intersective or nonintersective (e.g., *a beautiful dancer*).

Several factors are likely to conspire to promote one analysis over another. Some analyses might be preferred because they are inherently more amenable to incremental analysis. For example, in noun modification, there are reasons to suppose that intersective interpretations may be less costly and more compatible with rapid incremental interpretation than nonintersective interpretations, as, in principle, the interpretation of an intersective adjective need not be relativized to the head noun (Frazier, 1999). Currently, however, there is no clear evidence for a general preference for intersective modification (Frazier, 1999).

Lexical and pragmatic properties are clearly an important determinant. For example, in depictive structures, the semantic and pragmatic properties associated with an adjective can induce an interpretation where the adjective is taken to modify the subject (e.g., *The man ate the meat angry*) or, conversely, the object (e.g., *The man ate the meat raw*). In NP modification, the salience of the modified dimension appears to be important. Murphy (1990) found that nouns modified by adjectives were understood faster than nouns modified by other nouns or by adjectives derived from nominals (e.g., *corporate check*). He argues that these differences reflect the fact that adjectives (e.g., *blue sky*) often denote the dimension they modify, whereas one noun can often modify another on several dimensions (see Murphy, 2002, Chapter. 12, for an excellent overview).

Given the rather ubiquitous effects of frequency on other components of the language processing system, it would be surprising if compositional operations were not sensitive to differences in frequency at various grains of analysis. For example, Pykkänen and McElree (2004) found that resultatives were read faster and more accurately than both subject depictives (e.g., *The teenage boy painted the wall turquoise...* versus *The teenage boy painted the wall bored...*) and object depictives (e.g., *The artist knocked the frame crooked...* versus *The artist returned the frame crooked...*), even though the adjectives were matched on relevant properties such as frequency, length, and similarity to the modified noun. These differences could reflect architectural differences, but a corpus analysis of 1462 V–NP–AP sequences in the Penn Treebank corpus (Marcus, Santorini, & Marcinkiewicz, 1993) revealed that resultatives are nearly an order of magnitude more frequent than depictives (370 versus 39 instances). Thus, it is quite likely that the frequency of an operation facilitated performance. Additionally, it is reasonable to expect that finer grain co-occurrence frequencies will also affect performance. For example, in a corpus analysis, Spivey-Knowlton and Sedivy (1995) found that *with*-PPs are more likely to modify a verb than a direct object with action verbs (e.g., *smashed*), but that the opposite pattern held with psychological state and perception verbs (e.g., *hoped* and *heard*). Self-paced reading times reflected these differences: *with*-PPs were read faster when the semantic properties of the PP were consistent with the more frequent modifiee.

Finally, discourse constraints are likely to induce different compositional operations. In a visual world paradigm, in which eye movements were monitored, while comprehenders interact with visual displays, Sedivy, Tanenhaus, Chambers, and Carlson (1999) reported striking evidence that comprehenders rapidly interpret scalar adjectives (e.g., *the tall glass*) contrastively when the referential domain supported this interpretation (see Phillips, 2001, for discussion of the implications of these findings for theories of incremental compositional interpretation). Other findings also indicate that basic interpretive operations interact with general contextual and discourse factors. For example, discourse information can eliminate the processing disadvantage for adjuncts. Livesiedge et al. (1998) found that, in isolation, locative by-phrases following passive verb forms (e.g., *The shrubs were planted by the greenhouse*) were more difficult to process than agentive by-phrases (e.g., *The shrubs were planted by the apprentice*), replicating the basic argument advantage. However, the adjunct disadvantage disappeared when context set up an expectation for a location (e.g., *The man was wondering where to plant the shrubs*). Collectively, these results point to the need to systematically investigate how different types of constraints can modulate processing costs and the likelihood of pursuing various compositional operations.

### 3. CHALLENGES FOR COMPOSITIONALITY

Although the interpretive rules sketched in Section 2 have significant empirical coverage, many semantic phenomena seem to call for more powerful interpretive mechanisms. We provide an overview of constructions that challenge the simple view described so far. First, we discuss cases of *type mismatch*, where elements can combine even though they do not constitute an appropriate structural environment for either Functional Application or Predicate Modification. Second, we review cases where a part of the meaning of a sentence is not overtly expressed by any of its syntactic constituents, and the unexpressed meaning cannot be generated by any basic compositional operation.

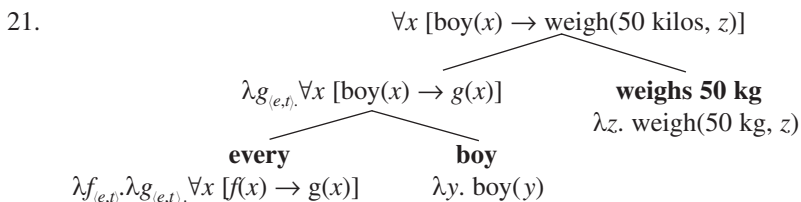
#### 3.1. Type Mismatch

##### 3.1.1. Quantifier interpretation

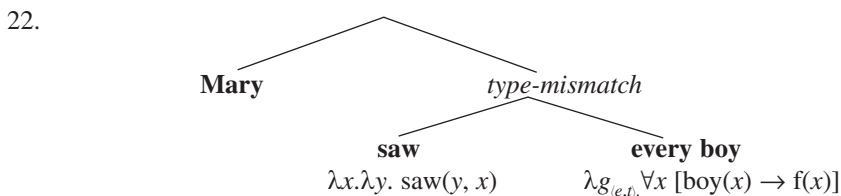
The system described so far cannot account straightforwardly for even the simplest cases of quantification. We have described a view where verbs are functions that can take individuals, such as *the boy*, as their input. Since quantified NPs such as *every boy* combine with verbs without problem (*every boy ran*), the approach suggests that quantifier phrases denote individuals. However, such an approach gets the meaning of quantifier phrases wrong in many ways. Consider an example from Sauerland and van Stechow (2001): If a quantified phrase such as *every boy* denoted a group consisting of every boy, then it should be possible to predicate properties such as *weighs 50 kg* of that group. However, (20) clearly does not mean that the group of boys collectively weighs 50 kg, rather that every (contextually relevant) individual who is a boy weighs 50 kg.

20. Every boy weighs 50 kg.

To derive this interpretation, the notion that *weighs* takes *every boy* as its argument is relinquished. Instead, *every* is treated as a function that takes two predicates of individuals as its arguments: the first one (e.g., *boy*) restricts the domain of quantification, and the second one attributes a property (e.g., the property of weighing 50 kg) to those individuals satisfying the restrictor predicate. This is shown in the tree diagram in (21).



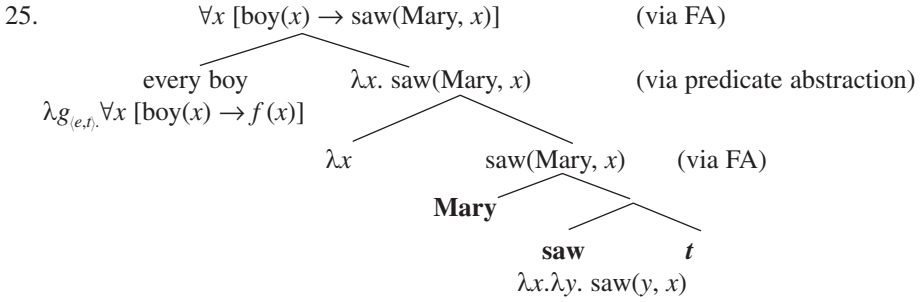
This standard treatment of subject quantifiers essentially dates back to Frege (1879). By treating quantifier phrases as higher-order predicates that take their verbs as their argument, instead of vice versa, we have an account of quantified subjects. However, this account cannot straightforwardly explain how we compose an interpretation of quantified *objects*. (22) illustrates the problem. If the object is quantified, as in (22), then the second argument of the quantifier needs to be the transitive verb. But the lexical entry for *every* given in (21) requires the second argument of the quantifier to be of type  $(e,t)$ , i.e., an intransitive verb. A type mismatch results.



Two main approaches to this type of type mismatch have been proposed. One involves simply shifting the object to a higher type. In this account, there is a freely available type shifting rule that serves to lift the type of every (and other quantifiers) from (23) to (24).

- 23. **every**<sub>1</sub>:  $\lambda f_{(e,t)}. \lambda g_{(e,t)}. \forall x [f(x) \rightarrow g(x)]$  (basic type)
- 24. **every**<sub>2</sub>:  $\lambda f_{(e,t)}. \lambda g_{(e,et)}. \lambda x. \forall y [f(y) \rightarrow g(y, x)]$  (lifted type)

The other approach involves syntactic movement, or quantifier raising (QR) (Heim & Kratzer, 1998). The object is raised out of the VP, and in doing so it creates an abstract predicate above the VP. In other words, the movement *creates* a predicate that is of a suitable type for the quantifier.

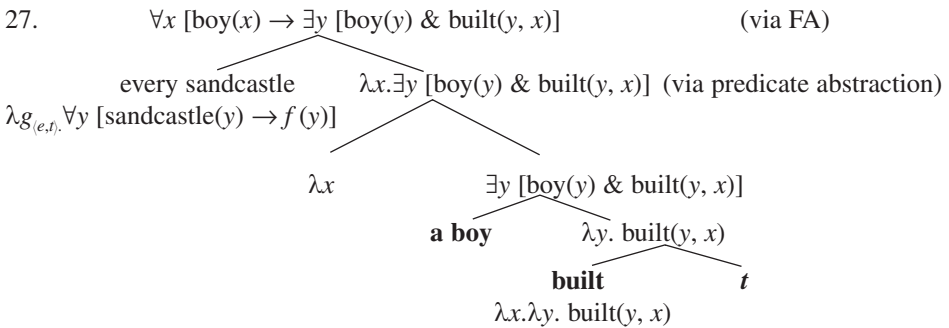


For our purpose, the essential point is that all extant theories involve some additional operations to account for the interpretation of object quantifiers.

Matters are complicated further if two or more quantifiers are present, as in (26), which includes an indefinite NP. Cases such as this are inherently ambiguous, as the quantifiers can take different scope with respect to each other.

26. A boy built every sandcastle.
- i. Surface scope: There exists an  $x$  such that  $x$  is a boy and  $x$  built a sandcastle.
  - ii. Inverse scope: For all  $x$ , such that  $x$  is a sandcastle, there exists an individual  $y$ , such that  $y$  is a boy and  $y$  built  $x$ .

How is inverse scope computed? Under the assumption that object quantifiers can raise, inverse scope is easily obtained.



Under a QR approach, we must assume that the subject quantifier phrase can also move, in order to account for surface scope. Thus, surface scope would be computed by further raising the subject quantifier over the object.

3.1.2. *Real-time processing of quantified expressions*

Only a few studies have investigated the interpretation of quantified expressions. So far, researchers have not addressed the basic question of whether quantifiers in object position are more costly to process than quantifiers in subject position, as most theories

would predict (but see Varvoutis & Hackl, 2006). Instead, studies have focused on doubly quantified structures to determine whether one scope assignment is preferred or less costly to compute (e.g., Anderson, 2004; Filik, Paterson, & Liversedge, 2004; Gillen, 1991; Kurtzman & MacDonald, 1993; Tunstall, 1998). The guiding intuition is that the latter should be preferred or consume less processing resources.

However, because doubly quantified sentences are ambiguous, researchers can only infer what interpretations comprehenders consider while processing the quantifiers. A typical research strategy has been to co-opt an experimental protocol used to investigate syntactic ambiguities. This is illustrated by (28) and (29), taken from a study by Kurtzman and MacDonald (1993).

28. A kid climbed every tree. The kid was full of energy.  
 29. A kid climbed every tree. The kids were full of energy.

The singular continuation in (28), *the kid*, forces a surface scope interpretation of the first sentence, whereas the plural continuation in (29) forces an inverse scope interpretation. Differential difficulty at one of these continuations provides evidence for the relative dominance of one of the scope readings. Conversely, no difficulty at this region suggests one of three options: either comprehenders are capable of computing both readings in parallel (e.g., Kurtzman & MacDonald, 1993), comprehenders compute an underspecified interpretation compatible with both readings (e.g., Gillen, 1991; see also Sanford & Sturt, 2002), or comprehenders stochastically compute either readings with approximately equal frequency.

Some studies have reported faster reading or judgment times for a singular continuation such as (28) than for a plural continuation such as (29), consistent with a preference for the computationally simpler surface scope reading (Anderson, 2004; Tunstall, 1998). Unfortunately, other results suggest that this difference may not reflect an overarching surface scope preference. Tunstall (1998) found an advantage in response times for grammaticality judgments of singular continuations in sentences analogous to (28), but comparable latencies for singular and plural continuations when the indefinite quantifier (*a*) followed the universal quantifier (*every*) in surface order. In a self-paced (sentence-by-sentence) reading time study, Gillen (1991) found that, irrespective of surface order, singular continuations were read more quickly than plural continuations. Filik et al. (2004) also found a general advantage for singular continuations in an eye-tracking study that factorially varied the type of quantifier, surface order of the quantifier, and ordering of grammatical arguments (direct and indirect object). These results undermine the claim that the advantage for singular continuations in structures like (28) reflect a strong preference for surface scope.

Interestingly, Anderson (2004) conditionalized reading times for doubly quantified sentences on responses to post-sentence questions, which were designed to indicate whether readers interpreted the sentences with surface or inverse scope. Sentences conditionalized on inverse responses were read more slowly than those conditionalized on

surface responses. Hence, although there might not be a general preference for surface scope, this finding suggests a cost for computing a scope configuration other than what is represented in surface syntax.

However, Filik et al. (2004) found across-the-board differences in quantifier orders, with indefinite-universal orders being more costly to process than universal-indefinite orders. This finding is consistent with a prediction of Fodor (1982; see also Ioup, 1975) that comprehenders will initially assume that an indefinite NP refers to one entity. In an indefinite-universal ordering, this initial bias will limit comprehenders' ability to compute an inverse interpretation when the universal quantifier is encountered, as they would need to revise their interpretation of the indefinite to denote multiple entities. Note that no such bias would be found in a sentence with a universal-indefinite ordering. The implication of the Filik et al. (2004) finding is that the observed difficulty of computing an inverse scope reading in indefinite-universal orders can only be attributed to the cost of computing inverse scope relations *per se* if there are also measurable effects in constructions with universal-indefinite orders. Otherwise, we cannot rule out the possibility that such effects represent reanalysis of initial semantic commitments.

Unfortunately, Anderson (2004) did not examine universal-indefinite orders in her conditional analysis. Nonetheless, she did find that, with highly constrained contexts, sentences with both relatively unambiguous inverse scope readings (e.g., *A different member tested every recipe*) and ambiguous inverse scope readings (e.g., *A member tested every recipe*) were read slower than sentences with both relatively unambiguous surface scope readings (e.g., *Every member tested a different recipe*) and ambiguous surface scope readings (e.g., *Every member tested a recipe*). To the degree that the indefinite NP, *a different member*, would be unlikely to be interpreted as one entity in a context that introduced several members (e.g., *...Members who nominated recipes were required to test the recipes to make sure that the instructions were correct. A different member...*), these results suggest that computing inverse scope is more costly than surface scope.

A host of factors may affect the likelihood that comprehenders commit to an initial interpretation of an NP when the sentence is ambiguous. This makes it rather difficult to construct clean tests of whether inverse scope is costly to compute across different constructions. Anderson's (2004) results with relatively unambiguous structures are suggestive. An alternative approach would be to investigate unambiguous structures with a singly quantified expression in object position. Such cases would provide a simpler test of whether the hypothesized type-shifting operation is taxing to perform.

### 3.1.3. Type mismatch in conjunctions

As so far described, NPs can be of two different types: individuals (type *e*) or higher order functions taking a predicate as their argument (type  $\langle\langle et \rangle, t \rangle$ ). The higher order functions, such as quantifiers, take verbs as their arguments, whereas the *e*-type NPs serve as the arguments of verbs. Under this hypothesis, one might expect that *e*-type and higher order NPs should not be able to conjoin, as this would create a conflict for the

directionality of Functional Application. But this prediction is clearly not born out: Proper names and definite NPs can conjoin with quantified NPs without problem as in (30) and (31).

30. The teacher and every student left the classroom.

31. Sally and some boy skipped class.

How is type mismatch in conjunctions resolved? Two main approaches have been proposed. One is a so-called “generalizing to the worst case” solution where all NPs are treated as higher order predicates, or as so-called *generalized quantifiers* (Montague, 1970). The consequence of this is that noun phrases *always* take verbs as their arguments rather than vice versa. The other approach is to shift the types of *e*-type NPs to a higher type only when necessary (Partee & Rooth, 1983). Thus, the type-lifting rule shown in (32) would serve to shift the *e*-type argument to a higher type in both (30) and (31).

32. lift:  $e \rightarrow \langle\langle e, t \rangle, t \rangle$   
           Sally  $\rightarrow \lambda\lambda f_{\langle e, t \rangle}. f(\text{Sally})$

Partee and Rooth (1983) propose type shifting as a last resort processing strategy to resolve type mismatches. Type-shifting is a purely semantic operation with no syntactic reflex and therefore sacrifices strong compositionality. Although most semantic theories incorporate type-shifting rules in some form or other, their existence is controversial (Heim & Kratzer, 1998). The lack of consensus about type-shifting may be at least partly due to the fact that it is difficult to find linguistic evidence that sharply distinguishes between different approaches. Given the controversy, type shifting would be a natural domain for psycholinguistic investigation. However, there have been no experimental studies investigating whether these traditional cases of type-shifting such as (30) and (31) are costly in real-time processing. Instead, an increasing body of experimental work has researched a closely related phenomenon, called “coercion.” In coercion, the semantic types of syntactic sisters mismatch, and in addition, some part of the meaning of the sentence is syntactically unexpressed.

### 3.2. Coercion

Psycholinguistic research on compositionality has mainly investigated the processing of two types of expressions: *complement coercion* and *aspectual coercion*. Both complement coercion and aspectual coercion involve a semantic mismatch and have been argued to challenge strong compositionality. Work on these constructions has focused either on their representation or on their processing with little synthesis. In what follows, we consider both traditional linguistic evidence and psycholinguistic and neurolinguistic data in evaluating to what extent these cases indeed require nonsyntactic interpretive mechanisms.

#### 3.2.1. Complement coercion

So far we have only considered verbs that take individuals, such as *John* or *the book*, as their arguments. However, many verbs select, not for individuals, but for verb phrases (VPs). Aspectual verbs such as *begin* and *finish* are good examples of VP-selecting



verbs. They denote functions that, when given a VP describing an event as their input (such as *writing a book*), return a predicate that denotes the initial or final part of that event (*began/finished writing a book*). Given these selectional restrictions, strong compositionality would seem to predict that these verbs should be ungrammatical if combined with non-event-denoting objects. But interestingly, these verbs are entirely natural and grammatical with objects that denote individuals rather than events, as shown in (33) and (34).

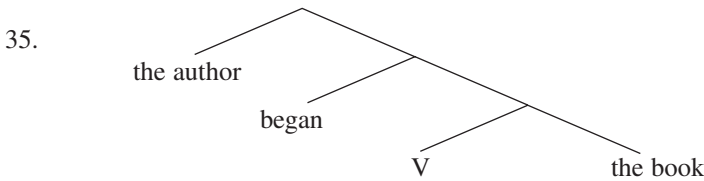
- 33. The author began the book.
- 34. The student finished the thesis.

Indeed, corpus analyses demonstrate that expressions such as (33) and (34) are greater than 9 to 1 more frequent than forms that fully specify the event (e.g., *The author began writing a book*) when, as here, the event is commonly associated with the noun (e.g., reading or writing a book). Fully specified events frequently occur with less predictable activities, such as *translate the book* (Briscoe, Copestake, & Boguraev, 1990; Lapata & Lascarides, 2003; Lapata, Keller, & Scheepers, 2003).

Processing of these sentences has been hypothesized to involve a process of *complement coercion*, a nonsyntactic process that converts the entity-denoting object into an event description that satisfies the selectional demands of the verb (Jackendoff, 1997; Pustejovsky, 1995). In (33) and (34), the end result of the coercion process would be an interpretation in which, for example, the author began writing the book and the student finished writing the thesis. Complement coercion has been taken to be an obvious disconfirmation of strong compositionality, but possible syntactic mechanisms for repairing the semantic mismatch have not been seriously considered. In what follows, we sketch a syntactic solution to coercion and evaluate its feasibility.

3.2.2. *Coercion as VP insertion*

A possible solution would be to assume that a syntactically silent VP has been inserted into the structure, as shown in (35), and that this circumvents the need for an extra-syntactic interpretive mechanism.<sup>5</sup>



<sup>5</sup> Much of natural language has been hypothesized to involve syntactic heads that do not have an overt pronunciation and thus there is nothing exotic about (35). For example, although English present tense is morphologically unmarked in all persons except the third singular, no one would claim that the syntactic tree of a sentence such as *I love Lucy* does not involve a tense node.

3.2.2.1. *Evidence in favor of VP insertion.* If coercion is based on VP insertion, it should be blocked exactly in those environments where overt VP complements are ungrammatical. Many event-selecting verbs participate in the causative-unaccusative alternation, as shown in (36) and (37). (37) shows that an event-denoting nominal object can raise to subject position if an agent does not already fill that position. However, this is possible only when the direct object is an NP, not when it is a VP, as illustrated in the gerundive and infinitival VP complements in (39) and (41).<sup>6</sup>

Event-denoting NP-complement:

36. We began the war. (causative)  
 37. The war began. (unaccusative)

Event-denoting VP-complement (gerund):

38. We began reading the book.  
 39. \*Reading the book began.

Event-denoting VP-complement (infinitive):

40. We began to read the book.  
 41. \*To read the book began.

If coercion is purely semantic, there should be no particular reason for it to be blocked with unaccusatives. However, the data patterns clearly favor a VP-hypothesis. While (42) can be used to assert that Sarah began a book-reading event, (43) cannot be used to assert that a reading of a book began. Similarly, (44) can be used to convey that Bill continued writing a thesis, but (45) cannot be used to express that a thesis-writing event continued.<sup>7</sup>

42. Sarah began the book.  
 43. \*The book began.  
 44. Bill continued the thesis.  
 45. \*The thesis continued.

These data speak against an *a priori* exclusion of a syntactic mechanism for coercion. A further distributional test can be devised by embedding event-selecting verbs inside adjectives. In (46), the event-selecting verb *finish* is combined with an entity

<sup>6</sup> The ungrammaticality of (39) and (41) is not due to a general constraint against gerundive and infinitival subjects (c.f., *Reading this book will be enjoyable* or *To read this book would be a mistake*), nor to a general prohibition against event-denoting subjects with unaccusatives (c.f., *The reading of the verdict began*).

<sup>7</sup> (43) and (45) are only grammatical under the irrelevant reading where *the book* and *the thesis* are interpreted as texts: *the book began with a prologue*.

entity-denoting argument. Here, composition is mediated by the adjectival suffix *able* (and a semantically vacuous copula). The natural interpretation of this sentence is that some activity involving the book is finishable. Thus it appears that, in contrast to unaccusatives, coercion is possible in the subject position of deverbal adjectives (for supporting processing evidence, see Section 3.2.3.1).

46. This book is finishable.

*It is possible to finish [Ving] this book.*

Now, if coercion occurs in the subject position of (46), and if coercion is VP-insertion, overt VP-arguments should also be possible in the subject position of able-adjectives. In other words, sentences such as those (47) and (48) should be well formed.

47. Reading this book is finishable.

48. Climbing this wall may be survivable.

Most native speakers seem to judge these sentences as marginal. Importantly, however, they appear better than the unaccusative examples in (43) and (45). If sentences such as (47) and (48) are indeed acceptable, there would seem to be a correlation between the distribution of coercion and the distribution of possible VP arguments: When VP arguments are ungrammatical, coercion is blocked, and when VP arguments are possible, so is coercion. The hypothesis that coercion is VP insertion derives this pattern for free.

3.2.2.2. *Evidence against VP-insertion.* Although the VP-hypothesis fares well in accounting for some distributional properties of coercion, it runs into problems with modification. If the structure of coerced expressions is as in (35), there should be two attachment sites for VP modifiers: the higher VP and the lower VP. The examples in (49) and (50) illustrate the availability of the lower VP for VP modification in those cases where the VP is overt. The lower VP is modified by the manner adverb *slowly* in (49), and by an instrumental phrase in (50).

Overt VP: Adverbial can modify the lower VP

49. We finished eating the meal slowly.

*True if a slow meal comes to an end quickly.*

50. I started cutting a loaf of bread with a knife.

*True if an event of cutting the loaf of bread with a knife was initiated.*

If the structure in (35) is right for coercion, these VP-modifiers should be able to take lower scope in coerced expressions as well. However, this seems impossible, as shown in (51) and (52).

Coercion: Adverbial cannot modify the “lower VP”

51. We finished the meal slowly.

*False if a slow meal comes to an end quickly.*

52. #I started a loaf of bread with a knife.

*False if an event of cutting the loaf of bread with a knife was initiated.*

*The knife must be an instrument of the initiation.*

The VP-hypothesis also makes wrong predictions about passivization. With overt VP complements, raising of the object of the lower VP to the subject position of the matrix clause results in deviance, as shown by the (a) examples below. If the coerced versions of these sentences are syntactically identical to the overt versions, they should also be ill formed in the passive. But, contra this prediction, the coerced versions are entirely natural.

53. a. ?The book was begun to be written by the author.

b. The book was begun by the author.

54. a. ?The album was completed to be recorded by the artist.

b. The album was completed by the artist.

55. a. ?The problem was attempted to be solved by the minister.

b. The problem was attempted by the minister.

In sum, if coerced expressions involve a silent VP, that VP for some reason cannot be adverbially modified and it does not constitute an intervener for A-movement in passivization. In other words, with regard to these phenomena, the VP would have to behave as if it was not there. Thus, despite the positive results regarding the distribution of coercion, the VP hypothesis is problematic in other domains. Consequently, a type-shifting analysis of coercion may be necessary, even if at the cost of sacrificing strong compositionality. Collectively the evidence suggests complement coercion constitutes a strong candidate for a purely semantic interpretive process.<sup>8</sup>

If coercion requires an extra-syntactic process that is not part of the default repertoire of interpretive rules, coerced sentences should be more costly to process than noncoerced controls. We next turn to processing measures of whether coerced sentences involve on-line construction of an event predicate from an entity-denoting NP. Further, if the resolution of the type mismatch in coercion involves *composition* of an event structure rather than just retrieval of a suitable activity, it should be possible to obtain a processing delay that is not simply due to activity retrieval.

---

<sup>8</sup> A rigorous formal examination of the representation of coercion is beyond the scope of this chapter; for proposals, see Egg (2003), Jackendoff (1997), and Pustejovsky (1995).

### 3.2.3. Real-time processing of complement coercion

3.2.3.1. *Basic findings.* Reading time investigations of complement coercion have repeatedly found that they are indeed more difficult to process than various types of control expressions, consistent with the idea that they require costly operations to repair the type mismatch between an event-selecting verb and an entity-denoting object. (McElree et al., 2001; McElree, Frisson, & Pickering, 2006; Pickering, McElree, & Traxler, 2005; Traxler, Pickering, & McElree, 2002; Traxler, McElree, Williams, & Pickering, in press). Initial demonstrations compared sentences such as (56) and (57).

56. The carpenter began the table during the morning break. (coerced)

57. The carpenter built the table during the morning break. (control)

Coerced expressions such as *began the table* were compared to control expressions, which used the verb that readers most often ascribed to the eventive interpretation of the coerced expression (determined by completion norms, see McElree et al., 2001). Reading time measures indicated that processing (56) was more costly than (57), even though the sentences were rated as equally plausible: in self-paced reading, participants took longer reading *table* and *during* in (56) than in (57) (McElree et al., 2001); in eye tracking, reliable differences first emerged at the NP (e.g., *the table*; Pickering et al., 2005) or on the two words following the NP (e.g., *during the*; Traxler et al., 2002).

These studies provided the initial evidence that expressions requiring complement coercion are indeed taxing to process. However, conclusions based on comparisons between two different types of constructions can be confounded by uncontrolled factors. In what follows, we outline findings indicating that the slower reading of coerced expressions is *not* due to uncontrolled semantic properties, differences in cloze probabilities, or differences in co-occurrence frequencies, and that the effect clearly reflects the slower computation of a coerced interpretation.

Contrasts such as (56) and (57) are not fully synonymous. Minimally, they differ in aspectual properties. However, Pickering et al. (2005) found the same coercion effect when (56) was compared to (58), which overtly expresses the interpretation that comprehenders report giving to the coerced form.

58. The carpenter began building the table during the morning break.

Eye-tracking measures showed that control expressions such as (57) and (58) did not differ at *the table* or beyond, but readers spent longer processing *the table* in (56) than both (57) and (58).

Another concern is that eventive verbs such as *begin* could be generally more semantically complex than control verbs such as *build* (c.f., Gennari & Poeppel, 2003), and

hence they might be more taxing to process with any type of complement. A related concern is that cloze probabilities are likely to be higher for an expression such as *built the table* than for *began the table*. Indeed, in the Traxler et al. (2002) materials, the average cloze values were 0.03 for coerced expressions such as (56) and ranged from 0.14 to 0.19 for controls such as (57). These values are all quite low, well below the range that typically affects eye-tracking measures (Rayner & Well, 1996), but they do trend in the direction of the processing cost.

Traxler et al. (2002) addressed both concerns by using both self-paced and eye-tracking measures to examine the processing of eventive verbs (e.g., *started*) paired with entity-denoting NP complements (e.g., *the puzzle*) and event-denoting NP complements (e.g., *the fight*). *The boy started the puzzle* was more difficult to process than *The boy started the fight*, despite the fact that cloze values were nearly identical (0.03 and 0.02 for event and entity nouns). Moreover, entity- and event-denoting NPs did not differ as the complements of verbs such as *saw* (e.g., *The boy saw the puzzle/fight*), which can semantically combine with either argument type. The form of this interaction indicates that the observed processing cost is localized to the combination of an eventive verb and entity-denoting complement, and this strongly suggests the effect is linked to the kind of compositional operations comprehenders use to interpret the VP.

Cloze measures may not reflect subtler differences in the frequency with which various constituents co-occur in the language user's experience. Corpus analysis is not useful for assessing the frequencies of particular verb–complement pairings, as the data for many pairings of constituents will be sparse. One solution is to use similarity measures derived from latent semantic analysis (LSA) (Landauer, Foltz, & Laham, 1998), which gives an approximation to co-occurrence – the cosine between pairs of text – reflecting the degree to which the constituents appear in similar contexts even if they never appeared together. Using the Traxler et al. materials, we estimated the co-occurrence of event- and entity-denoting NPs (e.g., *fight* versus *puzzle*) following the subject and either an eventive verb (e.g., *The boy started...*) or a verb that selects either an event or entity complement (e.g., *The boy saw...*). From a 300-factor LSA database reflecting usage up to 1st year college, the mean cosine for the entity-denoting NPs was estimated at 0.285 (sd=0.145) following eventive verbs and 0.280 (sd=0.167) following neutral verbs. Overall, event-denoting NPs had *lower* cosine values than entity-denoting NPs, 0.178 (sd=0.131) following eventive verbs and 0.169 (sd=0.161) following neutral verbs. This analysis suggests that the observed differences cannot be attributed to co-occurrence patterns, as the results are exactly opposite this type of account: the condition that showed the greatest cost – the one which required complement coercion (e.g., *The boy started the puzzle*) – had the highest cosine value (0.285).

Coercion effects also appear to generalize to other types of NPs and to languages other than English. Traxler et al. (2002) found a coercion cost for complements with both indefinite and definite NPs (*a book* and *the book*). This suggests that the effect cannot be attributed to the particular types of pragmatic accommodations that might be required by a definite NP. McElree, Frisson, and Pickering (2006) found reliable coercion effects

with proper nouns, ...*began Dickens* versus ...*met/read Dickens*. Eye-movement measures during reading indicated that metonymies (...*read Dickens*) were not more costly to interpret than conventional expressions (...*met Dickens*), but expressions that required coercing *Dickens* into an event were more taxing to interpret than both. Scheepers, Mohr, Keller, and Lapata (2004) replicated the contrast between *began/read the book* in German, with materials tightly matched on overall plausibility ratings and the predictability of the object noun.

Finally, a recent speed-accuracy tradeoff (SAT) study indicates that reading time measures reflect differences in the time needed for comprehenders to build an event interpretation of the complement (McElree, Pykkänen, Pickering, & Traxler, 2006). A notable short-coming of reading time measures is that longer reading times might indicate that readers take longer to interpret one expression than another *or* that readers are simply less likely to accurately process all the information necessary for an interpretation (McElree, 1993; McElree & Nordlie, 1999). The SAT procedure provides a means of directly measuring processing time in the presence of concomitant differences in accuracy (Bornkessel, McElree, Schlesewsky, & Friederici, 2004; McElree, 1993; McElree & Nordlie, 1999).

McElree et al. (2006) used the SAT procedure to investigate the processing of materials such as (56) and (57) above, as well as constructions such as (59) and (60).

59. The climber proved the ice survivable. (coerced)

60. The climber proved the fall survivable. (control)

These additional constructions were formed around verbs that select for small clause complements, consisting of an NP subject and an adjectival predicate. Crucially, the adjectives (e.g., *survivable*) were morphologically derived from eventive verbs. As such, they semantically required eventive subject NPs, just as eventive verbs require eventive complements (see Section 3.2.2.1). Composition should be simpler when a subject NP denotes an event (e.g., *the fall*) than when it does not (e.g., *the ice*), as the adjective in the latter case should trigger coercion of the NP into an eventive interpretation (e.g., “climbing of the ice”).

In a trial of the SAT procedure, a sentence was presented phrase by phrase, and the participants were required to decide whether a critical constituent represented a sensible continuation of the sentence. Participants were trained to respond to a response signal – a tone – presented at several times following the onset of the critical expression (*the table* in examples such as (56) and (57) above, or *survivable* in examples such as (59) and (60)). The response signal was varied across a range of times (0–4900 ms) to fully measure how the different interpretations unfolded over time.

Coerced expressions yielded lower overall levels of performance than minimally contrasting controls, whether coercion was triggered by an event-selecting verb (V–NP) or adjective (NP–AP). This suggests that comprehenders were less likely to compute a

sensible eventive interpretation of the subject or object NP when coercion was required. More importantly, however, measures of how the interpretations unfolded over time demonstrated that comprehenders computed plausible interpretations of the coerced expressions more slowly than control expressions (123 ms slower in complement coercion and 158 ms slower in NP–AP constructions). These measures provide direct evidence that comprehenders required more time to construct a sensible interpretation of the coerced complement.

3.2.3.2. *Locus of the effect.* Why do comprehenders require more time to interpret coerced complements? One explanation of these differences is that they simply reflect disruptions arising from the detection of a semantic type mismatch. There are at least two reasons to question this account. First, it leaves unanswered the question of how comprehenders actually succeed in recovering a suitable interpretation of these expressions. Second, although disruption in processing does occur when an interpretation does not make sense (e.g., Clifton, 1993; Pickering & Traxler, 1998; Traxler & Pickering, 1996), this type of explanation is not completely consistent with the delayed and sustained effects found in the reading time studies. One might have expected a mismatch effect to be evident when readers first fixate the complement, but such effects have not been found in any of the reading time studies.

Data from a recent magnetoencephalographic (MEG) study of complement coercion provides a more direct means of examining this hypothesis (Pylkkänen, Llinás, & McElree, 2004, submitted). Crucially, complement coercion does not modulate the same brain activity found in clear cases of mismatching semantic relations between a verb and its complement. Relative to control expressions such as *The author wrote the book*, anomalous expressions such as *The author amused the book* increased the activity in a left temporal source at 300–400ms (M350) – the MEG analogue of an N400 event related potential (ERP) component. However, *The author began the book* generated the same activity levels in this source as the control expression, *The author wrote the book*. Complement coercion modulated a ventromedial prefrontal source in a later 350–500 ms time-window, generating more activity in this source than either the anomalous or simpler control sentences. These findings suggest that the increased activity associated with complement coercion must involve more than simple detection of mismatching semantic properties.

Another explanation of the processing time difference attributes it to the time needed to retrieve or infer the activity implicit in the event interpretation of the coerced complement. Traxler, McElree, Williams, and Pickering (2005) tested this notion by investigating the effects of context. The idea was to place the required activity in the preceding context under the assumption that this would eliminate the cost if the difficulty involved retrieving or inferring an appropriate activity. For example, *The carpenter began the table* is most often interpreted as “The carpenter began to build the table.” Traxler et al. (2005) examine whether the coercion cost would hold in contexts such as *The carpenter was building all morning. Before he began the table....*



Importantly, this type of context manipulation did *not* eliminate the coercion cost. This speaks against attributing the cost to the time needed to retrieve the action implicit in the event sense. The findings also speak against attributing the cost to *selecting* an activity from a set of plausible actions. If expressions such as *began the book* are taxing because it is possible to interpret them in several ways [e.g., began reading, writing, translating (etc.) the book], a constraining context should have reduced the ambiguity and eliminated the cost (c.f., Altmann & Steedman, 1988; Binder & Morris, 1995; Hess, Foss, & Carroll, 1995; Pickering & Traxler, 1998).<sup>9</sup>

These results are consistent with the idea that coercion is more taxing because comprehenders must undertake more complex compositional operations to build a representation for the event sense of the complement. In other experiments, Traxler et al found that the cost was eliminated if the context provided the full event sense of the complement (e.g., *The student started/read a book in his dorm room. Before he started the book...* or *The student started/read a book in his dorm room. Before he started it...*). These experiments show that the cost of coercion can be circumvented if a relevant event sense is in the immediate discourse. Traxler et al. suggest that comprehenders were able to circumvent the costly act of building an event sense for the target expression by linking its interpretation to a relatively abstract event representation in the discourse.

In summary, the evidence indicates that these expressions are costly to process because comprehenders need to engage additional, extra-syntactic processes to generate an interpretation of the mismatched constituents. The evidence is most consistent with the idea that the cost arises from the on-line composition of an event structure.

#### 3.2.4. Aspectual coercion

Both complement coercion and type mismatch in quantifier interpretation involve semantic mismatches between a predicate and its argument. The grammaticality of these cases suggests that natural language has some set of mechanisms for fitting an argument to meet the selectional restrictions of its predicate. Does this also occur in modification?

Let us first revisit cases where Predicate Modification applies straightforwardly. In (61), the predicates (a–e) *gray*, *furry*, *four-legged*, and *creature* are all intersected to yield one complex predicate of type  $\langle e,t \rangle$ . This works because all the predicates are predicates of individuals, so they are functions that require as their input the same type of thing. Similar reasoning applies in the verbal domain. In (62), the VP *Brutus stabbed Caesar* and the modifiers (a–e) *violently*, *in the back*, and *with a knife* are all predicates of events (see Section 2.2) and can therefore happily intersect. In contrast, mixing predicates that are unambiguously predicates of individuals with predicates that are unambiguously predicates of events results in anomaly, as shown in (63).

---

<sup>9</sup> Traxler et al. (2005) provide other arguments against both of these alternative explanations.

61. Fido is a gray, furry, four-legged creature.  
 Entailments:  
 a. Fido is gray.  
 b. Fido is furry.  
 c. Fido is gray.  
 d. Fido is four-legged.  
 e. Fido is a creature.  
 Etc...
62. Brutus stabbed Caesar violently in the back with a knife.  
 a. Brutus stabbed Caesar violently.  
 b. Brutus stabbed Caesar in the back.  
 c. Brutus stabbed Caesar with a knife.  
 d. Brutus stabbed Caesar violently with a knife.  
 Etc...
63. #Fido is four-legged violently.

Now consider (64), which several researchers have argued to involve a type mismatch (Chierchia, 1984; Jackendoff, 1997; Klein & Sag, 1985; Partee & Rooth, 1983; Pustejovsky, 1991, 1995). Here the punctual verb *jump* is modified by a durative adverb. One might expect this combination to be ungrammatical, but instead, comprehenders appear to invoke a process or iterative interpretation of the verb *jump*. This reinterpretation is generally called ‘aspectual coercion’.

64. The girl jumped for three hours.

Crucially, however, (64) does not involve a type mismatch in the framework we have laid out, as both *jump* and *for three hours* are predicates of events. Consequently, they should be able to intersect. The problem though is that such an intersection does not yield an iterative reading but rather the transparent meaning in (65), which asserts that there exists a single jumping-event that lasted for three hours. This obviously clashes with our real-world knowledge about people’s jumping abilities.

65.  $\exists e$  [jumping(*e*) & agent(*e*, the girl) & for(*e*, three hours)]

Thus, in the framework we have described, aspectual coercion and complement coercion are crucially different. In complement coercion, transparent semantic composition is impossible. In aspectual coercion, transparent semantic composition is possible but it yields an anomalous output. Since sentences involving aspectual coercion are nevertheless judged felicitous, there must be some process that repairs this anomaly and changes the sort of the event described by the verb from a punctual event to a process. Given this difference between complement and aspectual coercion, comparison of their on-line processing is clearly of great interest to theories of the syntax–semantics interface. In what

follows, we first review psycholinguistic results on aspectual coercion and then return to representational issues pertaining to coercion and type-shifting more generally.

### 3.2.5. *On-line processing of aspectual coercion*

To test whether aspectual coercion imposes a significant processing load, Piñango, Zurif, and Jackendoff (1999) examined cross-modal lexical decisions to unrelated words at different points during the reading of sentences such as (66) and (67).

66. The insect hopped effortlessly until it reached the far end of the garden...

67. The insect glided effortlessly until it reached the far end of the garden...

At the durative adverb *until*, lexical decision times were longer for punctual verbs such as *hop* than for temporally unbounded verbs such as *glide*. They argued that this would not be the case if verbs such as *hop* were lexically polysemous between point-action and unbounded-activity interpretations. They suggest that the increased load reflects the deployment of on-line operations to generate a nonlexicalized, iterative sense of the verb.

Todorova, Straub, Badecker, and Frank (2000) pointed out some limitations of the Piñango et al. study. One concern was that all the sentences in the critical condition had iterative interpretations, whereas sentences in the control condition did not. It is possible that iterative readings *per se* impose a higher processing burden than noniterative readings.

Todorova et al. employed conditions illustrated in (68)–(71).

68. Even though Howard sent a large check to his daughter for many years, she refused to accept his money.

69. Even though Howard sent large checks to his daughter for many years, she refused to accept his money

70. Even though Howard sent a large check to his daughter last year, she refused to accept his money

71. Even though Howard sent large checks to his daughter last year, she refused to accept his money

Sentence (68) illustrates the condition predicted to engage more complex compositional processes, as it included a punctual verb (*send*) and a singular direct object (*a large check*) followed by a durative adverb (*for many years*). Todorova et al. point out that when a bare plural direct object (*large checks*) is an argument of the verb, as in (69), it appears to impose an iterative reading on the predicate immediately. Sentences such as (69) served as one contrast to (68). They also employed two other controls, which were analogues of (68) and (69) with adverbials (*last year*) compatible with both iterative and noniterative predicates. Processing was measured with a “stop-making sense” procedure (e.g., Boland, et al., 1995; Maunder, Tanenhaus, & Carlson, 1995), in which participants evaluated whether the sentence continued to make sense as they paced themselves through the sentence phrase by phrase.

Results indicated that participants were over twice as likely to reject sentences such as (68) than any of the control sentences (19% versus 7%, 8%, and 9%), and significantly more likely to reject these sentences at the onset of the region containing the durative modifier. Additionally, the associated self-paced times were significantly longer at the durative modifier for (68) as compared to (69), with no comparable differences evident in (70) and (71).

These findings indicate that iterative readings *per se* are not more taxing to process than noniterative readings, which alleviates one of the concerns with the Piñango et al. study. Importantly, like the Piñango et al. study, a processing cost was evident in (68) where the predicate was incompatible with the durative adverb. This pattern, however, is not consistent with the Piñango et al. claim that the source of the difficulty lies with the lexical representation of the verb. Punctual verbs paired with bare plurals (...*sent large checks*) in (69) were not difficult to interpret. This finding indicates that comprehenders can assemble without cost an interpretation compatible with the durative adverb when the verb is paired with an appropriate argument (see section 3.3 for further discussion). Todorova et al. suggest that comprehenders commit to an aspectual interpretation of the predicate only after both the verb and its arguments have been processed. This result suggests that aspect is not a lexical property stored with the verb, but instead is derived compositionally using features from the argument (see also Seegmiller, Townsend, DeCangi, & Thomas, 2004).

Do these results implicate a coercion process for aspectual mismatches analogous to what has been found with complement coercion? Conceivably they might if aspectual coercion is accomplished by introducing new semantic structure, such as an iterative operator that is not implicit in the verb's representation. However, as outlined in Section 3.2.4, there are formal reasons to doubt that the same compositional processes underlie the two phenomena. Additionally, there are other salient differences in the on-line experiments. Note that aspectual studies show evidence of a cost *only* when an interpretation of the predicate turns out to be incompatible with later material in the sentence, namely the durative modifier. In this respect, it appears more correct to view these effects as a type of semantic “garden-path.” In contrast, the cost in complement coercion appears to reflect the time it takes to establish an initial interpretation of the complement, not whether that interpretation is compatible with some material occurring later in the sentence. Crucially, if aspectual coercion were processed like complement coercion, effects should have been evident in cases like (69), where comprehenders were forced to unify the argument *large checks* with a “mismatching” punctual verb such as *sent*.

There are also other empirical grounds on which to question the commonality of the processes. Crucially, aspectual coercion does *not* engender any detectable differences in standard reading tasks (Pickering, McElree, Frisson, Chen, & Traxler, in press). We conducted two self-paced reading studies and one eye-tracking study with the Piñango et al. materials and found no measurable differences in self-paced times or any aspect of the eye movements for the conditions represented in (66) and (67). Additionally,

we collected self-paced reading times and eye-tracking measures for these sentences when the adverbial phrase (PP) was fronted, as in (72) and (73).

72. Until it reached the garden, the insect hopped effortlessly...

73. Until it reached the garden, the insect glided effortlessly...

If Piñango et al.'s arguments are correct, there should be a coercion cost associated with generating an iterative sense of *hop*. Hence, analogous to cases of complement coercion, there should be evidence of difficulty (longer times or more regressive eye movements) at and after *hopped* as compared to *glided*. We found no indication of any difficulty with (72).

Finally, we examined the Todorova et al. materials, (68)–(71) in another eye-tracking study experiment. Again, there were no differences evident in any of the conditions. Importantly, all these experiments included constructions involving complement coercion, as part of other ongoing research. In each case, we observed reliable effects of complement coercion. Hence, there is no concern that the null results for aspectual “coercion” reflect an aberrant participant population or faulty measurement procedures.

### 3.2.6. *Neurolinguistic studies of coercion*

Another way to investigate whether aspectual and complement coercion engage similar interpretive operations is to study their respective brain bases. Piñango and Zurif (2001) investigated the comprehension of both types of coercion by Wernicke's and Broca's aphasics. In the aspectual coercion experiment, three Wernicke's aphasics and three Broca's aphasics listened to coerced and transparent sentences such as *the horse jumped for an hour yesterday* and *the horse jumped over the fence yesterday*, respectively. Each sentence was followed by a comprehension question querying whether the horse jumped once or many times. Wernicke's aphasics performed at chance for coerced sentences and at 87% accuracy for transparent sentences. In contrast, Broca's aphasics performed significantly above chance for both conditions and showed no effect of sentence type. Given the small number of subjects, it is important to note that Broca's aphasics' performance was in fact numerically better on the coerced sentences than on the transparent sentences. Thus, it appears that left temporal areas are indeed important for aspectual coercion, in a way that Broca's area is not. Crucially, however, these deficit-lesion data do not discriminate between the hypothesis that aspectual coercion *occurs* in Wernicke's area and the hypothesis that Wernicke's area *provides crucial input* for aspectual coercion.

For complement coercion, two Wernicke's and three Broca's aphasics again listened to coerced (*The boy began reading the book*) and transparent sentences (*The boy began the book*) but now in a picture-matching task. The picture depicted either the correct scenario of a boy reading a book or an incorrect scenario of a boy buying a book. Contra the aspectual coercion experiment, in this study both patient groups performed numerically worse on coerced than transparent sentences although this effect only reached significance for the two Wernicke's aphasics. Thus these data are more suggestive of a main effect of coercion than an interaction. Nevertheless, the authors concluded “that

aspectual and complement coercion involve computations requiring the integrity of the left posterior cortical region associated with Wernicke’s area, but not the integrity of the left anterior cortical region associated with Broca’s area” (idib, p. 307). Although the aspectual coercion data are quite consistent with this hypothesis, the conclusion is premature for complement coercion.

Further, the complement coercion task may be difficult for Wernicke’s aphasics for reasons that are independent of coercion. In the author’s procedure, the auditory sentence and the two pictures were presented simultaneously, which means that the pictures had potentially been processed by the time the complement of the verb occurred. Consequently, at the coercion site, the listener may have had the conceptual representations of both reading and buying active. Interpretation of *the boy began the book* would then require inhibition of the competing event description. Given this, a patient population such as Wernicke’s aphasics, who have general problems with inhibitory processing (Milberg, Blumstein, Katz, & Gershberg, 1995; Wiener, Connor, & Obler, 2004), might perform poorly even if their coercion mechanism was intact. Further, interpretation of *the boy began the book* as *the boy began buying the book* is not entirely impossible. Consider, for example, an auction setting where buying is construed as a process. The fact that the “buying” interpretation is not strictly ungrammatical might increase the likelihood of patients with inhibitory problems to choose that scenario.

Although deficit-lesion methodology can tell us which areas are necessary for a certain process, identifying the brain areas responsible for that process requires methods that allow localization of function in the intact brain. As noted in Section 3.2.3.2, Pykkänen et al. (2004, submitted) investigated MEG responses to coerced sentences such as *the author began the book* and transparent sentences such as *the author wrote the book*. Contra Piñango and Zurif’s hypothesis that complement coercion occurs in Wernicke’s area, left temporal sources showed no effect of coercion. Instead, coercion elicited larger amplitudes in a ventromedial prefrontal area. This area is known to receive direct input from left superior temporal cortex (Rolls, 2004). Thus, these results would in fact be entirely consistent with Wernicke’s aphasics having trouble with complement coercion. In our study, the area that was found to be sensitive to coercion is clearly not a traditional language area. Thus, our results suggest that semantic noncompositionality might in some cases engage processes that are not specifically “linguistic.”

### 3.3. Towards a Taxonomy of Type-Shifting Rules

We have discussed several cases of nontransparent composition. Experimental investigation of these phenomena has only recently begun, and clearly many questions remain. Nonetheless, as we have emphasized, we believe that the study of the on-line processes of semantic composition should be guided by the rich and detailed hypothesis space provided by theoretical semantics. To that end, we briefly revisit each of the phenomena in Section 3 to evaluate them on representational grounds. To what extent can we predict similar or different processing operations based on what we know and can reason about their representation?

One way to classify type-shifting rules would be along the following two dimensions:

- A. Does the shift change the complexity of the type?
- B. Does the shift invoke an ontological change?

By “ontological change” we mean a change in the basic ontological category that the word or phrase denotes or predicates over. For example, complement coercion involves an ontological change from an *individual*  $\langle e \rangle$  to a predicate of *events*  $\langle s, t \rangle$ .

Type-shifting rules such as those in quantifier interpretation are shifts that change the complexity of the type but do not invoke an ontological change. For example, the sole purpose of Partee and Rooth’s (1993) “lift” rule, repeated in (74), is to change the directionality of Functional Application: An individual that would be interpreted as an argument of a function by default is converted to a higher order predicate that can take a verb as its argument. This introduces no new semantic content, hence no deep change in the meaning of the constituent.

74. Lift (enables an individual to take an intransitive verb as its argument)

$$\begin{array}{ll} \text{Sally} \rightarrow & \lambda f_{\langle e, t \rangle}. f(\text{Sally}) \\ e \rightarrow & \langle \langle e, t \rangle, t \rangle \end{array}$$

Complement coercion invokes an ontological shift *and* a change in type complexity. Therefore, one might expect it to be more costly than pure type-shifting rules such as lift.

75. Complement coercion

$$\begin{array}{ll} \text{the book} & \rightarrow \lambda e. \theta(e, \text{the book}) \\ e & \rightarrow \langle s, t \rangle \end{array}$$

How does aspectual coercion fit into this framework? As discussed in Section 3.2.4, aspectual coercion does not involve a type mismatch in the framework we have proposed, as our ontology does not distinguish between events on the basis of their temporal properties. Research on the internal temporal properties of events, called *aktionsart* or *situation aspect*, has a long history, going back to Aristotle. However, the precise way these properties should be incorporated into a type-driven system of interpretation has not been fully worked out.

Our knowledge about the internal temporal structure of events is largely knowledge about the world. Understanding the interaction of linguistic knowledge and world knowledge is no trivial matter, but an analysis of aspectual coercion requires us to make specific assumptions about it. Pustojevsky (1991, 1995), whose research has informed and inspired much of the psycholinguistic work on coercion, does not distinguish between linguistic and world knowledge but rather builds them both into complex lexical entries. In a series of papers, Dölling (1992, 1993, 1995, 1997, 2003) has developed an alternative approach that treats world knowledge as presuppositions on lexical entries.

Dölling introduces the notation in (76) for this purpose. Here, for example, the adjective *widespread* is treated as a function from individuals to truth-values, but the sorts of individuals that constitute a felicitous input for the function are restricted to kinds (e.g., *the tiger*). If the input does not denote a kind, a presupposition failure results: *#John is widespread*.

76.  $\lambda x: x$  is a kind. *widespread*( $x$ )

In Dölling's theory, these types of lexical entries figure into a two-stage model of interpretation, in which strictly compositional interpretation is followed by the fixing of various contextual parameters on the basis of world knowledge. An integral part of the second stage is the application of a rich set of productive *sort coercions*. For example, the sortal restrictions of the verb in (77) cause the semantic sort of the subject NP to shift from an institution to a person.

77. The newspaper telephoned

This type of shifting has generally been called metonymy, and Dölling is by no means the first or only person to have systematically discussed it. Although it is questionable whether the two-stage aspect of Dölling's theory is strictly applicable to real-time processing, what is appealing about Dölling's approach from a representational standpoint is that it incorporates lexical–semantic properties within a standard type-driven model of sentence–level interpretation whose basic operations are fully compositional.

In the context of this framework, it is natural to conceptualize aspectual coercion as a sortal shift within the basic category of events. How might this relate to the observed processing differences? The lack of reading time effects for aspectual coercion comparable to complement coercion suggests that sortal shifts are, unequivocally, simpler or easier for the processing system to deal with than ontological shifts. Indeed, the contrast between aspectual and complement coercion parallels a similar contrast between metonymy and complement coercion, which likewise suggests that other types of sortal shifts are not costly. Metonymic expressions such as *...read Dickens*, where *Dickens* must be interpreted as any of a number of books written by Charles Dickens, are not more taxing to process than literal expressions such as *...met Dickens*, where *Dickens* can be interpreted as the person Charles Dickens (Frisson & Pickering, 1999). This appears to be true of a wide range of shifts, including PLACE-FOR-EVENT (e.g., *protested during Vietnam*) and PLACE-FOR-INSTITUTION (e.g., *talked to the school*) metonyms (Frisson & Pickering, 1999). Crucially, as noted previously, McElree et al. (2006) showed that coercing *Dickens* to mean reading the works of Dickens, e.g., *began Dickens*, is costly relative to both *...met Dickens* or *...read Dickens*.

We suggest that the fundamental reason why aspectual coercion and other types of metonymy are not as taxing to process as complement coercion is that sortal shifts do not block initial composition. In contrast, the view developed here is that the ontological mismatch, such as those found in complement coercion, does block basic composition, and hence immediately mandates the on-line deployment of a costly coercion process.



The Todorova et al. (2000) study suggests that comprehenders do not have difficulty interpreting a predicate in either an iterative or process manner; rather, difficulty occurs *only* when the interpretation turns out to be incompatible with a durative modifier downstream. In this respect, the pattern is similar to what has been found in investigations of collective and distributed NP interpretations. Frazier, Pacht, and Rayner (1999) used eye-tracking procedures to investigate collective and distributed NP interpretations in structures such as (78)–(79).

78. Lyn and Patrick *each* saved 1000 dollars to pay for their honeymoon.

79. Lyn and Patrick *together* saved 1000 dollars to pay for their honeymoon.

80. Lyn and Patrick saved 1000 dollars *each* to pay for their honeymoon.

81. Lyn and Patrick saved 1000 dollars *together* to pay for their honeymoon.

When disambiguating evidence – either *each* or *together* – occurred immediately after the coordinate NP, there was no difference between (78) and (79). This indicates that distributed readings are as easy to compute as collective readings. However, when the disambiguating evidence occurred late, there was clear evidence for a bias for a collective reading of the coordinate NP: (80) was more difficult than (81). Apparently, comprehenders do not commit to a collective reading of the coordinate NP until after the predicate has been processed. In cases such as (80), this creates a semantic garden-path at *each*, which requires reinterpreting the subject and the predicate to derive the distributive reading.

Aspectual coercion may differ from collective/distributed interpretations and from complement coercion in that comprehenders might be able to underspecify aspectual interpretations to a greater degree. Underspecification has been argued to be possible in several domains, including quantifier scope relations (Filik et al., 2004; see Sanford & Sturt, 2002) and the processing of polysemous and metonymic expressions (e.g., Frazier & Rayner, 1990; Frisson & Pickering, 1999, 2001; Pickering & Frisson, 2001). The reading times result on aspectual coercion suggest that it might be possible for comprehenders to generate a conceptually underspecified representation that does not fix all of the contextual parameters of the predicate's interpretation. Comprehenders may be less inclined to do so in tasks that involve intrusive decisions, such as acceptability judgments (Todorova et al., 2000) or secondary tasks (Piñango et al., 1999). In short, comprehenders may need to be pressured to make commitments to see the effects of sortal incompatibilities.

#### 4. CLOSING COMMENTS

Our goal in this chapter has been to propose a framework for investigations of semantic composition. Just as formal work in syntactic theory has informed theories of parsing, we believe that the rich and substantial theorizing in semantics has the potential to guide and stimulate psycholinguistic and neurolinguistic research on real-time semantic composition.

**REFERENCES**

- Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, 18, 129–144.
- Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity*. Ph.D. thesis, Department of Linguistics, Northwestern University.
- Barker, C. (2002). Continuations and the nature of quantification. *Natural Language Semantics*, 10, 211–242.
- Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1186–1196.
- Boland, J. E. (2005). Visual Arguments. *Cognition*, 95, 237–274.
- Boland, J. E., & Boehm-Jernigan, H. (1998). Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language*, 39, 684–719.
- Boland, J., Tanenhaus, M., & Garnsey, S. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, 29, 413–432.
- Boland, J., Tanenhaus, M., & Garnsey, S., & Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from Wh-questions. *Journal of Memory and Language*, 34, 774–806.
- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from relational structure. *Journal of Memory and Language*, 51, 495–522.
- Briscoe, E., Copestake, A., & Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. *Proceedings of the 13th international conference on computational linguistics*, Helsinki, Finland (pp. 42–47).
- Carlson, G., & Tanenhaus, M. (1988). Thematic roles and language comprehension. In : W. Wilkins (Ed.), *Thematic relations*. New York, NY: Academic Press.
- Chierchia, G. (1984). *Topics in the syntax and semantics of infinitives and gerunds*. Unpublished Ph.D. dissertation, University of Massachusetts at Amherst.
- Clark, H., & Haviland, S. (1977). Comprehension and the given-new contrast. In: R. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Hillsdale, NJ: Erlbaum.
- Clifton Jr., C. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology*, 47, 222–246.

- Clifton, Jr., C. Speer, S., & Abney, S. (1991). Parsing arguments: Phrase structure and argument structure determinants of initial parsing decisions. *Journal of Memory and Language*, 30, 251–271.
- Davidson, D. (1967). The logical form of action sentences. In: N. Reschler (Ed.), *The logic of decision and action*, (pp. 81–95). University of Pittsburgh Press. Pittsburgh:
- Dölling, J. (1992). Polysemy and sort coercion in semantic representations. In: P. Bosch, & P. Gerstl (Eds), *Discourse and lexical meaning*. Proceedings of a Workshop of the DFG Sonderforschungsbereich 340, Stuttgart, pp.61–78.
- Dölling, J. (1993). Commonsense ontology and semantics of natural language. *Zeitschrift für Sprachtypologie und Universalienforschung*, 46, 133–141.
- Dölling, J. (1995). Ontological domains, semantic sorts and systematic ambiguity. *International Journal of Human–Computer Studies*, 43, 785–807.
- Dölling, J. (1997). Semantic Form and Abductive Fixation of Parameters. In: R. van der Sandt, R. Blutner & M. Bierwisch (Eds), *From Underspecification to Interpretation*. Working Papers of the Institute for Logic and Linguistics, (pp. 113–138). Heidelberg: IBM Deutschland.
- Dölling, J. (2003). Aspectual (re-)interpretation: Structural representation and processing. In: H. Härtl, & H. Tappe (Eds), *Mediating between concepts and grammar* (pp. 303–322). Berlin: Mouton de Gruyter.
- Egg, M. (2003). Beginning novels and finishing hamburgers – remarks on the semantics of to begin. *Journal of Semantics*, 20, 163–191.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164–203.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Filik, R., Paterson, K. B., & Liversedge, S. P. (2004). Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin and Review*, 11, 953–959.
- Fodor, J. D. (1982). The mental representation of quantifiers. In: S. Peters, & E. Saarinen (Eds), *Processes, beliefs, and questions* (pp.129–164) Dordrecht: D. Reidel.
- Fodor, J. A., & Lepore, E. (2002). *The compositionality papers*. Oxford: Clarendon Press.
- Frazier, L. (1999). *On sentence interpretation*. Dordrecht: Kluwer Academic Publishers.
- Frazier, L., & Clifton Jr., C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L. & Rayner, K. (1982). Making and Correcting Errors During Sentence Comprehension: Eye movements in the analysis of structurally ambiguous sentences, *Cognitive Psychology*, 14, 178–210.
- Frazier, L., Pacht, J. M., & Rayner, K. (1999). Taking on semantic commitments, II: Collective versus distributive readings. *Cognition*, 70, 87–104.

- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, 29, 181–200.
- Frege, G. (1879). *Begriffsschrift, eine der Arithmetischen Nachgebildete Formelsprache des Reinen Denkens*. Verlag von Louis Nebert, Halle.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*, 100, 25–50 (Trans by M. Black as ‘On sense and reference’). In: P. Geach, & M. Black (Eds), *Translations from the Philosophic Writings of Gottlob Frege*, (pp. 56–78) Oxford: Basil Blackwell, (1960).
- Frisson, S., & Pickering, M. J. (1999). The processing of Metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1366–1383.
- Frisson, S., & Pickering, M. J. (2001). Figurative language processing in the underspecification model. *Metaphor and Symbol*, 16, 149–171.
- Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, 89, B27–B41.
- Gillen, K. (1991). *The comprehension of doubly quantified sentences*. Unpublished Ph.D. dissertation, University of Durham.
- Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. New York, NY: Oxford University Press.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell Publishers.
- Hendriks, H. (1988). Type change in semantics: the scope of quantification and coordination. In: E. Klein, & J. Van Benthem, (Eds), *Categories, polymorphism and unification*, (96–119) Amsterdam: ITLI.
- Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124, 62–82.
- Ioup, G. (1975). Some universals for quantifier scope. In: J. Kimball (Ed.), *Syntax and semantics* (Vol. 4, pp. 37–58). New York: Academic Press.
- Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
- Jacobson, P. (1999). Towards a variable-free semantics, *Linguistics and Philosophy*, 22, 117–184.
- Kamp, H., & Partee, B. H. (1995). Prototype theory and compositionality. *Cognition*, 57, 129–191.
- Kennison, S. M. (1999). Processing agentive “by”-phrases in complex event and nonevent nominals. *Linguistic Inquiry*, 30, 502–508.

- Kennison, S. M. (2002). Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research*, 31, 65–81.
- Klein, E., & Sag, I. (1985). Type driven translation. *Linguistics and Philosophy*, 8, 163–201.
- Kratzer, A. (1996). Severing the external argument from its verb. In: J. Rooryck, & L. Zaring (Eds), *Phrase structure and the lexicon*. Dordrecht: Kluwer Academic Publishers.
- Kratzer, A. (in press). *The event argument and the semantics of voice*. Cambridge, MA: MIT Press.
- Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, 48, 243–279.
- Lapata, M., Keller, F., & Scheepers, C. (2003). Intra-sentential context effects on the interpretation of logical metonymy. *Cognitive Science*, 27, 649–668.
- Lapata, M., & Lascarides, A. (2003). A probabilistic account of logical metonymy, *Computational Linguistics*, 29, 263–317.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax–lexical semantics interface*. Cambridge, MA: MIT Press.
- Liversedge, S. P., Pickering, M. J., Branigan, H. P., & van Gompel, R. P. G. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 461–475.
- Liversedge, S. P., Pickering, M. J., Clayes, E. L., & Branigan, H. P. (2003). Thematic processing in adjuncts. *Psychonomic Bulletin and Review*, 10, 667–675.
- Marcus, M., Santorini, B., & Marcinkiewicz, M. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313–330.
- Maurer, G., Tanenhaus, M., & Carlson, G. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language*, 34, 357–82.
- McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, 32, 536–571.
- McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, 30, 115–124.
- McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, 6, 486–494.
- McElree, B., Pykkänen, L., Pickering, M. J., & Traxler, M. (2006). The time course of enriched composition. *Psychonomic Bulletin and Review*, 13, 53–59.

- McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched semantic composition. *Cognition*, 78, B15–B25.
- McKoon, G., & Macfarland, T. (2000). Externally and internally caused change of state verbs. *Language*, 76, 833–858.
- McKoon, G., & Macfarland, T. (2002). Event templates in the lexical representations of verbs. *Cognitive Psychology*, 47, 1–44.
- Milberg, W., Blumstein, S. E., Katz, D., & Gershberg, F. (1995). Semantic facilitation in aphasia: Effects of time and expectancy. *Journal of Cognitive Neuroscience*, 7, 33–50.
- Montague, R. (1970). Universal grammar. *Theoria*, 36, 373–398.
- Montague, R. (1974). *Formal Philosophy: Selected Papers of Richard Montague*. Edited and with an introduction by Richmond Thomason. New Haven, CT: Yale University Press.
- Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, 29, 259–288.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Natsopoulos, D. (1985). A verbal illusion in two languages. *Journal of Psycholinguistic Research*, 14, 385–397.
- Parsons, T. (1985). Underlying events in the logical analysis of English. In: E. LePore, & B. McLaughlin (Eds), *Actions and events: Perspectives on the philosophy of Donald Davidson* (pp. 235–267). Oxford:Blackwell.
- Parsons, T. (1990). *Events in the semantics of English: A study of subatomic semantics*. Cambridge, MA:MIT Press.
- Partee, B. H., & Rooth, M. (1983). Generalized conjunction and type ambiguity. In: R. Bäuerle, C. Schwarze, & A. von Stechow (Eds). *Meaning, use, and interpretation of language*. (pp. 361–368) Berlin: Walter de Gruyter.
- Partee, B. H., ter Meulen, A., & Wall, R. E. (1990). Mathematical methods in linguistics. *Studies in Linguistics and Philosophy* (Vol. 30). Dordrecht: Kluwer Academic Publishers.
- Phillips, C. (2001). Mechanisms for rapid use of focus information: Review of Sedivy (1997). *GLOT International*, 5, 25–32.
- Pickering, M. J., & Frisson, S. (2001). The semantic processing of verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 556–573.
- Pickering, M. J., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de Almeida. *Brain and Language*, 93, 1–9.
- Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J., (in press). Aspectual coercion and underspecification. *Discourse Processes*.

- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 940–961.
- Pylkkänen, L., Llinás, R., & McElree, B. (2004). Distinct effects of semantic plausibility and semantic composition in MEG. In: E. Halgren, S. Ahlfors, M. Hämäläinen, & D. Cohen (Eds), *Biomag 2004: Proceedings of the 14th international conference on biomagnetism*, Boston, USA: Biomag.
- Pylkkänen, L., Llinás, R., & McElree, B. (submitted). An MEG study of silent meaning.
- Pylkkänen, L., & McElree, B. (2004 ). Processing secondary predicates: Does locality matter? *Annual CUNY Sentence Processing Conference*, College Park, NY, March.
- Piñango, M. M., & Zurif, E. (2001). Semantic operations in aphasic comprehension: Implications for the cortical organization of language. *Brain and Language*, *79*, 297–308.
- Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax–semantics interface. *Journal of Psycholinguistic Research*, *28*, 395–414.
- Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*, 47–81.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review*, *3*, 504–509.
- Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain and Cognition*, *55*, 11–29.
- Rothstein, S. (1998). *Events and grammar*. Dordrecht: Boston: Kluwer Academic Publishers.
- Sauerland, U. & A. von Stechow. (2001) The syntax-semantics interface. In: N. Smelser, & P. Baltes (Eds), *International encyclopedia of the social and behavioural sciences* (pp. 15412–15418). Oxford, Great Britain: Pergamon.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*, 382–386.
- Scheepers, C., Mohr, S., Keller, F., & Lapata, M. (2004). The cost of enriched composition: Eye-movement evidence from German. *17th annual CUNY sentence processing conference*, College Park, Maryland.
- Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*, 409–431.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.
- Seegmiller, M. S., Townsend, D. J., DeCangi, D. & Thomas K. (2004). Comprehenders use event structure to develop discourse-level representations. *Annual architectures and mechanisms for language processing conference*, Aix-en-Provence, France.

- Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory and Cognition*, 26, 965–978.
- Spivey-Knowlton, M., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55, 227–267.
- Tenny, C., & J. Pustejovsky. (2000). *Events as grammatical objects*. Stanford: CSLI Publications.
- Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. *Proceedings of the twenty-second Annual Conference of the Cognitive Science Society* (pp. 3–8). Philadelphia, PA.
- Traxler, M., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye-movements. *Journal of Memory and Language*, 53, 1–25.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454–475.
- Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 47, 530–547.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–553.
- Tunstall, S. L. (1998). *The interpretation of quantifiers: Semantics and processing*. Unpublished Ph.D. dissertation, University of Massachusetts, Amherst.
- Varvoutis, J., & Hackl, M. (2006). Parsing Quantifiers in Object Position. *The 19th annual CUNY sentence processing conference*, CUNY Graduate Center, New York, NY.
- Wason, P., & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, 31, 591–597.
- Wiener, D. A., Connor, L. T., & Obler, L. K. (2004). Inhibition and auditory comprehension in Wernicke’s aphasia. *Aphasiology*, 18, 599–609.



This page intentionally left blank

## Chapter 15

### Constraint Satisfaction Accounts of Lexical and Sentence Comprehension

*Maryellen C. MacDonald and Mark S. Seidenberg*

Consider what it takes to understand an ordinary sentence such as *The man bought a tie with tiny white diamonds*. Part of your understanding includes that *the man* is the agent of the action *bought* and that *a tie* is the thing being bought. To get this far, you also need to understand that *man* and *tie* are nouns and not verbs, although the verb usage of these words is possible in other contexts, such as *man the boats* or *tie your shoes*. You also need to understand that *a tie* refers to neckware, not to a game with equal scores, and that *tiny white diamonds* are an attribute of the tie, and not the currency used to purchase the tie (cf. *The man bought a tie with his credit card*). Despite these and many other possibilities where interpretation could go wrong, the odds are in favor of your interpreting this sentence correctly. For example, *man* is more common as a noun than a verb, so a comprehender who unconsciously goes with the best odds will get to the right interpretation here. Similarly, words that follow determiners such as *the* and *a* are far more likely to be nouns than verbs, and tiny white diamonds are unlikely to be offered in trade for haberdashery, at least in most cultures. Comprehenders who follow the most likely alternatives will get to the correct interpretations of these aspects of the sentence. The idea that language comprehension is a process of following likely alternatives to derive an interpretation of ambiguous input forms the basic claims of constraint satisfaction, or constraint-based, theories of language comprehension. As in these examples, what is a likely alternative depends on properties of both individual bits of information (e.g., the frequency with which a word is used as a noun or verb) and combinations of bits of information (e.g., *the + man* or *a + tie*). Constraint-based theories emphasize how people learn, represent, and use such probabilistic information. This chapter will provide an overview of this approach, including its history, how it compares to alternative views, and a description of the kinds of computational mechanisms that are thought to underlie learning and using such constraints.

#### 1. TRADITIONAL VIEWS OF LEXICAL AND SYNTACTIC AMBIGUITY

As the sentence about the man and his new tie illustrated, ambiguity is ubiquitous in language. This chapter will focus on two main types of ambiguity: *lexical ambiguity*, illustrated by the multiple meanings of words such as *tie*, and *syntactic ambiguity*,

illustrated by the alternative interpretations of *with tiny white diamonds* as something describing the tie (thus modifying a noun) or describing the method of buying (modifying a verb). (See the Pickering and van Gompel, Kluender, and Tanenhaus and Trueswell chapters for discussions of other types of ambiguity.) The two kinds of ambiguity can interact; for example, adopting the noun vs. verb interpretation of *man* affects how one interprets the syntactic structure of a sentence containing this word. Despite the close relationship between these two types of ambiguity, for much of the history of modern psycholinguistics they have been studied independently. This division reflected differing views about lexical and syntactic representations (MacDonald, Pearlmuter, & Seidenberg, 1994). The meanings and other properties of words have often been thought to be stored in the *lexicon*, a person's mental dictionary. On this view, interpreting words involves looking up, or *accessing*, information in the lexicon. This process was thought to be autonomous, proceeding in the same way regardless of the context in which a word occurred (Tanenhaus, Leiman, & Seidenberg, 1979; Swinney, 1979). It was also thought to make minimal demands on limited capacity working memory and attentional resources, allowing multiple meanings of words to be accessed in parallel. This led to a two-stage model of lexical ambiguity resolution. In the first stage, the lexical system accessed the common meaning or meanings of words; in the second stage, information derived from the linguistic and extra-linguistic contexts and the comprehender's background knowledge were used to select the appropriate meaning and integrate it into the developing representation of the sentence (see Simpson, 1981, for review).

Syntactic structures, in contrast, were standardly assumed to be *constructed* by a mental parser on the basis of grammatical rules. Deriving sentence structure was assumed to place demands on working memory and attentional resources that are limited in capacity (Frazier, 1987; Gibson, 1998; Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992). These memory demands caused the parser to pursue only a single interpretation of syntactic structure at a time. This also led to a two-stage model. In the first stage, general parsing principles were used to assign a candidate syntactic structure online; in the second stage, other types of knowledge were utilized to flesh out this representation (e.g., interpret it semantically) and to revise the initial analysis if it were discovered to be incorrect.

Both lexical and syntactic accounts were motivated in part by appeals to the notion that language consists of distinct modules involving different types of information and processes; however, in the lexical case, this resulted in multiple alternatives being considered in a parallel process, whereas in the syntactic case, this resulted in a single interpretation being considered in a serial process (see MacDonald et al., 1994; Tanenhaus & Trueswell, 1995, for reviews). The modular view was consistent with distinctions between the lexicon and syntax in grammatical theories that were prominent at the time the two-stage accounts were being developed (Newmeyer, 1980). The two-stage approach was also justified on the basis of assumptions about processing capacity limitations and the need to analyze the linguistic input very rapidly. The route to efficient interpretation was thought to be via a two-stage system in which the preliminary first-stage analysis prevented the input from being lost from working memory; the burden on working memory limitations was reduced because processing at this stage was limited to certain types of information,

e.g., syntactic structure. The initial interpretation could then be refined, corrected, and elaborated in the second-stage analysis.<sup>1</sup> This attention to the time pressures of language comprehension and to the notion that processing may proceed through several distinct stages was reflected in the use of behavioral measures that were closely time-locked to the language input (e.g., tracking eye-movements, cross-modal priming). In various forms, two-stage approaches formed the dominant theoretical framework for word and sentence comprehension through the 1980s, and the focus on the time course of processing continues to this day.

The alternative view, which came to be called constraint-based language comprehension (or comprehension via probabilistic constraints), emerged in the 1990s. This approach challenged essentially every major tenet of the two-stage accounts. Whereas the two-stage theories held that comprehension consists of discrete stages at which different types of information and processes are used, constraint-based theories viewed comprehension as continuous and homogeneous, with the same types of information and processes in use at all times. Whereas the two-stage theories assumed that processing limitations restrict the types of information that initially guide the comprehension process, constraint-based theories emphasized the richness of the linguistic signal, the capacity of language users to learn this information over time, and the comprehender's capacity to bring this information rapidly to bear on the input during real-time comprehension processes.

## 2. SOURCES OF THE CONSTRAINT-BASED APPROACH

The constraint-based approach emerged from advances in several areas, including linguistic theory, corpus linguistics, psycholinguistics, and computational modeling.

### 2.1. Changing Views about Linguistic Structure

Whereas two-stage models reflect early approaches within generative grammar in which lexical and syntactic information were held to be separate, the constraint-based approach to comprehension is more closely related to work within linguistics in which (to varying degrees) lexical and syntactic representations are closely related (e.g., Bresnan, 1982; Chomsky, 1981; Joshi, 1985). The lexical representation of a word might include not only information about its spelling, pronunciation, and meaning(s) but also its grammatical functions and the types of syntactic structures in which it participates. It is a small step to then envision this information as part of a large interactive network (MacDonald et al., 1994). Under this scenario, the computation of both "lexical" and "syntactic" information in sentence comprehension is governed by a common set of lexical processing mechanisms.

---

<sup>1</sup> This two-stage process is reminiscent of practices in machine translation, in which an automatic but limited first pass analysis (by machine) is then corrected and elaborated in a second-stage analysis (by a human translator).

## 2.2. Changing Views about Context

Language is comprehended essentially as it is perceived (Marslen-Wilson, 1975), and so a central question is what types of information can be brought to bear on decoding and interpreting the incoming signal. Studies of the role of language context in comprehension have also undergone a significant shift over the years. Research in the two-stage era focused on the use of real-world knowledge in guiding the comprehension process, and on the difficulties inherent in accessing relevant information online (Kintsch & Van Dijk, 1978). People know a vast amount about the world; as research on natural language processing in artificial intelligence suggested, it is a difficult problem to design a comprehension system that accesses relevant information from an enormous database of facts (Hayes-Roth & Jacobstein, 1994). Moreover, several studies emphasized the ineffectiveness of context, suggesting that comprehenders were limited in their application of real-world knowledge during comprehension (Forster, 1979), that context facilitated lexical processing only when words were highly predictable (Fischler & Bloom, 1979), and that this very strong degree of contextual constraint is rare in naturally occurring texts (Gough, Alford, & Holly-Wilcox, 1981). These results led to the conclusion that context-based prediction was not an important component of comprehension.

Complementary findings emerged from the study of lexical ambiguity resolution (e.g., Swinney, 1979; Tanenhaus et al., 1979). Many words are ambiguous between semantically distinct meanings (e.g., WATCH: a timepiece, to look; BANK: a monetary institution, the ground bordering a river). These early studies examined the processing of ambiguous words for which there are two main meanings that are used approximately equally often in the language (“equibiased” ambiguities). The main finding was that subjects initially activated multiple meanings, even in contexts that were highly disambiguating. For example, the contexts in (1) and (2) clearly disambiguate the word ROSE. Yet subjects showed priming (facilitation compared to an unrelated control) for target words related to both of the main meanings (e.g., FLOWER, STOOD) presented immediately following each sentence (Tanenhaus et al., 1979). Results such as this were taken as evidence that comprehenders initially activated the common meanings of ambiguous words and within about 250 ms selected the correct meaning based on the context. Here too the processing of words seemed to be independent from processes involved in integrating a sequence of words into a meaningful, syntactically structured representation.

1. They all ROSE.
2. He bought a ROSE.

The research on predictability effects and lexical ambiguity resolution led many researchers to conclude that context effects are relatively weak, with the result that theories instead emphasized bottom-up aspects of processing – how words are identified. The ambiguity research played an important role in Fodor’s (1983) development of his concept of modularity. The lexicon was seen as a paradigmatic example of an autonomous module in the comprehension system.

Subsequent research has led many of these conclusions to be revised. Whereas the word predictability studies initially argued for a limited role of context, later work suggested that context effects could operate at levels other than predicting specific words. Studies of semantic priming, for example, suggested that the processing of a word is facilitated when preceded by a word with which it shares semantic features (e.g., McRae, de Sa, & Seidenberg, 1997). Here the target words are not predictable, but facilitation occurs nonetheless. Moreover, it is worth noting that most of the studies which suggested that context effects are limited in scope examined reading rather than spoken language. Written language does not exhibit many of the properties that make speech perception such a difficult computational problem (e.g., variability with respect to rate, pitch, accent; relatively lower signal – noise ratios; co-articulation and the absence of definitive markers for phoneme or word boundaries). The spoken code seems inherently more context bound, insofar as the mere perception of sounds depends on the contexts in which they occur (e.g., Samuel & Pitt, 2003).

As with the context research, the lexical ambiguity research was similarly reexamined. Whereas initial studies had argued for activation of multiple meanings of ambiguous words independent of context, subsequent research yielded a more complex picture. Several studies showed that contextual information could result in only one meaning of an ambiguous word being considered online (e.g., Simpson & Kreuger, 1991). However, other studies showed that context could not override all aspects of lexical knowledge, in particular the relative frequencies of the meanings: there was still an ambiguity effect (computation of multiple meanings) when contexts favored the less-frequent meaning of an ambiguous word (Duffy, Morris, & Rayner, 1988). Thus, the system is apparently neither strictly modular nor completely context-bound. Kawamoto (1993) developed a computational model that provided insight about these results. His system was not inherently modular, insofar as nothing architectural prohibited contextual information from affecting meaning activation. However, in practice lexical information was activated more rapidly, limiting the effects of context. This is because there is a much closer relationship between the spelling or sound of a word such as ROSE and its meanings than there usually is between either of the meanings and the contexts in which they occur.

Finally, researchers began to question a key assumption underlying much of the research on lexical ambiguity: that words have discrete meanings that can be accessed like entries in a mental dictionary. The meaning of a word routinely shifts as a function of the context in which it occurs. Consider a word such as *piano*. It has a seemingly simple, unambiguous common meaning: large keyboard instrument with steel wires struck by felt-covered hammers (we are ignoring here the secondary musicological sense meaning “soft in volume”). Yet different shades of meaning are involved in pushing a piano (where weight is relevant but musical properties are not) vs. playing a piano (where the opposite is true; Merrill, Sperber, & McCauley, 1981). How to properly characterize meanings is a difficult issue that has been addressed from many theoretical and disciplinary perspectives (Margolis & Laurence, 1999). Here it is sufficient to note that it may be an essential property of word meaning that it is computed in a context-dependent manner every time a word is comprehended. This type of computation seems inherently at odds with a modular lexicon that automatically and independently

activates stored meanings and passes them along to other comprehension systems. The creation of novel meanings from proper nouns (Clark & Clark, 1977) and the interpretation of novel noun compounds (Gagne & Shoben, 1997) raise similar issues.

We do not have a general theory of lexical ambiguity resolution in hand; to have one would be to solve a good part of the problem of language comprehension. However, this research made it clear that a broad range of factors involving properties of both words and contexts affect lexical ambiguity resolution, and that the interactions among these many factors determine the outcomes that are observed.

### 2.3. Changing Views about Language Statistics

Languages exhibit statistical structure – variations in the distributions of elements such as sounds, words, and phrases. Despite the existence of this structure, for many years statistical analyses of language attracted little interest within mainstream linguistics and psycholinguistics, principally because Chomsky (1957) compellingly argued that language exhibits important properties that are not captured by mere statistics (as “Colorless green ideas sleep furiously” illustrated). According to the probabilistic constraints approach, however, comprehension essentially *is* the process of exploiting statistical regularities of many kinds. Learning and using language seem like difficult problems (ones that necessarily require innate grammatical knowledge, or learning or parsing mechanisms) only because this statistical information was systematically excluded from theorizing.

The ground-breaking studies that expanded notions about the range of information that might be used in sentence comprehension were Bever (1970) and Ford, Bresnan, and Kaplan (1982). In a classic article, Bever (1970) made a number of important observations concerning syntactic complexity and ambiguity and the factors that can make sentence comprehension difficult. Bever suggested that comprehenders are guided by *perceptual strategies* that assign interpretations based on frequency and plausibility. He described a specific strategy whereby comprehenders interpret noun-verb-noun sequences as agent-action-object. Violating this expectation (as in Bever’s example “The horse raced past the barn fell”) creates a misanalysis, which came to be known as a “garden path” effect (Frazier, 1978). Ford, Bresnan, and Kaplan (1982) provided an early investigation of the effects of lexical knowledge on sentence comprehension. They proposed that comprehenders initially adopt an analysis of a syntactic ambiguity that incorporates the most frequent subcategorization of the sentence’s verb (see also Fodor, 1978). Verb subcategorization refers to the noun phrase arguments a verb may take; for example *move* may or may not have a direct object noun phrase. Ford et al. provided evidence consistent with the idea that the several subcategorization options were ordered by frequency, and that comprehenders consider sentence interpretations in the corresponding order.

Although their importance was widely recognized, the Bever and Ford et al. articles did not immediately generate a program of research. One problem that inhibited further progress was that the research tools that were available did not make it easy to calculate

robust language statistics from large samples of text or discourse. This problem was largely obviated in the 1990s, when resources such as the *Wall St. Journal* corpus (Marcus, Santorini, & Marcinkiewicz, 1993) became publicly available and could be analyzed using desktop computers. This methodological advance made it possible to conduct behavioral studies examining the use of various types of statistical information in comprehension (discussed below). A second problem was the absence of a theory that could explain which language statistics are relevant, and how they could be learned, represented in memory, and efficiently used in processing. In the absence of such a theory, it was not obvious how the Bever and Ford et al. findings could be extended. This problem also began to be addressed in the 1990s, with advances in the theory of statistical learning within the connectionist framework, to which we now turn.

#### 2.4. Development of the Connectionist Paradigm

The term “connectionism” refers to a broad, varied set of ideas, loosely connected (so to speak) by an emphasis on the notion that complexity, at different grain sizes or scales ranging from neurons to overt behavior, emerges from the aggregate behavior of large networks of simple processing units. Our focus is on the parallel distributed processing (PDP) variety developed by Rumelhart, McClelland, Hinton and others in the 1980s (McClelland, Rumelhart, & Hinton, 1986). This approach includes a variety of concepts that are potentially relevant to language. In brief, PDP networks consist of large numbers of simple processing units that take on activation values. The connections between units carry weights that determine how activation is passed between units. The network is configured to perform a task (such as recognizing a word or object, or predicting the next word in a sentence). Learning involves gradually adjusting the weights on connections. The problem is to find a set of weights that yields performance that corresponds to human performance on the task (e.g., with respect to accuracy, generalization, developmental trajectory). Several algorithms can be employed for this purpose; they vary in how closely they mimic properties of learning at neural or behavioral levels (see Harm & Seidenberg, 2004, for discussion). Network performance is determined by several main factors: (1) the architecture of the system (e.g., the configuration of units and connections); (2) the characteristics of the input and output representations; (3) characteristics of the patterns used in training the model; and (4) characteristics of the learning algorithm. In other words, the model’s performance depends on its initial state, what it experiences, and how it learns from those experiences.

This theoretical framework has been discussed extensively elsewhere; here we focus on three properties that inform the probabilistic constraints approach to comprehension.

First, the networks incorporate a theory of statistical learning. The main idea is that one way that people learn (there may be others) is by gathering information about the frequencies and distributions of environmental events. This type of learning is thought to be general rather than language specific. Many nonhuman species are also capable of rudimentary forms of statistical learning (Estes, 1955); humans may be distinct with respect to the power of their statistical learning capacities. Language, for example, requires



tracking correlations and covariation across multiple types of linguistic information within and across modalities (e.g., a speech signal and the context in which it is uttered).

The applicability of these ideas to language was initially explored in the context of learning inflectional morphology (Rumelhart & McClelland's, 1986, past tense model) and learning to read (Seidenberg & McClelland, 1989; Plaut McClelland, Seidenberg, & Patterson, 1996; Harm & Seidenberg, 2004). The reading models in particular developed the idea that lexical knowledge consists of statistical relations between orthographic, phonological, and semantic codes. Learning then involves acquiring this statistical knowledge over time. Subsequent research on statistical learning in infants and adults has provided strong evidence consistent with this view. A wealth of studies now attest to humans' robust abilities to learn statistical patterns that inhere in diverse types of stimuli (Saffran & Sahni, in press). The domain-generalty of statistical learning is suggested by studies showing that infants are equally good at learning the statistical structure in a series of spoken syllables and a series of pure tones (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999), and by similarities across auditory (Saffran et al., 1996) and visual (Kirkham, Slemmer, & Johnson, 2002) modalities. This learning mechanism provides a way to derive regularities from relatively noisy data, a property that is likely to be highly relevant to the child's experience in learning language. Although some researchers have argued that specifically grammatical relationships are not acquired by statistical learning (e.g., Marcus et al., 1999; Peña, Bonatti, Nespor, & Mehler, 2002), these claims have been challenged (Perruchet, Tyler, Galland, & Peereman, 2004; Seidenberg, MacDonald, & Saffran, 2002).

Second, the models provide a basis for understanding why particular types of statistics are relevant and not others. Above we described the main factors that determine a model's behavior (and, by hypothesis, a person's). Note that this description did not include a specification of which *types* of statistics a model should compute. It is not necessary to stipulate this in advance because this aspect of a model's behavior falls out of the other factors. In practice, what a model learns is heavily determined by the nature of the representations that are employed. These representations (e.g., of phonology or semantics) are intended as (simplified) claims about what people know and bring to a task such as language learning. This knowledge may be innate or may itself be learned by processes to be explored in other models. The goal is to endow a model with exactly the knowledge and capacities that people (infants, children, adults) bring to learning a task, although this ideal is only approximated in any implemented model. Given the properties of these representations, other aspects of the model architecture (e.g., number of units or layers; patterns of connectivity between layers), and a connectionist learning algorithm, the model will pick up on particular statistical regularities implicit in the examples on which the model is trained. Thus, motivating the various elements of a model and how it is trained is very important, but the model itself determines which statistics are computed.

This discussion is relevant to a concern that is often voiced about connectionist models, that they are too powerful – capable of learning regularities that humans cannot learn. In fact, what such models learn is highly constrained. Constraints on what is

learned arise not so much from the learning algorithm itself as from other aspects of the network, particularly properties of the representations that are used. For example, models that represent articulatory or acoustic primitives in a realistic way are constrained by facts about what people can say or hear.

Third, the framework provides a powerful processing mechanism, the exploitation of multiple simultaneous probabilistic constraints. Information in a network is encoded by the weights. The weights determine (“constrain”) the output that is computed in performing a task. Processing involves computing the output that satisfies these constraints. This output changes depending on what is presented as input (e.g., the current word being processed in a network that comprehends sentences word by word).

This type of processing, known as constraint satisfaction, has several interesting properties. One is that the network’s output is determined by all of the weights. Such models illustrate how a large number of constraints can be utilized simultaneously without imposing excessive demands on memory or attention. Constraint satisfaction is passive – activation spreads through a network modulated by the weights on connections – rather than a resource-limited active search process. Another important property is that the constraints combine in a nonlinear manner. Bits of information that are not very informative in isolation become highly informative when taken with other bits of information. Much of the power and efficiency of the language comprehension system arises from this property. Languages exhibit many partial regularities. Different types of information are correlated, but weakly. The comprehender cannot wholly rely on any one type of information, but combinations of these partial cues are highly reliable. This concept may seem paradoxical at first. If individual cues are unreliable, wouldn’t combinations of these cues be even more unreliable? No, not if cue combination is nonlinear. The informativeness of each cue varies as a function of other cues. This point is easy to grasp by illustration. Someone is thinking of an object – guess what it is. The cues are it is a fruit, it is yellow, and its name begins with B. In isolation, each cue only weakly constrains the answer. The combination of cues, however, makes it very likely that the object is *banana*.

The same process can occur on a sentence or discourse level. In the context of a discussion of shopping and the syntactic environment of the determiner *a*, the word *tie* probably refers to neckwear. This contingency holds despite the fact that all the simple probabilities are quite low – by itself, a shopping context doesn’t demand that neckware be discussed, the occurrence of *a* does not predict the word *tie*, and *tie* in isolation affords several more frequent interpretations than the neckware one.

The bases of constraint satisfaction systems have been explored extensively in the computational literature. Connectionist models provide one way of implementing this process, but there are symbolic systems that perform similarly (Mackworth, 1977). In the psycholinguistic literature, the basic idea was introduced in Bates and MacWhinney’s (1989) Competition Model. Bates and MacWhinney argued that language is comprehended by following “cues” that compete with one another and are weighed as a function of their effectiveness in past comprehension events. The Competition Model incorporated

the important ideas that many linguistic cues are learned and language-specific; that cues could conflict and be differentially weighed; and the importance of integrating syntactic and nonsyntactic information during comprehension. The probabilistic constraint approach can be seen as coupling many ideas embodied by the Competition Model with proposals about the statistical basis of cues (“constraints”), and how multiple constraints are learned, represented, and exploited in processing. The Competition Model has been very important in research on how children acquire knowledge of language-specific cues, how languages differ with respect to the relative prominence of different cues (e.g., word order vs. inflectional morphology), and how cue competition affects the final interpretation of a sentence. The model had less to say about the integration of many simultaneous probabilistic cues, or about online processes in comprehension (see Elman, Hare, & McRae, 2004, for discussion). Also, in the connectionist models we have described, different alternative interpretations do not directly compete. The same weights are used in processing all input patterns. The performance of the model (or person) depends on the aggregate effects of exposure to many examples. There is nothing like parallel activation of multiple alternatives, just the computation of the best-fitting output. “Competition” is realized only implicitly, because alternatives have affected the weights, not by explicitly computing and comparing alternatives.

In summary, the probabilistic constraints approach emphasizes the role of statistical information concerning the occurrence and co-occurrence of different types of linguistic and nonlinguistic information in language comprehension. Learning a language involves acquiring this information from the large sample of utterances to which every learner is exposed. The theory assumes that humans are born with (or soon develop) capacities to perceive particular kinds of information (e.g., in listening), to engage in statistical learning, and to encode what is learned in networks of neurons. Familiar types of linguistic representation such as phonemes, syllables, morphemes, words, and constituents are not represented directly in memory; rather these terms are approximate descriptions for higher level statistical generalizations that emerge with experience (e.g., Seidenberg & Gonnerman, 2000). On this view, the learner builds (or “bootstraps”) a language out of statistical relations among different types of information, and skilled language comprehension involves using these statistical generalizations in processing utterances. These ideas have been extensively explored in the context of syntactic ambiguity resolution, to which we now turn.

### 3. PROBABILISTIC CONSTRAINTS AND SYNTACTIC AMBIGUITY RESOLUTION

Syntactic ambiguities arise when a sequence of words is compatible with more than one sentence structure. Often the syntactic ambiguity coincides with a lexical ambiguity of some sort. For example, in (3), there is an ambiguity between interpreting *Carol* as the noun phrase (NP) direct object of the verb *saw* or the beginning of a sentential complement (often termed the NP/S ambiguity). This ambiguity is linked to lexical ambiguity in the verb, which can optionally take either a direct object NP or a sentential complement. The example also illustrates another common feature of syntactic ambiguities, at least in

English, that they may be triggered by the omission of an optional word or phrase. Thus, the sentential complement sentence in (1c) could be introduced with *that* (Shanta saw that Carol ...), which would remove the temporary ambiguity.

- 3a. Temporary ambiguity: Shanta saw Carol ...
- 3b. NP direct object interpretation: Shanta saw [Carol], but Carol didn't see her.
- 3c. Sentential complement interpretation: Shanta saw [Carol would be late].

A dominant concern in syntactic ambiguity resolution has been the *timecourse* over which information is brought to bear on the ambiguity. The modular two-stage account is exemplified by Frazier and colleagues' Garden Path Model (e.g., Frazier, 1987; Frazier & Clifton, 1996), in which the first-stage *parser* (the syntactic interpretation component of the comprehension system) develops a syntactic structure for the input, guided by only the lexical categories of the input words (*noun*, *verb*, etc.), the syntactic rules of the language, and by structure-based heuristics (Minimal Attachment and Late Closure) that direct structure building when more than one alternative structure is afforded by the input. At some later point, a second stage integrates semantic and contextual information into the representation, and if this information conflicts with the initial interpretation built by the parser, the conflict may trigger a revision and reanalysis of the input.

The constraint-based view argues that the preference for one interpretation over another during comprehension of an ambiguous sentence stems not from global heuristics such as Minimal Attachment but from the rapid combination of many probabilistic constraints. A key observation concerning such constraints is that different types of information tend to be correlated; for example, a verb's meaning is strongly related to the kinds of noun arguments it tends to appear with in sentences (Hare, McRae, & Elman, 2003; Levin, 1993; Roland & Jurafsky, 2002). As a result, even weak cues can combine with other correlated cues and have a strong effect on interpretation preferences. Thus the approach links syntactic level information, such as knowledge about transitive sentence structures (those with a direct object in the verb phrase), to lexically specific information, such as the frequency with which a particular verb (*bought*, *say*) occurs with a direct object, the frequency with which a noun (e.g., *tie*) occurs as a direct object, and the conjoint frequency with which *bought* and *tie* co-occur in a verb/direct object configuration. The correlation of cues has an important role in understanding how abstract pragmatic constraints, often thought to be too complex to be brought to bear in online ambiguity resolution, could have a rapid effect on the process. For example, new entities introduced into a discourse are more likely to receive additional modification than are previously mentioned (or "given") noun phrases, thus affecting the probability that syntactically ambiguous prepositional phrase will modify this noun phrase. The given/new distinction is strongly correlated with the type of determiner used to introduce the noun phrase; new entities often occur with *a*, and given ones with *the*. Thus *a tie with* is more likely to have the *with* phrase modify the tie than is the sequence *the tie with...* (Spivey-Knowlton & Sedivy, 1995). This pattern reflects how discourses are structured and might require extensive computation in some cases, but the comprehender has a ready proxy in the simple co-occurrence of some determiners and the interpretation of *with*.

The contrast between two-stage and constraint-based accounts has often focused on the extent to which the separate processing stages posited by two-stage models are isolable. In the case of the Garden Path model, in which a purely syntactic first stage is followed by use of all other types of information in a second stage, the issue is the extent to which putatively second-stage nonsyntactic information could be shown to affect the operations of the first-stage parser. A significant body of work in the 1980s and 1990s used eye fixations during reading to address this issue, and a number of researchers suggested that the earliest eye fixations on a small region of text reflected operations of the first-stage parser, while later fixations were driven by second-stage semantic integration processes (e.g., Rayner, Carlson, & Frazier, 1983). This view was motivated in part by studies in which manipulations of semantic information in syntactically ambiguous sentences were found to affect late eye fixations, but not early ones (e.g., Ferreira & Clifton, 1986; Rayner et al., 1983). Subsequent studies suggested that the delayed effects of nonsyntactic information in these reading patterns were attributable to weak or infelicitous contexts or other biases in the ambiguous stimuli (Altmann & Steedman, 1988; Crain & Steedman, 1985). More robust manipulations of context have shown clear evidence of the use of nonsyntactic information in first pass reading measures (e.g., Garnsey, Pearlmuter, Meyers, & Lotocky, 1997; Trueswell, Tanenhaus, & Garnsey, 1994) and even in the first fixation durations on words (Speer & Clifton, 1998), a measure that has often been taken as the earliest processing evidence obtainable with eyetracking (Rayner et al., 1983). Moreover, as the nature of contextual effects received additional investigation, the number of potentially relevant constraints, and the interactions between them, grew more complex. This trend can be illustrated by considering one particular ambiguity in detail.

### 3.1. An Example: Main Verb vs. Reduced Relative Ambiguities

The structures considered here are probably the most thoroughly studied in psycholinguistics. The focus on these structures arose from Bever's (1970) observation that whereas the sentence *The horse raced past the barn fell* is taken to be gibberish by most speakers of English, it is readily comprehended when two optional words (a relative pronoun and a form of be, such as *that was*) are inserted marking the start of a relative clause, as in *The horse that was raced past the barn fell*. Another example, somewhat easier to comprehend, is given in (4). This is called the Main Verb/Reduced Relative (MV/RR) ambiguity because it is initially unclear whether the first verb, *raced* in Bever's example and *arrested* in (4), is the main verb of the sentence (as in 4b) or is introducing a reduced relative clause (4c). The clause is said to be "reduced" because of the omission of the optional relative pronoun and a form of *be*.

- 4a. Temporary Main Verb/Reduced Relative Ambiguity: The three men arrested...
- 4b. Main Verb Interpretation: The three men arrested the burglary suspects in a parking garage.
- 4c. Reduced Relative Interpretation: The three men arrested by the local police were wanted in connection with the jewel robbery.

Early studies of interpretation of this ambiguity manipulated the degree of contextual information consistent with the reduced relative (RR) interpretation and found strong misinterpretation or “garden-path” effects in reading patterns at all levels of contextual support, indicated by long reading times in the sentence region that disambiguated the ambiguity (Rayner et al., 1983; Ferreira & Clifton, 1986). These reading patterns were taken to indicate that comprehenders initially adopted the main verb (MV) interpretation (the one favored by the parsing heuristic Minimal Attachment) independent of context and were surprised when the disambiguation favored the reduced relative interpretation. Subsequent studies explored the nature of contexts in depth and suggested that interpretation is guided by a number of probabilistic constraints, with the difficulty of a given interpretation of the ambiguity typically a function of several constraints acting together. Some of the major categories of constraints are listed below.

- A. *Animacy of the pre-verbal NP* (e.g., *men*), as this affects the likelihood that this noun will be the agent vs. patient of an upcoming verb, in that animate nouns are more typical agents. This constraint is important because the noun is the agent of the next verb in the MV interpretation, and it is the patient of the verb in the RR interpretation (Trueswell et al., 1994; but Ferreira & Clifton, 1986 failed to find animacy effects).
- B. *The relative frequency of usage of the ambiguous verb* (e.g., *arrested*) in active vs. passive structures, as the MV interpretation is an active structure while the RR is a passive. Active/passive voice frequency is related to several intercorrelated properties of the verb, including the verb’s frequency of occurrence in the past tense (required for the active MV interpretation) vs. past participle (required for passives and the RR interpretation), and its relative frequency of uses in transitive (with a direct object) vs. intransitive (no direct object) constructions. The RR interpretation is always transitive, but the MV may be intransitive (Hare, Tanenhaus, & McRae, 2006; MacDonald, 1994; MacDonald et al., 1994; Trueswell, 1996).
- C. *Plausibility of the pre-verbal NP as an agent vs. patient of the ambiguous verb*, such as the plausibility that *men* would be the agent vs. patient of *arrested*, what McRae, Ferretti, & Amyote (1997) called *thematic fit*. It is an example of a *combinatorial* constraint, in that it integrates properties of at least two words and the information in constraints A-B above (McRae, Spivey-Knowlton, & Tanenhaus, 1998; Pearlmuter & MacDonald, 1992; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994). The power of such combinatorial constraints can be seen in several reanalyses of failures rapid effects of some simple constraint. For example, studies that found to find only minimal effects of noun animacy or other broader discourse plausibility factors (Ferreira & Clifton, 1986; Rayner et al., 1983) tended to use stimuli with verbs that strongly promoted an active, intransitive interpretation. In this situation, verb biases were working strongly in favor of the MV interpretation, and combinatorial constraints (over properties of both verbs and nouns) had little effect (MacDonald et al., 1994; Tanenhaus & Trueswell, 1995; Trueswell et al., 1994).
- D. *The basic frequency of the MV vs. RR structure*. Within two-stage models, the initial preference for MV structures stems from parsing heuristics such as Minimal Attachment, but within the constraint-based tradition, this effect emerges from the

fact that MV structures are more common than RR structures in the language (Bever, 1970; McRae et al., 1998).

- E. *The nature of the words after the onset of the ambiguity.* In most empirical studies, the first few words after the ambiguous verb constitute a prepositional phrase, such as *by the local police* in sentence (4c) above. Depending on their lexical properties and that of the preceding material, the post-ambiguity words may serve to promote one or the other interpretation of the ambiguity. The constraints here can be simple, such as the basic probability that *by* refers to an agent of an action (promoting the passive and thus an RR interpretation) vs. a location (less constraining for the two alternative interpretations), or the constraints may be combinatorial, such as properties of *by* given a particular preceding verb or NP, as in *by + the local police* (MacDonald, 1994; McRae et al., 1998). Following the prepositional phrase, the relative clause typically ends in most stimulus materials, and the true main verb of the sentence is encountered, as in *were wanted* in (4c). Researchers often assume that encountering the main verb completely disambiguates the string in favor of the RR interpretation, but the degree of disambiguation actually varies greatly with particular stimulus items. The major factor here is whether the main verb is itself ambiguous between a past tense and a past participle interpretation. A tense ambiguity at the main verb permits a second temporary MV/RR ambiguity in the stimulus sentence, as in *The witness examined by the lawyer turned out to be unreliable* (from Ferreira & Clifton, 1986). Here *turned* initially permits a RR modifying *lawyer* (as in *the lawyer turned in by the detective*), so that a definitive disambiguation is delayed. Stimuli with this second ambiguity are rare in most studies (including in Ferreira & Clifton), but they may be a source of additional noise in the reading data in some experiments. This additional ambiguity also serves to reinforce the point about the large number of constraints that can influence ambiguity resolution here.
- F. *The thematic role of the pre-verbal noun.* Relative clauses are more natural when the head noun is a theme of the action (*the flowers (that were) sent to the performer...*) than when the recipient of the action is the head noun (*the performer (who was) sent the flowers...*) (Keenan & Comrie, 1977). The rarity (or oddness) of modifying a Goal role decreases the likelihood of a reduced relative interpretation, and some studies that have found poor use of nonsyntactic constraints have tended to contain stimuli in which the goal role is relativized (e.g., Rayner et al., 1983), which strongly promotes the MV interpretation. This bias also interacts with the effect of post-verbal words described in point E, in that when the goal NP is modified as in *the performer sent the flowers*, the words after the ambiguous verb (*the flowers*) can strongly promote the MV interpretation (Tabor, Galantucci, & Richardson, 2004).
- G. *Constraints from the broader discourse that could promote either interpretations.* These constraints include whether the discourse makes it plausible to modify the first noun, which promotes an RR interpretation (Altmann & Steedman, 1988; Crain & Steedman, 1985; Ni, Crain, & Shankweiler, 1996; Sedivy, 2002), whether the tense of the verbs in prior discourse promotes interpretation of the ambiguous verb as a past tense or past participle (Trueswell & Tanenhaus, 1991), and factors affecting the likelihood of using a reduced vs. unreduced relative clause form in various discourse situations (McKoon & Ratcliff, 2003). The influence of these

discourse-level constraints may be modulated by more robust lexical-level constraints. For example, Filik, Paterson, and Liversedge (2005) found the extent to which attention-focusing words such as *only* influenced ambiguity resolution (Ni et al., 1996; Sedivy, 2002) varied with the range of alternative interpretations permitted by the ambiguous verb.

This and similar lists of potential constraints and their interactions (Townsend & Bever, 2001; Tanenhaus, Spivey-Knowlton, & Hanna, 2000) suggest why comprehension of the MV/RR ambiguity can sometimes succeed easily and other times fail miserably. Successful comprehension occurs when a variety of constraints strongly promote the ultimately correct reduced relative interpretation at an early point in the ambiguous string, and garden-pathing occurs when the evidence strongly favors the incorrect interpretation, as in *The horse raced past the barn fell*.

### 3.2. Computational Modeling of Constraint-Based Theory

The probabilistic constraints approach draws heavily on computational concepts such as constraint satisfaction. One way of exploring such concepts is by implementing simulation models. Modeling is a tool that can be used for different purposes; here we describe three ways computational models have been used in the development of the constraint-based approach.

1. As a tool for developing and illustrating novel theoretical concepts and analyses of the comprehension process. These models tend to be narrow in scope and tied to phenomena rather than the results of particular behavioral studies. Perhaps the most influential example is the work of Elman (1990), who developed the concept of a simple recurrent network in which the task is to predict the next word in a sentence given the current word and information about the prior context (in Elman's networks, the state of the hidden unit layer). Elman's models exhibited several interesting behaviors: they learned to predict words that were grammatical continuations of sentences; they formed representations of the grammatical categories of words; they encoded long-distance dependencies, not merely transition probabilities between adjacent words. The models introduced the important idea that sentence comprehension could be construed as following a trajectory in the state-space defined by a recurrent network. The models also helped to revive the idea that prediction might be an important component of sentence interpretation; recall that early results suggesting that words are not generally predictable from context (Gough et al., 1981) led to the view that contextual effects were weak and unhelpful. However, Elman's networks, and other connectionist approaches that emphasized distributed representations, suggested that comprehension processes might incorporate partial predictions where expectations are generated for certain semantic or syntactic properties of the upcoming input, even if exact words themselves are not predicted. There is now increasing evidence that human comprehension processes incorporate these predictive elements (e.g., Altmann, van Nice, Garnham, & Henstra, 1998; McRae, Hare, Elman, & Ferretti, 2005).



The St. John and McClelland (1990) model combined an Elman network with a distributed hidden layer of sentence meaning representations (the “sentence gestalt”). The model took simple sentences as input and was trained to answer queries about the thematic roles of noun phrases. The fact that the model was trained to represent meaning rather than predict upcoming words made it an interesting departure from other models of sentence processing, but it was very limited in scope, having a small number of words and thematic roles. The model also could not interpret multi-clause sentences, a serious limitation given the centrality of these constructions in syntactic and psycholinguistic research. Rohde (2002) adapted and expanded the St. John and McClelland architecture in a much larger model. His model replicated several key results in human sentence interpretation, but also behaved in some ways that differed from humans. Further research would be needed to determine whether these limitations could be overcome within this architecture.

A final example is Allen and Seidenberg’s (1999) model in which both comprehension and production were simulated within a single network. They used the model to illustrate a theory of how people make grammaticality judgments, and how, paradoxically, this ability could be maintained in aphasia (Linebarger, Schwartz, & Saffran, 1983). The model also illustrated why a sentence such as “Colorless green ideas sleep furiously” is judged grammatical even though the transition probabilities between words are low.

None of these models are “complete” in any sense, and the limitations on their scope allow the possibility that the results are not general. Nonetheless, these models are important as vehicles for introducing novel mechanisms, approaches, and analyses.

2. As a procedure for discovering the statistical regularities implicit in a large corpus of utterances, as discussed above. This application of modeling has been used mainly in studies of language acquisition; studies by Mintz (2003), Redington, Chater, and Finch (1998) and others show how representations of grammatical categories can be derived from distributional information. Cassidy, Kelly, and Shari (1999) used a simple feedforward network as a procedure for discovering phonological correlates of proper names. Haskell, MacDonald, and Seidenberg (2003) used a network to discover phonological properties associated with adjectival modifiers in English. Mirković, MacDonald, and Seidenberg (2005) developed a model that learned much of the complex inflectional system for nouns in Serbian, and showed that gender was cued by correlations between phonology and semantics.
3. As a way of accounting for behavioral data. This usage is akin to models of word reading that simulate the results of behavioral experiments (e.g., Harm & Seidenberg, 2004). Some simple recurrent networks (SRNs) have been used for this purpose. These SRNs are typically subject to additional analyses that allow their behavior to be linked to measures of human performance (Christiansen & Chater, 1999; MacDonald & Christiansen, 2002; Tabor, Juliano, & Tanenhaus, 1997; Tabor & Tanenhaus, 1999). For example, MacDonald and Christiansen computed an error

measure, called Grammatical Prediction Error, that provided a good fit with reading times in studies of relative clause comprehension. Tabor, Juliano, & Tanenhaus, (1997) and Tabor and Tanenhaus (1999) related model performance to reading times using a different approach. They coupled an SRN and a dynamical processor that was designed to relate hidden unit representations to a sentence interpretation. The dynamical processor's ability to settle on an interpretation varies with past experience, and processor time can be related to reading times in behavioral studies.

Another class of models directly addresses the process of constraint-weighting during comprehension and the linkage between these processes and behavioral data (Elman et al., 2004; McRae et al., 1998; Spivey & Tanenhaus, 1998). These models are not simulations of sentence comprehension per se but rather test claims for activation of alternative interpretations at different points in a sentence. Moreover, the models have no learning component but are instead hand-tuned to the properties of particular stimulus sentences from one or more existing behavioral study, potentially yielding a transparent relationship between the activation levels in the model and the pattern of behavioral data. For example, McRae et al. modeled the interaction of six constraints in the MV/RR ambiguity. They developed stimulus materials for an empirical study and used a combination of corpus analyses and questionnaire studies to assess the degree of constraint promoting MV vs. RR interpretations for each of their stimulus items. The constraints were implemented in a simple localist model in which each constraint and each interpretation of the ambiguity are represented by single nodes in a network. The model simulates the timecourse of constraint interaction because constraints are combined at each word position, and the alternative interpretations receive activation as a function of the combined constraint strength at that point. The alternative interpretations compete with one another, so that activation of one alternative drives down activation of the other. McRae et al. used the model to test two alternative accounts of ambiguity resolution for this structure by varying the time at which different constraints were available to the model. In one model, all syntactic and nonsyntactic constraints were available as soon as the relevant words were encountered in the sentence; this model corresponded to the claim that ambiguity resolution is accomplished through the rapid integration of multiple probabilistic constraints. In the second model, a syntactic constraint favoring the MV interpretation was allowed to have an early effect, and non-syntactic constraints (verb tense frequencies and the plausibility of a noun being an agent or patient of a verb) were delayed for several words. This model was designed to simulate predictions of a two-stage model, in which the MV interpretation is adopted in the first stage and use of non-syntactic information is delayed until the second stage. McRae et al. found that the more interactive model was a better fit to the data than the one in which non-syntactic constraints were delayed. They argued that the addition of modeling provides a much stronger test of alternative accounts than empirical work alone, in that the modeling effort forces commitments to particular claims about constraint interaction and its timecourse.

This brief summary serves to show that a wealth of ideas about sentence comprehension and related aspects of language have been explored using implemented connectionist models.

#### 4. STATE OF THE SCIENCE: CONTROVERSIES, UNRESOLVED ISSUES, AND FUTURE DIRECTIONS

The approach we have described is ambitious and yet still in the early stages of development, and so there are many gaps that researchers are attempting to address. We therefore conclude this chapter by considering a series of questions.

##### 4.1. Statistics All the Way Down?

A number of researchers have taken issue with constraint-based approaches to language comprehension processes (e.g., Frazier, 1995, 1998; Townsend & Bever, 2001). We have discussed some of these concerns elsewhere in this chapter, and our focus in this section will be on empirical studies that are designed to provide evidence that important aspects of language comprehension have a nonstatistical basis, contradicting a basic tenet of the approach. For example, McKoon and Ratcliff (2003, 2005) have challenged the constraint-based account of ambiguity resolution in the MV/RR construction, and more broadly, suggested that difficulty with RR sentences such as *The horse raced past the barn fell*, lie not in their temporary ambiguity but in the incompatibility between the construction and the verb's meaning (specifically, that *race* is a verb with internally caused changes of state). They suggest that the RR sentences and their "unreduced" counterparts, such as *The horse that was raced past the barn fell*, have subtly different meanings and uses, such that internally caused change of state verbs can appear in unreduced but not reduced relative clauses. This approach, in which certain reduced relatives are nonsensical rather than merely ambiguous, to some degree indicts all ambiguity resolution approaches to this construction. McKoon and Ratcliff's claims have been forcefully countered by McRae, Hare, and Tanenhaus (2005) and Hare et al. (in press). They trace the difficulty in McKoon and Ratcliff's examples to the frequency of passive uses of the ambiguous verbs (see constraint B in the list above), disentangle this property from meaning components, and provide additional evidence for a constraint-based account of this ambiguity.

Perhaps the most direct empirical challenge to the constraint-based accounts comes from work by Pickering, Traxler, Van Gompel, and colleagues (e.g., Traxler, Pickering, & Clifton, 1998; Van Gompel, Pickering, & Traxler, 2001; Van Gompel, Pickering, Pearson, & Liversedge, 2005), who have argued that constraint-based accounts make incorrect predictions about reading times for certain ambiguities. Specifically, they observe that constraint-based accounts predict that comprehension times should be longer for ambiguous sentences compared to unambiguous ones, owing to the fact that ambiguous sentences engender competition between alternative interpretations. In a series of studies using several different ambiguous constructions, these authors have found that globally ambiguous sentences are read more quickly, not more slowly, than unambiguous sentences. They suggest that these data argue against a constraint-based account and instead support a two-stage model in which multiple sources of information may affect a sentence's initial interpretation. However, Green and Mitchell (2006) found that the McRae et al's., (1998) computational model generally did not enter into an extended (period of competition for globally ambiguous sentences, and thus there is no prediction

for longer reading times for these items relative to disambiguated ones in these models. Green and Mitchell's simulations uncover behavior in the model that runs contrary to the assumptions that Traxler et al. (1998) and Van Gompel et al. (2001, 2005) made about model performance, and their simulation results emphasize the pitfalls of relying on intuition for how an implemented model will behave. More generally, these behavioral results and simulations will serve to push alternative accounts to be more precise, especially about predictions for sentences (like globally ambiguous ones) for which people find it difficult to compute an interpretation online. Fodor (1982) and Kurtzman and MacDonald (1993) discussed the possibility that certain global quantifier scope ambiguities may never be fully resolved by comprehenders, and perhaps global syntactic ambiguities also may not always receive a definitive resolution (see also Ferreira, Ferraro, & Bailey's, 2002, account of "good-enough" sentence interpretation and brief remarks about strategic effects in reading below).

#### 4.2. Which Statistics?

Languages exhibit many properties that can be counted; some, such as how often verbs follow nouns, seem more relevant than others, such as the frequency distribution for words in the third position in sentences. In a fully specified theory of language comprehension, it would be clear which statistical regularities people encode and use in processing, and why. Clearly, we do not have anything like that kind of theory in hand; we have some evidence about the use of particular statistics that supports the general theoretical framework. As discussed above, in principle it should not be necessary to specify "the statistics that are relevant to language" a priori because that information should fall out of an automatic procedure: a neural network (or similar formalism) that processes language, subject to constraints imposed by the architecture, representations, and input. This procedure also approximates the experience of the child, for whom the relevant statistics are learned rather than pre-specified. Above, we summarized modeling research that represents important progress toward this ideal, and some models (in limited domains) have generated testable predictions. However, there are practical limits on building large-scale models, and analyzing the behavior of a complex dynamical system becomes difficult. These conditions make it difficult to use a computational model as an independent, hands-off way of determining which statistics are relevant. For a skeptic, the absence of a complete model creates the possibility that the statistical approach is vacuous because it can explain any result. No matter how an experiment turns out, the argument goes, a researcher can find a statistic or combination of statistics that can account for the pattern of results. The approach is therefore not merely unfalsifiable (i.e., able to fit all patterns that do occur); it can also fit patterns of data that never occur.

While it is important to acknowledge the limits of current knowledge, these concerns are not realistic. First, the methodology used in this research does not involve collecting behavioral data and then finding statistics to fit the results. Rather, researchers test hypotheses developed from several sources: linguistic theory; existing empirical findings; close analyses of examples (e.g., data mining a corpus); and other types of theorizing (e.g., about why languages exhibit particular kinds of statistical regularities; see below). Second, the strategy

of tailoring one's statistical analysis to fit a particular set of data would be self-defeating, because it results in overfitting: the results will not generalize to other data of the same sort (see Seidenberg & Plaut, 2006, for a discussion of this issue with respect to models of word reading). Third, the grain of the behavioral data is such that the number of constraints that actually account for detectable variance is relatively small (although not easily accommodated by factorial designs). The theory states that performance is determined by the aggregate effects of all of one's experiences with language; out of this highly variable set of data, some very strong regularities arise, and these end up being the ones that can account for observable behavior.<sup>2</sup> Finally, the idea that there will always be a statistic or combination of statistics to fit any data pattern greatly misjudges the extent to which language structure is constrained. There aren't unlimited degrees of freedom in accounting for the data because there aren't unlimited degrees of freedom in how a language can be structured. It is true that the number of language statistics that can be calculated is nearly infinite, but most of them are meaningless. The fact that we can calculate language statistics that do not account for data is not a problem if there are other bases for determining which statistics *are* relevant.

### 4.3. Different Models for Different Phenomena?

Above we suggested that implemented models have been important in developing and testing the probabilistic constraints approach. Such models are as yet limited in scope, and many important linguistic phenomena have yet to be addressed. A deeper concern is that every model is different, i.e., different models have been applied to different phenomena. Where is the integrative model that would subsume the broad range of phenomena that have as yet been investigated using many different models?

We are sympathetic to the concerns that are raised by using different models for different phenomena. It would be a problem if the principles that explain one phenomenon, studied using one architecture, are incompatible with the principles that explain some other phenomenon, studied using another architecture. Perhaps such models can only succeed when narrowly focused, as might be seen if a more general model were attempted. There probably is no simple way to address this issue, or a simple, preferable alternative. In the early stages of developing this approach it has been necessary to merely demonstrate that it is sufficient to account for interesting phenomena, given the general climate of skepticism about statistical methods and connectionist models in the study of language. Further progress would be achieved if, as additional models are developed, researchers were able to identify which general computational properties

---

<sup>2</sup> Perhaps a good analogy is the analysis of evoked potentials (Kutas, van Petten, & Kluender, this volume). This methodology involves gathering many samples of apparently noisy data: every brain wave is different from every other one. Many different aspects of these waves could be measured and counted, and there is no independent theory of how the waves are generated to indicate which elements are important prior to looking. Averaging across many data samples, however, certain regularities emerge (i.e., systematic displacements of the waveform such as P300, N400 and others). Language may exhibit a greater number of regularities, and we also want a better theory of their sources, but the similarities are noteworthy.

are crucial. These principles are more important than the characteristics that differentiate implemented models. This approach has achieved some success in the domain of single word reading. Researchers have identified a small set of critical computational principles, which have been explored in a succession of models. Each model is slightly different than the others (because of advances in understanding network properties or because they focus on different phenomena), but they are governed by the same principles. It is the set of principles and how they apply to a set of phenomena that constitute the explanatory theory, not the properties of individual models.

Achieving this deeper level of understanding language comprehension requires much more research: more models addressing a broader range of phenomena; comparing different architectures with respect to the same phenomena; analyzing models to identify the properties that are critical to achieving human-like performance. This is an ambitious agenda and it is not clear whether there are enough researchers with the sufficient technical skills and interest in the approach to achieve these goals. Moreover, it is not clear whether it is either feasible or desirable to develop a genuinely integrative model of broad scope. As Seidenberg and Plaut (2006) observed,

The concept of a complete, integrative model is a non sequitur, given the nature of the modeling methodology, particularly the need to limit the scope of a model in order (a) to gain interpretable insights from it and (b) to complete a modeling project before the modeler loses interest or dies. The goal of the enterprise, as in the rest of science, is the development of a general theory that *abstracts away from* details of the phenomena to reveal general, fundamental principles (Putnam, 1973). Each model serves to explore a part of this theory in progress.

We think it's important to keep in mind that models are tools, not the goal of the theoretical enterprise. The limitations of individual models are tolerable if they yield insights about puzzling phenomena, generate testable hypotheses, and promote theoretical development.

#### 4.4. Where Do Language Statistics Come From?

Within the constraint satisfaction account, a fine grained characterization of the statistical regularities constraining the interpretation of ambiguities is important to capturing behavioral data. As much of the above discussion suggests, the complexity of the system makes this accounting a nontrivial enterprise. Some insight into the constraints, and a broader account of language performance, may emerge from addressing the question of the origin of the statistical regularities in language. That is, why do languages exhibit certain statistical properties and not others? At least three forces may modulate the statistics of a language. First, some statistical regularities may be shaped by conceptual structures (McRae, Ferretti, & Amyote, 1997), so that aspects of our (nonlinguistic or pre-linguistic) thinking constrains the form of utterances. Second, statistics may be shaped by language producers' sensitivity to limits on our comprehension abilities, so that producers tailor their utterances to those that are more easily understood, in the process creating statistical

regularities in the language. The extent to which speakers are sensitive to listener needs is not fully resolved, but certainly there are at least some clear examples of speakers tailoring speech to their audience, such as the broad differences in the character of child- and adult-directed speech. Finally, some statistics may emerge from the production process itself. MacDonald (1999) and Gennari and MacDonald (2004) argued for this approach, termed the Production–Distribution–Comprehension (PDC) account, which suggests that certain statistical patterns emerge from language producers’ needs to maximize fluency during production. For example, speakers appear to adopt syntactic structures for their utterances at least in part to yield an utterance in which more highly “accessible” words are uttered early, where accessibility here refers to a variety of conceptual, lexical, and perhaps articulatory properties that affect the ease of articulating a particular word or phrase (Bock, 1987). MacDonald (1999), Gennari and MacDonald (2004) and Race and MacDonald (2003) have applied this logic to several different comprehension issues and have argued that comprehenders’ preferences to interpret ambiguities in favor of one vs. another alternative structure can be linked to the relative frequency of those alternatives in the language, owing to speakers’ and writers’ syntactic choices during the production process. These choices in turn stem from biases inherent in the production system, such as to place shorter sentence elements (words or phrases) before longer ones, or to place pauses or small optional function words before sections of high-production complexity. If this view is on the right track, then an increased understanding of constraint satisfaction in sentence comprehension will emerge from a better grasp of how the production process promotes certain production choices (word orders, word-structure co-occurrences, structure-discourse co-occurrences, etc.) and discourages others.

#### 4.5. Where To Go Next?

In presenting this approach, we have already mentioned several important directions for future research. We will close by mentioning three more. First, as detailed in the chapter by Trueswell and Tanenhaus in this volume, researchers are beginning to expand the range of constraints that comprehenders consider by investigating the extent to which comprehenders integrate the visual scene and other aspects of conversational interaction. This work allows an investigation of comprehension of speech, in contrast to the vast majority of studies discussed in this chapter, which have investigated written language. Second, returning to the written language realm, something that would benefit all theoretical perspectives is to increase our understanding of reading data and its relationship to computational accounts of comprehension processes. Researchers from many theoretical perspectives agree that the theorizing and the data are not well matched, in that certain reading patterns are compatible with radically different interpretations of ambiguity resolution processes (e.g., Lewis, 2000; Tanenhaus, 2004; Van Gompel et al., 2001). This situation may be traced to some combination of imprecision in theoretical claims, inability of reading or other dependent measures to resolve fine-grained predictions about timecourse, insufficient consideration of the possibility that reading and other dependent measures may reflect comprehenders’ strategies so that the data may not be a “pure” reflection of the ambiguity resolution processes. That is, we all know that a novel, a newspaper, and a chemistry textbook elicit different reading behaviors, yet there is very

little appreciation among researchers of the degree to which reading strategies might vary with the nature of filler items or comprehension questions in experiments and the extent to which these strategic components could be affecting reading patterns that are attributed to “automatic” sentence processing operations.

Third, the constraint-based approach affords the opportunity to investigate the relationship between acquisition and skilled performance. The focus in adult comprehension has been on timecourse, specifically the speed with which comprehenders can bring constraints to bear on linguistic input, and there has been relatively little discussion of the learning mechanisms by which comprehenders come to possess the relevant constraints. The claim that the learning is inherently statistical invites research into the extent to which there is continuity between acquisition and adult performance and the extent to which a statistical learning account will prove adequate to explain the child’s rapid mastery of language. These questions link to an enormous and ongoing research enterprise in child language acquisition, one with its own controversies and struggles to match theory and data. It is therefore an exciting possibility for theoretical development that the studies of the adult state and the acquisition process in the child might be mutually informative and constraining.

## REFERENCES

- Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist nets. In: B. MacWhinney (Ed.), *The Emergence of Language*. Hillsdale, NJ: Erlbaum.
- Altmann, G. T. M., van Nice, K., Garnham, A., & Henstra, J. A. (1998). Late closure in context. *Journal of Memory and Language*, 38, 459–484.
- Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In: B. MacWhinney, & E. Bates (Eds), *The crosslinguistic study of sentence processing*. Cambridge, UK: Cambridge University Press.
- Bever, T. G. (1970). The cognitive basis for linguistic structure. In: J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–360). New York: Wiley.
- Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structure. *Journal of Memory and Language*, 26, 119–137.
- Bresnan, J. (1982). The passive in lexical theory. In: J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 3–86). Cambridge, MA: MIT Press.
- Cassidy, K. W., Kelly, M. H., & Sharoni, L. J. (1999). Inferring gender from name phonology. *J. Exp. Psychology: General*, 128, 362–381.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.



- Chomsky, N. (1981). Knowledge of language: Its elements and origins. *Philosophical Transactions of the Royal Society of London*, 295, 223–234.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language. An introduction to psycholinguistics*, New York: Harcourt Brace Jovanovich.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In: D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural Language Parsing*. Cambridge: Cambridge University Press.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Elman, J. L. (1990). Representation and structure in connectionist models. In: G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 345–382). Cambridge, MA: MIT Press.
- Elman, J., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In: M. Tomasello, & D. Slobin (Eds), *Beyond nature-nurture: Essays in honor of Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, 62, 369–377.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Filik, R., Paterson, K. B., & Liversedge, S. P. (2005). Parsing with focus particles in context: Eye movements during the processing of relative clause ambiguities. *Journal of Memory and Language*, 53, 473–495.
- Fischler, I. S., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18, 1–20.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9, 427–473.
- Fodor, J. D. (1982). The mental representation of quantifiers. In: S. Peters, & E. Saarinen (Eds), *Processes, beliefs, and questions* (pp. 129–164). Dordrecht: Reidel.

- Ford, M., Bresnan, J., & Kaplan R. M. (1982). A competence-based theory of syntactic closure. In: J. Bresnan (Ed.), *The mental representation of grammatical relations*, Cambridge, Massachusetts, MIT Press, pp. 727–796.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In: W. Cooper, & E. C. T. Walker (Eds), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Mahwah, NJ: Lawrence Erlbaum Associates.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. thesis, University of Connecticut.
- Frazier, L. (1987). Sentence processing: A tutorial review. In: M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559–586). Mahwah, NJ: Lawrence Erlbaum Associates.
- Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, 24, 437–468.
- Frazier, L. (1998). Getting there (slowly). *Journal of Psycholinguistic Research*, 27, 123–146.
- Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Gagne, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 71–87.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Gennari, S., & MacDonald, M. C. (2004). Relating comprehension and production of relative clauses. Paper presented at the 17th Annual CUNY conference on human sentence processing, College Park, MD.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gough, P. B., Alford, J. A., & Holley-Wilcox, P. (1981). Words and contexts. In: O. J. L. Tzeng, & H. Singer (Eds), *Perception of print: Reading research in experimental psychology* (pp. 85–102). Mahwah, NJ: Lawrence Erlbaum Associates.
- Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution, *Journal of Memory and Language*, 1–17.
- Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization references. *Journal of Memory and Language*, 48, 281–303.
- Hare, M., Tanenhaus, M. K., & McRae, K. (2006). Understanding and producing the reduced relative construction: Evidence from ratings, editing and corpora. Manuscript submitted for publication.

Harm, M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.

Haskell, T., MacDonald, M. C., & Seidenberg, M. S. (2003). Language learning and innateness: Some implications of *compounds research*. *Cognitive Psychology*, *47*, 119–163.

Hayes-Roth, F., & Jacobstein, N. (1994). The state of knowledge-based systems. *Communications of the ACM*, *37*, 26–39.

Joshi, A. K. (1985). How much context-sensitivity is required to provide reasonable structural descriptions: Tree adjoining grammars. In: D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural Language Parsing*. Cambridge, NY: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, *32*, 474–516.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar: *Linguistic Inquiry*, *8*, 63–99.

Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignment. *Psychological Review*, *99*, 349–364.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Kurtzman, H., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, *48*, 243–279.

Kutas, M., Van Petten, C., & Kluender, R. (this volume). Psycholinguistics electrified II: 1995–2005. In: M. Traxler, & M. A. Gernsbacher (Eds), *Handbook of psycholinguistics* (2nd ed.). New York: Elsevier.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: The University of Chicago Press.

Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, *29*, 241–248.

Linebarger, M., Schwartz, M., & Saffran, E. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361–393.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.

- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In: B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *89*, 483–506.
- Mackworth, A. K. (1977). Consistency in networks of relations. *Artificial Intelligence*, *8*, 99–118.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.
- Marcus, M. P., Santorini, B., & Marcinkiewicz, M. (1993). Building a large annotated corpus of english: The Penn Treebank. *Computational linguistics*, *19*, 313–330.
- Margolis E., & Laurence. S. (Eds). (1999). *Concepts: Core readings*. Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226–228.
- McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In: D. E. Rumelhart, & J. L. McClelland (Eds), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I). Cambridge, MA: MIT Press.
- McKoon, G., & Ratcliff, R. (2003). Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review*, *110*, 490–525.
- McKoon, G., & Ratcliff, R. (2005). “Meaning through syntax” in sentence production and comprehension: Reply to McRae et al. (2005). *Psychological Review*, *112*, 1032–1039.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.
- McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*, 1174–1184.
- McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176.
- McRae, K., Hare, M., & Tanenhaus, M. K. (2005). Meaning Through Syntax is insufficient to explain comprehension of sentences with reduced relative clauses: Comment on McKoon and Ratcliff (2003). *Psychological Review*, *112*, 1022–1031.

- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Merrill, E. C., Sperber, R. D., & McCauley, C. (1981). Differences in semantic encoding as a function of reading comprehension skill. *Memory & Cognition*, 9, 618–624.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91–117.
- Mirković, J., MacDonald, M. C., & Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes*, 20, 139–168.
- Newmeyer, F. J. (1980). *Linguistic theory in America: The first quarter-century of transformational generative grammar*. New York: Academic Press.
- Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes*, 11, 283–334.
- Pearlmutter, N. J., & MacDonald, M. C. (1992). Plausibility and syntactic ambiguity resolution. *Proceedings of the 14th Annual Conference of the Cognitive Society* (pp. 498–503). Mahwah, NJ: Lawrence Erlbaum Associates.
- Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604–607.
- Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133, 573–583.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Putnam, H. (1973). Reductionism and the nature of psychology. *Cognition*, 2, 131–146.
- Race, D. S., & MacDonald, M. C. (2003). The use of “that” in the production and comprehension of object relative clauses. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Boston, MA: Erlbaum.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning & Verbal Behavior*, 22, 358–374.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425–469.
- Rohde, D. L. T. (2002). A Connectionist Model of Sentence Comprehension and Production. Unpublished Ph.D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

- Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In: S. Stevenson, & P. Merlo (Eds), *The lexical basis of sentence processing: Formal, computational, and experimental issues* (pp. 325–346). Amsterdam: John Benjamins.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In: J. L. McClelland, D. E. Rumelhart, and the PDP research group (Eds). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II*. Cambridge, MA: MIT Press.
- Saffran, J. R., & Sahni, S. D. (in press). Learning the sounds of language. In: M. Joanisse, M. Spivey, & K. McRae (Eds), *Cambridge handbook of psycholinguistics*, Cambridge, UK: Cambridge University Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48, 416–434.
- Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Memory and Language*, 46(2), 341–370.
- Seidenberg, M. S., & Gonnerman, L. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4, 353–361.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, 298, 553–554.
- Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. In: S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Hove, UK: Psychology Press.
- Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning & Verbal Behavior*, 20, 120–136.
- Simpson, G. B., & Kreuger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, 30, 627–643.
- Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition*, 26(5), 965–978.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1521–1543.

- Spivey-Knowlton, M., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.
- St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217–257.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.
- Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, *23*, 491–515.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory & Language*, *50*, 355–370.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*, 211–271.
- Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In: C. Umiltà, & M. Moscovitch (Eds), *Attention and performance 15: Conscious and nonconscious information processing* (pp. 589–615). Cambridge, MA: MIT Press.
- Tanenhaus, M. K. (2004). Sentence processing: Past, present, and future. In: M. Carreiras, & C. Clifton, Jr. (Eds). *On-line measures in sentence processing: Eye movements, ERPS and beyond*. Sussex, UK: Psychology Press.
- Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*, 427–440.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. E. (2000). Modeling thematic and discourse context effects with a multiple constraints approach: Implications for the architecture of the language comprehension system. In: M. W. Crocker, M. Pickering, & C. J. Clifton (Eds), *Architectures and mechanisms for language processing* (pp. 90–118). Cambridge, UK: Cambridge University Press.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In: J. Miller, & P. Eimas (Eds), *Handbook of Perception and Cognition, Vol. 11: Speech, Language and Communication* (pp. 217–262). New York: Academic Press.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558–592.
- Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory & Language*, *35*, 566–585.

Trueswell, J. C., & Tanenhaus, M. K. (1991). Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes*, 6, 303–338.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory & Language*, 33, 285–318.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, 52, 284–307.

Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45, 225–258.



This page intentionally left blank

## Chapter 16

### Eye-Movement Control in Reading

*Keith Rayner and Alexander Pollatsek*

#### 1. INTRODUCTION

Psychologists interested in language processing have increasingly turned to the use of eye movement data to examine moment-to-moment processing (Liversedge & Findlay, 2000; Rayner, 1998; Rayner & Liversedge, 2004; Starr & Rayner, 2001).<sup>1</sup> This, in our view, is not surprising because eye movements represent one of the best ways to study language comprehension processes. In comparison to other available techniques (see Haberlandt, 1994; Rayner & Sereno, 1994a), eye-movement data provide a relatively natural, on-line method for investigating critical psycholinguistic issues. Importantly, eye movements during reading are not part of an artificially induced task – they are part of the normal reading process. In addition, monitoring readers' eye movements does not perturb their normal reading rate. Although it has been the case in the past that studies utilizing eye-movement data have typically required the position of the head to be fixed (by the use of a bitebar or chin rest), this is not necessarily the case at the present time (although the most accurate eye-tracking data invariably result from having a participant's head fixed). This constraint is sometimes viewed as introducing an unnatural component to reading. However, our view (see also Rayner & Sereno, 1994a) is that participants in eye-movement experiments read quite normally. This is supported by data reported by Tinker (1939) that indicate that reading rate and comprehension do not differ when readers read text in a laboratory situation with their eye movements recorded and when they read in normal conditions (i.e., without a fixed head).

We do not think that the only way to study reading is by examining the eye-movement record. In building a theory of language comprehension, it is necessary to obtain converging evidence from various sources. Thus, other tasks, such as word-by-word self-paced reading, rapid serial visual presentation (RSVP) of sentences, event-related

---

<sup>1</sup>In this context, it is interesting to note that Table 2 in the Rayner (1998) review article lists 113 articles that were reported between 1978 and 1998 in which eye-movement data were used to examine some aspect of language processing. As of mid-2005, 88 articles have appeared in 7.5 years since the 1998 review article was published.

potentials (ERP), and so on, have their place and many of them undoubtedly probe the nature of readers' mental representations and provide useful information about on-line processing of language. However, we think that among the current methodologies, the eye-movement technique does the best job of revealing moment-to-moment processes in reading.

A second general point that we wish to make is that psycholinguists interested in language processing often use eye-movement data without understanding some of the basic issues underlying the technique (see Rayner, 1998; Rayner & Liversedge, 2004). That is, it is now relatively easy to obtain eye-movement data (as many companies that market eye-trackers also provide software for data analysis), and many researchers seem to be primarily interested in seeing if their manipulation has an effect on the eye-movement record. However, we would argue that it is quite important for such researchers to know something about eye movements per se since properties of the oculomotor system could well be influencing the results obtained. In this chapter, we will provide an overview of what is known about eye movements in reading, and the relationship between eye movements and cognitive/linguistic processing. We will focus primarily on eye movements and lexical processing, though we will also touch on some research dealing with parsing and discourse processing. In large part, we will argue that the movements of the eyes through the text is primarily driven by lexical processing, with higher-order processing intervening when something does not compute well. We will also describe some recent models of eye-movement control in reading, though we will focus primarily on our own model (the E-Z Reader model).

## 2. BASIC FACTS ABOUT EYE MOVEMENTS IN READING

During reading, we typically have the impression that our eyes are gliding smoothly across the page. However, this is an incorrect impression; instead the eyes make a series of rapid movements (called *saccades*) separated by periods of time when the eyes are relatively still (called *fixations*). It is only during the fixations that new visual information is encoded from the text because vision is functionally suppressed during the saccades. Fixations typically last about 200–250 ms, although individual fixations in reading can be as short as 50–100 ms and as long as 500 ms. Distributions of fixation durations look like normal distributions (with the mean around 200–250 ms) that are skewed to the right. Typically, saccades last roughly 20–40 ms; the duration of the saccade depends almost exclusively on the size of the saccade. Saccades moving from the end of one line to the next (called *return sweeps*) typically last longer than the movements that progress along a line, and they also tend to undershoot the intended target. Thus, a return sweep will often be followed by a corrective movement to the left (when reading English). Nevertheless, the first fixation on the line is typically 5–7 letter spaces from the beginning letter on the line; likewise, the last fixation on a line is also typically 5–7 letters from the last letter in the line. Thus, only about 80% of the text typically falls between the extreme fixations.

While the two eyes begin moving at about the same time, it turns out that the eyes do not land in exactly the same place in a word. Liversedge, White, Findlay, and Rayner

(2006; see also Heller & Radach, 1999) recently demonstrated that in as much as 50% of the cases, the two eyes are not aligned on the same character. Nevertheless, despite this divergence (which does vary as a function of line location), the effect of linguistic processing is still apparent in the eye-movement record (Juhasz, Liversedge, White, & Rayner, 2006).

The saccades per se serve the function of bringing a given region of text into foveal vision for detailed analysis. A given line of text falling on the reader's retina can be divided into three different regions with respect to the reader's point of fixation: foveal, parafoveal, and peripheral. The foveal region corresponds to the central 2° of visual angle around the fixation point (for text at a normal viewing distance, 1° of visual angle is equivalent to roughly 3–4 letters); the fovea is specialized for processing detail. The parafoveal region of a line extends from the foveal region out to about 5° of visual angle to each side of fixation. Readers are able to acquire some useful letter information from this region (see Section 3.1). The peripheral region includes everything on the line beyond the parafoveal region. Beyond the fovea, acuity drops off markedly and words that are not located in the fovea are difficult to identify. Indeed, reading on the basis of non-foveal information is difficult if there is parafoveal information and impossible if only peripheral information is available (Rayner & Bertera, 1979). Although readers are aware of the ends of lines and other gross aspects of the text, information in peripheral vision tends to be of little use in reading.

The average saccade size in reading is about 7–9 letter spaces.<sup>2</sup> However, just as with fixation durations, there is quite a bit of variability in saccade size: some saccades are as short as one letter and some can be over 20 letter spaces (though the longest saccades typically follow a regression and take the eyes to a point ahead of the point at which the regression was launched). The variability that exists in both fixation duration and saccade size is related to processing activities: when text is difficult, readers make longer fixations and shorter saccades. Furthermore, when text is difficult, readers move their eyes backwards in the text (these backwards movements are called *regressions*). Regressions occur about 10% of the time in skilled readers. Many regressions are short (back to the word just to the left of the current fixation) and probably reflect either oculomotor irregularities or else word recognition difficulties; other regressions are longer, and probably reflect comprehension difficulties. Interestingly, there also appears to be an inhibition of return component to regressions as fixations preceding saccades to previously fixated words are longer than saccades to skipped words (Rayner, Juhasz, Ashby, & Clifton, 2003).

Much of the research on eye movements and reading has focused on fixation time on a word (or on reading time for larger segments of text). However, both the probability of a regression from a word and the probability of skipping a word are often examined as well.

---

<sup>2</sup>In reading, the appropriate metric for how far readers move their eyes is character spaces and not visual angle (see Morrison & Rayner, 1981).

Part of the variability in saccade length discussed above is related to word skipping (generally, skipped words are processed even though they are not fixated). Skipping is not random, as short words (three or fewer letters) are skipped fairly frequently, six-letter words are usually fixated, and words that are eight letters or longer are rarely skipped (Brysbart & Vitu, 1998; Rayner & McConkie, 1976). Other factors influence skipping as content words are typically fixated about 83% of the time, whereas function words (which obviously tend to be shorter) are fixated only 19–38% of the time (Carpenter & Just, 1983; Rayner & Duffy, 1988). (We will return to factors that influence skipping below.)

While eye-movement data are very informative with respect to lexical processing and understanding reading, they are not perfect reflections of the mental activities associated with comprehension. There is a purely motoric component of eye movements, and low-level visual and oculomotor factors can also influence fixation time and saccade length. Nevertheless, very useful information can still be obtained from the eye-movement record.

### **3. CRITICAL ISSUES IN USING EYE-MOVEMENT DATA TO STUDY READING**

If one is interested in using eye-movement data to study some aspect of language comprehension during reading, there are a number of issues inherent in using eye movements that need to be addressed. We will briefly discuss the following issues: the perceptual span, integration of information across saccades, control of eye movements, and measures of processing time.

#### **3.1. The Perceptual Span**

How much information do readers process on each fixation? What is the size of the effective field of view? These questions are clearly related to issues of acuity that we discussed above. Clearly, for readers of English, most of the useful information is confined to the foveal and parafoveal regions. Indeed, studies by McConkie and Rayner (1975), Rayner (1975), and Rayner and Bertera (1979) using gaze-contingent display change paradigms have confirmed this. With these techniques, either the global amount of information available to the reader can be precisely controlled (as in the moving window paradigm, McConkie & Rayner, 1975; Rayner & Bertera, 1979), or the amount and type of information in a specific region can be precisely controlled, as when a preview word is changed to a target word mid-saccade (as in the boundary paradigm, Rayner, 1975). These experiments, and many that followed (see Rayner, 1978, 1998 for reviews) have demonstrated that for readers of English (and other alphabetic writing systems), the span of perception (or region of effective vision) extends from 3–4 character spaces to the left of fixation (or the beginning of the currently fixated word) to 14–15 character spaces to the right of fixation. Furthermore, readers do not acquire useful information from lines below the one they are fixating (Pollatsek, Raney, LaGasse, & Rayner, 1993). Given that information from the rightmost part of the perceptual span is typically rather gross information, the region of word identification on the current fixation is even more restricted

(to typically no more than 7–8 letters to the right of fixation, although the exact size of the region varies as a function of the text being read).

The fact that the word identification span is restricted turns out to be very advantageous for researchers interested in using eye movements to study on-line language processing. If readers could process words from a wide range around their point of fixation, it would be difficult to know which word was being processed at any point in time and eye movements would not be particularly useful for studying language processing. It would be ideal for studying language processing if readers only processed the word they were fixating (making it easy to tie down what is being processed at any point in time). The reality is not quite that good, but much of the processing when a word is fixated is on the fixated word, especially the processing that occurs before the decision to move on to the next word in the text. We will return to this issue more fully in Section 6 when we discuss models of eye-movement control in more detail.

### 3.2. Integration of Information across Saccades

Readers do not obtain a chunk of information on one fixation and then a different chunk of information on the next fixation. Rather, there is overlap of information from fixation to fixation. That is, they usually obtain useful information from the word to the right of the currently fixated word (and occasionally from the word two to the right) and this information is used on the following fixation. So, if a reader is looking at word  $n$ , they identify the meaning of that word, but also obtain some *preview* information from word  $n+1$  that helps them identify it when they fixate it. In general, the size of this preview benefit is 30–40 ms (Hyönä, Bertram, & Pollatsek, 2004).

Research is still on-going to determine what levels of processing are responsible for this benefit from a preview of a word; however, much has been learned and we will paint the general results in rather broad strokes (see Rayner, 1998 for more precise details). First, it is clear that visual information is not integrated across fixations; if the case of the letters changes from fixation to fixation, readers do not notice the change and it has little effect on their reading (McConkie & Zola, 1979; Rayner, McConkie, & Zola, 1980). Second, semantic information is not the basis of the preview effect. Thus, *song* as a preview in the parafovea for *tune* does not result in preview benefit in reading (Altarriba, Kambe, Pollatsek, & Rayner, 2001; Rayner, Balota, & Pollatsek, 1986). In the Rayner et al. (1986) study, the same words that were ineffective as parafoveal primes produced a robust priming effect when the prime (*song*) and target (*tune*) were both presented foveally. Third, morphological information is also not a good candidate for facilitating preview benefit (Inhoff, 1989; Kambe, 2004; Lima, 1987). Fourth, letter information is important: information about the beginning letters of word  $n+1$  is critically important (Rayner et al., 1980; Rayner, Well, Pollatsek, & Bertera, 1982; Inhoff, Pollatsek, Posner, & Rayner, 1987). In this latter context, it is interesting that Miller, Juhasz, and Rayner (2006) recently reported that words with early orthographic uniqueness points do not yield stronger parafoveal preview benefits than words with late orthographic uniqueness

points (see also Lima & Inhoff, 1985, for further evidence that is inconsistent with COHORT types of effects in preview benefit). Further, it should be noted that information about ending letters is also important, but not as important as beginning letters (Briehl & Inhoff, 1995). More recently, Johnson, Perea, and Rayner (2006; see also Johnson, 2006) have demonstrated the importance of letter information by showing that transposed letters (*jugde* for *judge*) provide more preview benefit than letter substitutions (*jvbg* for *judge*), and almost as much benefit as identical previews. Fifth, phonological codes are important in integrating information across fixations (Ashby & Rayner, 2004; Chace, Rayner, & Well, 2005; Henderson, Dixon, Petersen, Twilley, & Ferreira, 1995; Miellet & Sparrow, 2004; Pollatsek, Lesch, Morris, & Rayner, 1992; Sparrow & Miellet, 2002). For example, in Pollatsek et al.'s study, a homophone of the target word (*beech*) presented as a preview in the parafovea (for *beach*) facilitated processing of the target word more than the preview of an orthographically related preview (*bench*). In summary, the basis for the parafoveal preview benefit effect appears to be some type of combination of abstract letter codes and phonological codes.

### 3.3. Eye-Movement Control

There are two components of eye-movement control: (1) what determines *where* to look next and (2) what determines *when* to move the eyes. We will discuss each in turn. But, we first want to make the point that both decisions are computed on-line on most fixations. The first unambiguous demonstration of this was provided by Rayner and Pollatsek (1981). In those experiments, the physical aspects of the text were varied randomly from fixation to fixation, and the behavior of the eyes mirrored what was seen on the current fixation. In the first experiment, the size of the window of normal text was randomly varied from fixation to fixation (so the size of the window might be 9 letters on fixation 1, 31 letters on fixation 2, 17 letters on fixation 3, and so on), and saccade length varied accordingly. In the second experiment, the foveal text was delayed after the onset of the fixation by a mask (with the time of the delay varying from fixation to fixation), and fixation durations varied accordingly (see also Morrison, 1984).

#### 3.3.1. *Where to move the eyes*

Low-level information (i.e., the spaces between words) is the primary determinant of where to look next (Morris, Rayner, & Pollatsek, 1990; Pollatsek & Rayner, 1982; Rayner, Fischer, & Pollatsek, 1998). When spacing information is absent, saccades are much shorter and readers are much more cautious in moving their eyes (Rayner et al., 1998). The length of the upcoming words is also important (O'Regan, 1979, 1980; Rayner, 1979). Although there is some variability in where the eyes land on a word, readers tend to make their first fixation on a word about halfway between the beginning and the middle of the word (McConkie, Kerr, Reddix, & Zola, 1988; Rayner, 1979; Rayner, Sereno, & Raney, 1996). A recent interesting, and seemingly counterintuitive, finding is that fixations tend to be longer when readers, fixations initially land near the middle of the word than when they land on the ends of words (Vitu, McConkie, Kerr, & O'Regan,

2001). Nuthmann, Engbert, and Kliegl (2005) have elegantly demonstrated that this *inverted optimal viewing position* effect is largely attributable to mislocalized fixations.

As indicated above, word skipping is heavily influenced by word length as shorter words are more likely to be skipped (Brysbaert & Vitu, 1998; Rayner, 1998). However, linguistic variables (particularly contextual constraint, but also word frequency to some extent) also have strong influences on where decisions, notably on whether a word is skipped (though there is little influence from contextual constraint on where in the word the eyes land, Rayner, Binder, Ashby, & Pollatsek, 2001). In particular, words that are highly predictable are much more likely to be skipped (Gautier, O'Regan, & LaGargasson, 2000; Rayner & Well, 1996). More frequent words also tend to be skipped more, although this effect is not as strong as that of contextual constraint (Rayner et al., 1996). (Note that both these effects hold even when the length of the word is controlled.) Our view is that words are largely skipped because they have been identified on the prior fixation and there is some evidence suggesting that fixations prior to skips are often inflated (Drieghe, Rayner, & Pollatsek, 2005; Kliegl & Engbert, 2005; Pollatsek, Rayner, & Balota, 1986; Pynte, Kennedy, & Ducrot, 2004). Thus, whereas low-level variables are largely determining where to fixate next, if the word to the right of fixation is identified on the current fixation, such identification will lead to a change in decision about which word to target next.

### 3.3.2. *When to move the eyes*

In Section 4, we will discuss a large number of variables (related to how easy or difficult a word is to process) that have been shown to influence fixation time on a word. In this section, we will limit ourselves to discussion of how quickly information gets into the processing system and its implications for when to move the eyes.

Rayner, Inhoff, Morrison, Slowiaczek, and Bertera (1981) first demonstrated that reading proceeds essentially normally if text is presented for at least 50 ms on each fixation before a masking pattern replaces the entire text (see Ishida & Ikeda, 1989). More recently, studies by Rayner, Liversedge, White, and Vergilino-Perez (2003), Liversedge et al. (2004), and Rayner, Liversedge, and White (2005) have demonstrated that if the text is available for 60 ms prior to either the fixated word disappearing or being masked, reading proceeds quite smoothly and normally. Of greater interest is that they also found that the frequency of the fixated word has just as strong an influence on how long the eyes remain in place when it disappears after 60 ms as when it does not disappear. This appears to be a strong evidence that the cognitive processes associated with understanding the fixated word is the primary force driving the eyes through the text.

## 3.4. Measures of Processing Time

We will first focus on the measures most commonly used to investigate the processing time associated with a given target word. These measures are: *first-fixation duration*



(the duration of the first fixation on a word), *single-fixation duration* (the duration on a word when only one fixation is made on the word), and *gaze duration* (the sum of the durations of all fixations on a word prior to moving to another word). In addition, the *total time* on a word (the sum of the durations of all fixations on a word including regressions) is often reported.

In Section 4, we will primarily discuss relevant studies in terms of the three first pass variables (first fixation, single fixation, and gaze duration), though studies dealing with specific target words typically also report the *spillover* time (typically measured as the fixation time on the word following the target word), the probability of fixating on the word, the probability of refixating the word (i.e., the probability of making additional fixations on the word following the initial fixation), and the probabilities of regressing back to the word and regressing back from the word. Another variable that has become increasingly used in studies using specific target words is *go-past time*, which is the time from first fixating on the word (including regressions back in the text) until a fixation is made to the right of it. This measure thus includes more than first pass time and can reasonably be construed as the time it takes upon reading the target word on first pass until it is successfully integrated with the on-going context.

While single-fixation duration, first-fixation duration, and gaze duration are the measures of choice for studying the time course of word recognition, a wider variety of measures is typically used in measuring processing associated with larger regions of text (as is typical in the types of studies we will discuss in Section 5). For the most part, in such studies, critical regions of text are identified, usually consisting of about 3–4 words, and the time it takes readers to read the regions of interest is measured. The standard measures are: *first-pass reading time* (the counterpart of gaze duration: the sum of all fixations in a region from first entering the region until leaving the region), *go-past* or *regression path duration* (the sum of all fixations in a region from first entering the region, including any regressions that are made, until moving to the right of the region), *regressions-out* (the probability of regressing out a region, generally limited to the first-pass reading of that region), *second-pass reading time* (the sum of all fixations in a region following the initial first pass time), and *total reading time* (the sum of all fixations in a region, both forward and regressive movements). First-fixation durations are also sometimes reported, especially when the region is short or when the researcher is interested in spillover effects from the previous region, but when regions are long and the disambiguating material is not likely to be included in the initial fixation, the first fixation is inappropriate.

Measures such as first pass time are generally referred to as early measures; second pass time (and total time, to the extent that it reflects second pass time rather than first pass time) are generally referred to as late measures (Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). The go-past and regressions-out measures are sometimes considered early measures (but sometimes as late measures); the occurrence of a regression probably reflects some difficulty in integrating a word when it is fixated, arguably an early effect, but the operation of overcoming this difficulty may well occur late in processing. Acutally, as Clifton, Staub, and Rayner (2006) pointed out, the terms early and late may

be misleading, if they are taken to line up directly with first-stage vs. second-stage processes that are assumed in some models of sentence comprehension (Rayner, Carlson, & Frazier, 1983; Frazier, 1987). Nonetheless, careful examination of when effects appear may be able to shed some light on the underlying processes. Effects that appear only in the late measures are in fact unlikely to directly reflect first-stage processes; effects that appear in the early measures may reflect processes that occur in the initial stages of sentence processing, at least if the measures have enough temporal resolving power to discriminate between distinct, fast-acting, processes.

Finally, it is ideally the case that a region of interest would consist of the same words. However, in psycholinguistic experiments this is not always possible and researchers often end up being forced to compare conditions that vary in specific words and/or the number of words. In such cases, a deviation from regression measure introduced by Ferreira and Clifton (1986) is typically used to attempt to correct (albeit imperfectly) for length differences.

#### **4. WORD RECOGNITION AND EYE MOVEMENTS**

One of the most robust findings in studies of eye movements and reading is that the ease or difficulty associated with understanding a word during reading clearly affects how long readers fixate on that word. In the remainder of this section, we will briefly review findings which have demonstrated effects due to word difficulty, contextual constraint, number of meanings (lexical ambiguity), phonological codes, semantic relations between words, morphological effects, and plausibility effects prior to moving to higher level effects. We will not provide an exhaustive review of all such studies. Rather, we will simply highlight the typical findings associated with each of these variables that plausibly have some relationship to how easy a word is to process.

##### **4.1. Word Difficulty**

There is a huge body of research on what makes individual words more or less difficult to process (in and out of context). Perhaps the most widely used standard index of word difficulty is word frequency (usually determined from corpus counts of adult reading materials). In reading, word frequency has a very reliable influence on how long readers look at a word (Just & Carpenter, 1980; Rayner, 1977). One problem in assessing the effect of word frequency is that it is fairly highly correlated with other variables, notably the length of a word. However, Rayner and Duffy (1986) and Inhoff and Rayner (1986) manipulated word frequency while controlling for word length and demonstrated that there was still a strong effect of frequency on fixation times on a word. The size of the frequency effect typically ranges from about 20 to 40 ms in first-fixation duration and from 30 to 90 ms in gaze duration (depending on the size of the difference in the actual frequencies in the stimuli). Since these initial reports, numerous studies have demonstrated frequency effects on fixation time measures (Altarriba, Kroll, Sholl, & Rayner, 1996; Calvo & Meseguer, 2002; Henderson & Ferreira, 1990, 1993; Hyönä & Olson,

1995; Kennison & Clifton, 1995; Kliegl, Grabner, Rolfs, & Engbert, 2004; Liversedge et al., 2004; Raney & Rayner, 1995; Rayner & Fischer, 1996; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner, Fischer, & Pollatsek, 1998; Rayner, Sereno, & Raney, 1996; Sereno & Rayner, 2000; Vitu, 1991). An interesting finding is that the frequency effect is attenuated as words are repeated in a short passage (Rayner, Raney, & Pollatsek, 1995) so that by the third encounter of a high-or low-frequency word, there is no difference between the two. (The durations of fixations on both high and low frequency words decrease with repetition, but the decrease is more dramatic for low frequency words.)

Is word frequency the only variable that affects how difficult a word is to process? Obviously, one can manipulate the visibility of the letters and get sizable increases in fixation time when the letters are harder to encode (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; Reingold & Rayner, 2006). But are there other, deeper, variables? One line of experimentation suggests that there is more to difficulty than frequency. This line of research has one set of participant's rate words as to their perceived familiarity, and then has another set read text in which target words are matched for (objective) frequency but different on rated familiarity. These experiments (Chafin, Morris, & Seely, 2001; Juhasz & Rayner, 2003; Williams & Morris, 2004) clearly demonstrated that familiarity influenced fixation times on words even when frequency was controlled (particularly for words that are low frequent). Thus, something is operative besides frequency, but then one wants to know what the objective variables are that are causing these differences in familiarity.

Another variable that has been examined in depth recently is *age-of-acquisition* (when a person is likely to have first encountered a word). Age-of-acquisition is determined either by corpus counts or by subjective ratings, and it has been shown to influence how long it takes to process a word (Juhasz, 2005). Juhasz and Rayner (2003, 2006) recently demonstrated that there was an effect of age-of-acquisition above and beyond that of frequency on fixation times in reading that was somewhat stronger than that of word frequency.

This effect raises several questions. First, is age-of-acquisition merely a cumulative frequency effect? That is, perhaps age of acquisition measures are merely better indicators of how frequently one has seen a word in text in one's lifetime than standard frequency measures. Instead, perhaps words that are learned earlier in life enjoy a special status. There is currently no resolution of this issue (see Juhasz, 2005). A second issue is whether effects such as familiarity effects are merely due to age-of-acquisition. Again, there is no clear resolution of this issue. A third issue is whether there are other variables that are confounded here. One obvious variable is the concreteness of a word, as words acquired early in life tend to be concrete and words acquired later in life tend to be abstract. Lastly, age-of-acquisition is also likely confounded with the frequency of a word in the spoken language, and given that phonological coding is important in reading (see below), this is another potentially important variable (Juhasz, 2005). The next 5 years of research will perhaps resolve these issues.

## 4.2. Contextual Constraint

In studies that manipulate predictability, sentence contexts are first prepared such that certain target words are either predictable or unpredictable from the context. A rating study is then performed, and the rating scores are used as the measure of predictability – how much the prior context constrains a given target word. Considerable research has demonstrated that words that are predictable from the preceding context are looked at for less time than words that are not predictable. Ehrlich and Rayner (1981) first demonstrated the effects of contextual constraint on fixation time, and the basic result has been confirmed a number of times (Ashby, Rayner, & Clifton, 2005; Balota, Pollatsek, & Rayner, 1985; Drieghe et al., 2005; Inhoff, 1984; Rayner Ashby et al., 2004; Rayner & Well, 1996; Schustack, Ehrlich, & Rayner, 1987). Not only are fixation time measures shorter on highly predictable words than low predictable words, readers also skip over highly predictable words more frequently than low predictable words (Binder, Pollatsek, & Rayner, 1999; Ehrlich & Rayner, 1981; Rayner & Well, 1996; Vitu, 1991).

One question about predictability is whether it is merely due to the objective *transitional probability* with which a given word follows another word in printed text (as determined via corpus counts). McDonald and Shillcock (2003a, 2003b) found that words with high transitional probability (e.g., *defeat* following *accept*) receive shorter fixations than words with low transitional probability (e.g., *losses* following *accept*). However, Frisson, Rayner, and Pickering (2005) subsequently found that differences in predictability were not merely due to transitional probability. In an experiment that had a highly controlled set of items, there was an effect of predictability (with transitional probability controlled), but no effect of transitional probability (with predictability controlled). This suggests that correlations between words in text have little influence unless people are conscious of them. In addition, Frisson et al. showed that predictability effects are detectable very early in the eye-movement record and between contexts that are only weakly constraining.

## 4.3. Number of Meanings (Lexical Ambiguity)

The number of meanings a word has influences fixation time on the word. Rayner and Duffy (1986), Duffy, Morris, and Rayner (1988), and Rayner and Frazier (1989) first demonstrated this lexical ambiguity effect, which has subsequently been replicated a number of times (Binder, 2003; Binder & Morris, 1995; Binder & Rayner, 1998; Dopkins, Morris, & Rayner, 1992; Folk & Morris, 2003; Kambe, Rayner, & Duffy, 2001; Rayner, Pacht, & Duffy, 1994; Sereno, 1995; Sereno, O'Donnell, & Rayner, 2006; Sereno, Pacht, & Rayner, 1992; Wiley & Rayner, 2000). The basic finding is that when a *balanced* ambiguous word (a word like *straw* with two approximately equally likely meanings) is encountered in a neutral context, readers look longer at it than an unambiguous control word matched on length and frequency, whereas they do not look any longer at a *biased* ambiguous word (a word like *bank* with one highly dominant meaning) in a neutral context than an unambiguous control word. In the former case, it appears that there is some sort of conflict between the two meanings. However, it appears that the subordinate meaning is not registered in the latter case; this is consistent with the finding

that if a subsequent disambiguating region makes clear that the subordinate meaning was intended, then there is considerable disruption to reading (long fixations and regressions). In contrast, when the disambiguating information precedes the ambiguous word, readers do not look any longer at the balanced ambiguous word than the control word. Apparently, the context provided sufficient information to guide the reader to the contextually appropriate meaning. However, in the case of biased ambiguous words when the subordinate meaning is instantiated by the context, readers look longer at the ambiguous word than the control word. This latter effect has been termed the *subordinate bias effect*. Rayner, Cook, Juhasz, and Frazier (2006) recently demonstrated that an adjective immediately preceding the target noun is a sufficient context to produce the effect.

All of the experiments mentioned above dealt with ambiguous nouns. In this context, results reported by Frazier and Rayner (1987) and by Pickering and Frisson (2001) are quite interesting. Frazier and Rayner (1987) found that syntactic category ambiguity (*trains* can be a noun or a verb) resulted in delayed effects; fixation time differences did not emerge on the target word itself (even with biasing context), but were delayed as if the system were trying to get additional information before committing to one meaning or the other. Pickering and Frisson (2001) likewise reported that with verbs with two meanings, the resolution of verb meaning ambiguity is delayed. Frazier and Rayner (1990) also found that nouns with different senses (e.g., *newspaper* meaning a publication or a physical object) yielded delayed effects in comparison to the typically reported results with nouns with two distinct meanings, and Frisson and Pickering (1999) found that metonymic expressions were treated differently from literal expressions (as reflected by the fixation time patterns on such expressions). The reasons for the different patterns of results are, as yet, far from clear, but they all show that lexical ambiguity influences the time that it takes to process a word in text.

#### 4.4. Phonological Coding

Words that are phonologically ambiguous (like *tear* and *wind*) have substantially longer gaze durations than unambiguous control words (Carpenter & Daneman, 1981) and words with two different spellings, but the same pronunciation (and two different meanings, such as *beech–beach* and *shoot–chute*), also have longer fixation times than unambiguous control words (Folk, 1999; Folk & Morris, 1995; Jared, Levy, & Rayner, 1999; Rayner, Pollatsek, & Binder, 1998). In addition, readers will often misinterpret the low frequency member of the pair as the higher frequency member if the context is highly constraining (Rayner et al., 1998, cf., Daneman & Reingold, 1993; Daneman, Reingold, & Davidson, 1995). Moreover, the finding previously mentioned (Pollatsek et al., 1992) that a parafoveal preview of a homophone of a target word provides greater preview benefit than a matched orthographic control indicates that phonological coding occurs early – even before a word is fixated.

More recently, Ashby and Clifton (2005) found that lexical stress influences how long readers look at a word. They further argued that implicit prosody (prosody generated internally by the reader), is a factor in how long readers look at a word (see also Hirotani, Frazier, & Rayner, 2006).

#### 4.5. Semantic Relationships between Words

Words that are semantically related to each other (and in close proximity to each other in the text) produce effects that appear to be analogous to semantic priming (Meyer & Schvaneveldt, 1971). Thus, the word *king* in the close proximity to *queen* results in shorter fixation times on *queen* than on an unrelated word in the same location (Carroll & Slowiaczek, 1986; Morris, 1994). Carroll and Slowiaczek (1986) found that such priming effects only occurred when the two words were in the same syntactic constituent, but Morris (1994) found some evidence for priming across constituents. Morris and Folk (1998) reported that this facilitation depends in part on whether the semantic associate of the target word is in linguistic focus (see Birch & Rayner, 1997). As we noted earlier, there are also repetition effects on fixation times in reading (Rayner et al., 1995; Raney & Rayner, 1995).

Several studies have demonstrated that specific kinds of semantic processing influence reading time on a word. Traxler, McElree, Williams, and Pickering (2005) and Traxler, Pickering, and McElree (2002) investigated the effect on readers' eye movements when the context forces a noun with no intrinsic temporal component to be interpreted as an event, as in the phrase *finish the book*. They found increased go-past time on the critical word or increased first pass time on the next region (see also Frisson & Pickering, 1999). Frisson and Frazier (2005) found that when a mass noun appears with plural morphology (e.g., *some beers*) or a count noun appears in the singular with a plural determiner (e.g., *some banana*), there is an increase in the duration of the first fixation on the critical word.

#### 4.6. Morphological Effects

Most research on word recognition has traditionally dealt with mono-morphemic words. This tradition has also been largely true of research on eye movements and word recognition. Recently, however, many studies have examined the processing of morphemically complex words (Inhoff, Radach, & Heller, 2000; Juhasz, Inhoff, & Rayner, 2005). This newer tradition started with studies (Hyönä & Pollatsek, 1998; Pollatsek, Hyönä, & Bertram, 2000) dealing with the processing of Finnish words (which by their very nature tend to be long and morphologically complex). Hyönä and Pollatsek (1998) found that the frequency of both the first and second constituent of two constituent compound words had large effects on the gaze duration on the compound word for long compound words (when the frequencies of the compound words were matched). However, Bertram and Hyönä (2003) found that the effects of the frequency of the first constituent were quite attenuated for shorter Finnish compound words. Similarly, somewhat smaller constituent frequency effects have recently been demonstrated with English compound words that were about the same length as the shorter Finnish compounds (Andrews, Miller, & Rayner, 2004; Juhasz, Starr, Inhoff, & Placke, 2003). Niswander-Klement and Pollatsek (2006) found a similar length-modulated constituent frequency effect for English prefixed words. That is, they found effects of the frequency of the root morpheme of the prefixed word (with the frequency of the words controlled), but that this effect was stronger for longer prefixed words.

Pollatsek and Hyönä (2005) recently demonstrated that semantic transparency (defined as whether the meanings of the constituents were related to the meaning of the word) had no effect on fixation times on Finnish compound words. There are conflicting findings in English, however. Juhasz (2006) found a main effect of transparency on gaze durations, whereas Frisson, Niswander-Klement, and Pollatsek (2006) obtained no effect. However, what is consistent across studies is that there is evidence for morphological decomposition for both opaque and transparent compounds, as there is an effect of the frequency of the first constituent for both.

#### 4.7. Plausibility Effects

Plausibility manipulations have been widely used in the context of studies of sentence parsing. In this section, we will briefly consider the extent to which plausibility/anomaly effects have immediate effects on eye movements. Although there are a few such studies (Braze, Shankweiler, Ni, & Palumbo, 2002; Murray & Rowan, 1998; Ni, Crain, & Shankweiler, 1996; Ni, Fodor, Crain, & Shankweiler, 1998), we will focus on a recent study by Rayner, Warren, Juhasz, and Liversedge (2004) because it has the virtue that the target word was identical across conditions. Rayner et al. had participants read sentences such as

1. John used a knife to chop the large carrots for dinner last night.
2. John used an axe to chop the large carrots for dinner last night.
3. John used a pump to inflate the large carrots for dinner last night.

Sentence 1 is a normal control condition; in sentence 2, it is somewhat implausible that John would use an axe to chop carrots (though one could conceive of such a scenario, as in a camping trip); sentence 3 creates an anomalous scenario. In all sentences, the target word is *carrots*. Rayner et al. found that while sentence 2 only caused mild disruption to reading (and it generally occurred after fixating on the target word), sentence 3 caused immediate disruption, but the disruption occurred rather late in processing the target word (i.e., after the first fixation). Rayner et al. concluded that the default situation in reading is for lexical processes to drive the eyes through the text in reading, but when something does not compute at a higher level (as with the anomalous sentences), then higher-order processes could intervene and cause the eyes to fixate longer (though the influence of such higher-order effects manifest themselves in relatively late processing) time measures on a target word (i.e., in the gaze duration and go-past measure, but not in the first fixation or single-fixation duration measure). What is still not clear, however, is how to objectively define the difference between the implausible and anomalous sentence contexts.

#### 4.8. Languages other than English

In all of our preceding discussions, we have focused largely on results of studies with English speaking readers. However, virtually all of the results that we have discussed hold for other alphabetic writing systems (and indeed some of the studies cited above have been done in other languages). But, we hasten to note that characteristics of the

writing system can have powerful influences on eye movements. Here, we mention two such writing systems: Hebrew and Chinese.

Hebrew is a more densely packed language than English, because the vowels are systematically deleted for skilled readers and function words are added as clitics to the end of content words. Of course, Hebrew is printed from right-to-left. These two facts result in the perceptual span being asymmetric to the left of fixation for Hebrew readers (Pollatsek, Bolozky, Well, & Rayner, 1981), as well as smaller than readers of English (for skilled readers of Hebrew, the span extends about 3 letters to the right of fixation to about 11 letters to the left). Interestingly, whereas morphology seems to have little influence on preview benefit for readers of English, readers of Hebrew do show morphological preview benefit (Deutsch, Frost, Peleg, Pollatsek, & Rayner, 2003; Deutsch, Frost, Pollatsek, & Rayner, 2000, 2005). Lexical variables that have been studied in Hebrew seem to yield similar results to English, and Hebrew readers show systematic landing position effects that are similar to English (Deutsch & Rayner, 1999).

Chinese is obviously even more densely packed than Hebrew. This results in a very small perceptual span, which extends from 1 character to the left of fixation to 2–3 characters to the right when reading from left-to-right (Inhoff & Liu, 1997, 1998), and much shorter saccades than English. While the concept of a word is not as well-defined in Chinese as it is in English (and Chinese readers often disagree concerning where word boundaries are located), most words consist of two characters (and most Chinese characters are like morphemes). It has recently been demonstrated that Chinese readers show frequency effects (Yan, Tian, Bai, & Rayner, 2006) and predictability effects (Rayner, Li, Juhasz, & Yan, 2005) that are quite comparable to readers of English.

#### **4.9. Summary of Eye Movements and Word Recognition**

Up to this point, we have reviewed some basic findings regarding how certain variables related to word-recognition mechanisms manifest themselves in the eye-movement record. In general, the primary assumption is that lexical factors play a large role in influencing when the eyes move. We do acknowledge that some of the effects we have discussed above are undoubtedly related to post-lexical processing. This raises the question of whether lexical or post-lexical processing, or both, is involved in the decision to move the eyes from one word to the next. Our bias is that many of the effects described above (though obviously not all of them) are reflecting lexical processing. As we will see later, the most successful models of eye-movement control are based on the premise that how long readers look at a word is influenced by the ease or difficulty associated with accessing the meaning of the word. Up to this point, word frequency and word predictability are primarily the indices that have been utilized in the models to predict fixation times. However, in the context of the E-Z Reader model, some effects of morphological complexity (Pollatsek, Reichle, & Rayner, 2003) and number of meanings (Reichle, Pollatsek, & Rayner, 2005b) have been modeled. Our bias is that lexical processing is the engine that drives the eyes and that higher-order effects most likely have influences when something does not compute (as with the anomaly study by Rayner et al., 2004 above). We turn now to the more



difficult issues of the effect of higher-order variables (such as parsing and discourse factors) on eye movements.

## 5. HIGHER-ORDER EFFECTS ON EYE MOVEMENTS: PARSING AND SYNTACTIC AMBIGUITY

Our review of the variables listed in section 4 suggests that a fairly clear, if incomplete, picture is developing with respect to how word processing/lexical factors influence eye movements during reading. However, the same is not true regarding higher-level factors (Clifton et al., 2006). Indeed, effects of parsing/syntactic ambiguity and discourse level variables seem to be highly variable in terms of how they influence eye movements. We will first discuss research on parsing and syntactic ambiguity, and then move to a discussion of the influence of discourse processing on eye movements.

Research on eye movements and syntactic ambiguity resolution has played a central role in the development of theories of sentence processing. It is beyond the scope of this chapter to discuss the extent to which a serial syntax-first type of theory (such as the Garden Path theory presented by Frazier, Clifton, Rayner, and colleagues) or a constraint-satisfaction type of theory (as championed by McDonald, Tanenhaus, and colleagues) can best account for sentence processing. Here, we will focus on the relationship between eye movements and parsing.

Some of the earliest eye-movement research on parsing and syntactic ambiguity held the promise that syntactic factors might have clearly identifiable influences on readers' eye movements. Frazier and Rayner (1982) examined the reading of sentences like (4) and (5), and found that first-fixation durations on the disambiguating region (underlined in the examples) were longer when a temporary ambiguity was resolved in favor of the un-preferred reading (in 4, when *this* was absent). This disruption persisted through the next several fixations, and also appeared as an increased frequency of regressions. Eye movements thus appeared to provide a clear window onto syntactic garden-path effects.

4. Since Jay always jogs a mile and a half (this) seems like a very short distance to him.
5. (The lawyers think) his/the second wife will claim the entire family inheritance (belongs to her).

Much of the disruption in (4) appeared in a region that followed the absence of an obligatory comma (or prosodic break), and disruption in (5) appeared in a sentence-continuation that had no counterpart in the non-disruptive control condition. These facts led to some concerns about this early work. But the force of the missing-comma criticism (i.e., that disruption was caused by the 'mistake' in the materials) is compromised by the fact that an equally obligatory comma was missing in the control condition, with no effect on reading times, and the lack of a closely matched control in (5) was corrected in later research (Rayner & Frazier, 1987).

Frazier and Rayner's results suggested that syntactic processing difficulty could be identified by quickly appearing disruptions in the eye-movement record. Rayner et al. (1983) further provided evidence for a similar conclusion about semantic processing difficulty. They found increased first-pass reading times for the disambiguating region (as well as increased durations of the first three fixations in this region) for sentences like (6), where the first noun is semantically anomalous under the presumably preferred initial analysis, compared to sentences like (7).

6. The kid hit the girl with a wart before he got off the subway.
7. The kid hit the girl with a whip before he got off the subway.

Another early demonstration of syntactic effects on eye movements was presented by Ferreira and Clifton (1986), who showed disruption in the disambiguating region of temporarily ambiguous sentences, both when the initial noun was animate (8) and when it was inanimate (9) and implausible as the subject of the following verb.

8. The defendant (who was) examined by the lawyer proved to be unreliable.
9. The evidence (that was) examined by the lawyer proved to be unreliable.

The disruption appeared in first-pass reading time, and it was argued that the semantic implausibility of the presumably preferred main clause analysis in (9) did not override readers' initial syntactic parsing preferences. However, Trueswell, Tanenhaus, and Garnsey (1994) argued that there were problems with some of Ferreira and Clifton's items and challenged their conclusion. They prepared what they considered to be more adequate sets of materials (which they carefully normed), and found that any effect of ambiguity on first-pass reading time was nonsignificant (indeed, nearly zero, in one experiment) in materials like (9), where semantic preferences weighed against the main clause analysis. They concluded that semantic factors could overturn syntactic preferences, favoring an interactive, constraint-satisfaction, model over the modular serial model favored by Ferreira and Clifton (1986).

Clifton et al. (2003) revisited this issue using materials taken from Trueswell et al. (1994). They varied parafoveal preview of the disambiguating information (since Trueswell et al. made interesting claims about the extent to which readers could use parafoveal information to disambiguate a temporary ambiguity) and participants' reading span. These two manipulations for the most part did not affect the magnitude of the disruption triggered by a temporary ambiguity and the first-pass time measures were similar to those reported by Trueswell et al. (1994). Semantic biases reduced the first-pass reading time measure of the temporary ambiguity effect to non-significance in sentences like (9) (although, similar to Trueswell et al., the interaction of semantic bias and temporary ambiguity was not fully significant, and, unlike Trueswell et al. the ambiguity effect did not go to zero). However, a very different pattern of results was observed for the go-past time and proportion of first-pass regressions out measures. These measures showed disruptive effects of temporary ambiguity that were at least as large in semantically biased inanimate-subject sentences like (9) as in animate-subject sentences like (8)

where no semantic bias worked against the presumed preference for a main clause analysis. Clifton et al. (2003) concluded that a full examination of the eye-movement record indicated that initial syntactic parsing preferences were not overcome by semantic biases, although such biases clearly affected overall comprehension difficulty for temporarily ambiguous and unambiguous sentences.

A subsequent analysis of the Clifton et al. (2003) data by Clifton et al. (2006) revealed that while an increase in regressions was responsible for the appearance of a garden-path effect in the inanimate subject condition, regressions were really quite infrequent, always <13% of the trials. This means that the garden-path effects that Clifton et al. (2003) observed in the inanimate-subject condition actually reflected eye-movement events that took place on a minority of the trials. On most trials in the inanimate-subject condition, eye-movements were not affected by temporary ambiguity. It is quite possible that the same holds true for the animate-subject condition: first-pass fixation durations may have been increased by temporary ambiguity on only a small minority of trials. This contrasts sharply with what is true of effects of lexical frequency on fixation durations, where the distribution shifts upwards for low-frequency words (Rayner, 1995; Rayner et al., 2003). No existing research on syntactic garden paths provides data on a large enough number of sentences to permit a convincing distributional analysis to be made (Clifton et al., 2006). It remains a challenge to researchers to devise a way of asking the question of whether first-pass reading times typically or exceptionally increase upon the resolution of a garden path.

In this section so far, we have focused on one difference in the literature on parsing with two studies that utilized the same manipulation, but which came to somewhat different conclusions, depending on which eye-movement measures were focused on. We suspect that this is not an isolated phenomenon (see Binder, Duffy, & Rayner, 2001; Clifton et al., 2006 for further discussion) and that exactly when a given effect will show up in the eye-movement record depends very much on the exact nature of the manipulation and the type of ambiguity present in the study.

In light of the findings we discussed earlier concerning lexical ambiguity resolution, an interesting question is whether the presence of two possible syntactic analyses slows reading, similar to when reading times are slowed when a word has two meanings that are roughly equivalent in frequency? Another question is how are eye movements affected when subsequent material reveals that the reader's initial analysis of a syntactic ambiguity is incorrect. Interestingly, the majority of studies on syntactic ambiguity have not reported any statistically significant effects on reading time in the ambiguous region itself (Staub & Rayner, 2006). A few studies (Frazier & Rayner, 1982; Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001) have found that an ambiguous region was in fact read more quickly than the corresponding region of an unambiguous control sentence. A few studies have also reported a slowdown in the ambiguous region compared to an unambiguous control (Clifton et al., 2003; Kennison, 2001; Ni et al., 1996; Paterson, Liversedge, & Underwood, 1999; Schmauder & Egan, 1998). However, an explanation other than ambiguity is often available (Staub & Rayner, 2006).

In sum, there is little evidence to indicate that syntactic ambiguity per se causes reading to slow down, and there seem to be circumstances in which ambiguity leads to faster reading times. Evidently, readers either do not consider multiple syntactic analyses in parallel (Frazier, 1978, 1987), or if they do, competition between these analyses does not disrupt processing (van Gompel et al., 2001, 2005). This conclusion stands in contrast with the conclusion from studies of the processing of lexical ambiguity, in which it has been clearly shown that competition between multiple word meanings slows reading.

Eye-movement data have been used to investigate the human parser's preferred analysis of many types of temporary ambiguity; an extensive list of references organized by the type of ambiguity under investigation appears in Clifton et al. (2006). Because there are reliable signs of disruption in the eye-movement record when an initial syntactic analysis is disconfirmed, it has been possible to test subtle and linguistically sophisticated hypotheses about how the parser constructs this initial analysis, and the factors that can influence this analysis. Eye-movement data have helped to reveal the parser's strategies for resolving "long-distance dependencies", in which a phrase appears some distance from the element from which it gets its thematic role, as in the question *Which boy did the teacher reward?*, where *which boy* is the object of the verb *reward* (Pickering & Traxler, 2001, 2003; Traxler & Pickering, 1996). They have also helped to reveal the processing implications of a phrase's status as an argument or adjunct of a verb (Clifton, Speer, & Abney, 1991; Kennison, 2002; Liversedge, Pickering, Branigan, & van Gompel, 1998; Liversedge, Pickering, Clays, & Branigan, 2003; Speer & Clifton, 1998).

There are open questions about the circumstances under which disambiguation results in a slowing down of forward saccades, regressive eye movements, or both (Altmann, 1994; Altmann, Garnham, & Dennis, 1992; Rayner & Sereno, 1994b, 1994c). However, both Frazier and Rayner (1982) and Meseguer, Carreiras, and Clifton (2002) demonstrated that when readers make regressive eye movements, they do not do so randomly. Instead, where these regressions go reflects some awareness of the point at which the reader's initial, incorrect analysis diverged from the correct analysis.

It is clear that eye-movement data have allowed researchers to probe the early stages of reading in a clear and direct fashion that is exceeded by no other technique. However, a survey reported by Clifton et al. (2006) showed that there was considerable variability in *when* a given manipulation had an effect; this often depended on the type of syntactic construction being used. Clifton et al. noted, and we would certainly agree, that eye-movement data have shown that much, if not quite all, of sentence comprehension is nearly immediate (within a fixation or so after encountering a critical word), as indicated by effects of syntactic or semantic anomaly or complexity and recovery from "garden paths". Eye-movement data have also shown that syntactic knowledge and at least some kinds of semantic, pragmatic, and real-world knowledge have effects even during fixations on the phrase that provides access to this knowledge. But their survey of the literature also clearly showed that the effects of such kinds of knowledge are more variable, even more ephemeral, than the effects that lexical frequency and lexical ambiguity have on eye movements. Fundamental questions, such as whether high-level

knowledge consistently affects fixation durations or affects them only now and then, remain unanswered. Furthermore, as Clifton et al. noted, there are disagreements about if and how one kind of knowledge (e.g., knowledge of the situation a sentence describes) modulates the effects of another kind of knowledge (e.g., knowledge of possible syntactic configurations), and disagreements about whether any such modulation is in turn modulated by differences in a reader's abilities and strategies.

In the end then, it is clear that higher-level variables that affect sentence processing and interpretation are much more complex, both in their definition and in their effect, than the variables that govern much of the variation in word recognition. It may be that understanding how these high-level variables operate is not something that can be induced from observations of eye-movement data (as has been true in large part in the domain of word recognition). Rather, as Clifton et al. (2006) noted, understanding must be guided by the development of more explicit theories than now exist of how syntactic, semantic, pragmatic, and real-world knowledge guide language processing.

## **6. HIGHER-ORDER EFFECTS ON EYE MOVEMENTS: DISCOURSE PROCESSES AND INFERENCES**

Whereas eye-movement data are often considered to be the gold standard in studies of sentence parsing, eye-movement data have had much less impact on studies dealing with discourse processes and inferences. In many respects this is quite surprising because it would seem important to determine exactly when readers make inferences as they read, and certainly eye-movement data hold the promise of revealing this type of temporal effect. In this section, we will review studies that have used eye movements in this manner. Our suspicion is that proportionally more eye-movement studies dealing with discourse processes and inferences will appear over the next few years.

In understanding text, readers must be able to integrate information within sentences and also make connections across sentences to form a coherent discourse representation. To what extent can eye-movement data reflect these processes? In this section, we will review research dealing with sentence and clause-wrap up, antecedent search, and on-line inferences.

### **6.1. Sentence and Clause Wrap-up**

Just and Carpenter (1980) found that fixation times on words that occurred at the end of a sentence were unusually long (in comparison to words that did not end a sentence) as measured by a regression analysis. Subsequently, Rayner et al. (1989) reported that when a target word ended a clause or a sentence, fixation times were inflated in comparison to when that same word did not end a clause or sentence. More recently, Rayner, Kambe, and Duffy (2000) confirmed this finding and further demonstrated that not only were fixations longer on clause and sentence final words, but that the next saccade was lengthened (see also Hill & Murray, 2000). So, readers slow down at clause and sentence boundaries, but

then send their eyes further into the next region of text as if processing capacity had been freed up once the wrap-up processes had been completed. Hirotani et al. (2006) have followed up on these findings and demonstrated that implicit prosody/intonation (imposed internally by the reader) is also very much involved in wrap-up effects.

## 6.2. Antecedent Search

The process of establishing a connection between an anaphoric element (such as a pronoun) and its antecedent in the text, antecedent search, is central to comprehending discourse. Pronominal reference and noun–noun reference are two such instances in which the correct linkage between discourse elements is required for text comprehension.

In pronominal reference, when a pronoun like *she* is encountered in the course of reading, the reader must identify an antecedent that matches it in number and gender. Sometimes, the process is trivially easy and no disruption is observed in the eye-movement record (Blanchard, 1987). If the pronoun involves violation of a gender stereotype (referring to a truck driver as *she*), fixations are inflated (Duffy & Keir, 2004; Sturt, 2003). If there is considerable distance between the pronoun (or anaphor) and the antecedent, readers' fixations are longer when the pronoun is encoded and the antecedent search may continue over the next couple of fixations (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994; O'Brien, Raney, Albrecht, & Rayner, 1997); if the distance between pronoun/anaphor and antecedent is close, fixations are not inflated as much. Of course, since pronouns are typically short words, readers skip over them quite frequently (thus making it difficult to determine exactly when the pronoun was encoded). Interestingly, van Gompel and Majid (2004) found that pronouns with infrequent antecedents yielded longer fixations in the encoding region than pronouns with more frequent antecedents; the effect did not occur on the pronoun itself but was slightly delayed to the region following the pronoun.

Just as a pronoun requires an antecedent, a definite Noun Phrase (NP) that does not directly refer to something outside the text requires a coreferring antecedent in the text. Thus, if a reader encounters the NP *the bird* after earlier reading about a robin (or vice versa), there is an antecedent link. Whereas pronouns carry little semantic information beyond gender and number, nouns typically have more semantic content, which facilitates the search for the antecedent. Duffy and Rayner (1990) found evidence that antecedent search time with anaphoric NPs was primarily localized on the target noun (see also van Gompel & Majid, 2004), so that there were no major spillover effects as with pronouns.

## 6.3. On-line Inferences

Within a discourse representation, the simplest kind of connection is one in which one word gains its reference through another word in the text, such as anaphor. However, elaborative inferences occur when information that has not been explicitly stated up to a

given point in the text is inferred by the reader. Eye-movement data have confirmed that elaborative inferences occur on-line during reading. These data have also served to differentiate between conditions in which readers make an inference and those in which they wait for more explicit information.

O'Brien, Shank, Myers, and Rayner (1988) asked readers to read short passages of text and fixation time was examined on a target word (*knife*) in the final sentence of a passage as in (10).

10. He threw the knife into the bushes, took her money, and ran away.

The target word was previously either explicitly mentioned in the text (as in the phrase *he stabbed her with his knife*) or only strongly suggested (as in *he stabbed her with his weapon*). O'Brien et al. (1988) found no difference in gaze duration on the target word across these two conditions. It thus appears that the concept *knife* had been inferred from the prior context when the word *weapon* was actually present. In contrast, when the text did not strongly suggest the concept of a knife (as in *he assaulted her with his weapon*), gaze duration on the target word *knife* was longer compared to when it had been explicitly mentioned or strongly suggested earlier in the passage. These results indicate that the longer fixation time on the target word was due to a memory search for its antecedent and that antecedent search begins immediately upon fixating the target word.

Although O'Brien et al. (1988) found evidence for on-line elaborative inferences, they also demonstrated that such inferences only occur when there is a "demand sentence" (which invited the reader to make the inference) just prior to the sentence containing the target word. A subsequent study by Garrod, O'Brien, Morris, and Rayner (1990) further constrained the conditions under which elaborative inferences occur. Their data suggested that the presence of a demand sentence invites the reader to actively predict a subsequent expression and elaborative inferences only occur when there is an anaphoric relationship between two nouns.

Other studies have observed rather immediate effects in the eye movement record of bridging inferences (Myers, Cook, Kambe, Mason, & O'Brien, 2000) and the integration of role fillers in scripted narratives (Cook & Myers, 2004; Garrod & Terras, 2000). The most interesting point in these studies is that higher-order variables show immediate effects in the eye-movement record. At some level, studies such as these provide a problem for models of eye-movement control in reading, which are largely based on the assumption that lexical processing is the engine that drives the eyes through the text in reading. We now turn to a discussion of such models.

## 7. MODELING EYE MOVEMENTS IN READING

Obviously, developing a quantitative model that could explain all the phenomena that have been observed in reading is a task that is beyond us at present, and it may be an

unreachable goal. That is, there are so many factors that influence reading, ranging from the legibility of the characters up through the frequency of the words in the language, the complexity of the syntax, the higher-order organization of the text, and the real-world knowledge shared by the author and the reader. If a model were to try to explain all of these factors, it would almost certainly either be hideously complex or degenerate into a multiple regression equation that merely re-affirmed that all of these variables (and others) play a part in reading.

As a result, our belief is that for a model of reading to be of some value at this point in time, it needs to be able to explain a significant part of the reading process, yet be simple enough so that it is a useful heuristic tool for understanding which phenomena it can explain and which it can not explain. At present, there are a number of programs of research that are developing quantitative models of eye movements in reading. Though they differ wildly in many respects, they all share certain features, as none are attempting any serious modeling of how text is parsed or how discourse structures are being constructed. Some focus on lexical processes primarily influencing eye movements, whereas others attempt to explain eye movements in reading largely by lower-level, perceptual and motor processing or via Ideal Observer procedures. The primary models are E-Z Reader (Pollatsek, Reichle, & Rayner, 2006; Rayner, Ashby, Pollatsek, & Reichle, 2004; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006b; Reichle, Rayner, & Pollatsek, 1999, 2003), SWIFT (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005; Kliegl & Engbert, 2003), Glenmore (Reilly & Radach, 2003), SERIF (McDonald, Carpenter, & Shillcock, 2005), Mr. Chips (Legge, Klitz, & Tjan, 1997; Legge, Hooven, Klitz, Mansfield, & Tjan, 2002), and the Competition/Interaction model (Yang & McConkie, 2001, 2004). These models differ on a number of dimensions, but space does not permit us to discuss the models in detail (though most of them were reviewed by Reichle et al., 2003). As a result, we will largely try to illustrate this modeling enterprise through outlining the modeling enterprise we are associated with (versions of the E-Z Reader model), and will briefly comment on how some of the other models differ and the points of controversy.

### **7.1. The E-Z Reader Model**

The E-Z Reader model focuses on trying to explain how lexical processing influences the progress of the eyes through the text. We think this is a justifiable focus, as a case can be made that higher-order variables (such as constructing discourse structures) have a more indirect influence on how readers make their way through the text. That is, individual fixation durations on words are typically about 250 ms and gaze durations typically average no more than 300–350 ms, and the motor programming time for an eye movement is far from instantaneous, as the time necessary to make an eye movement to the simplest visual stimuli (e.g., the onset of a point of light) takes close to 200 ms. Thus, it seems unlikely that readers are waiting for all levels of processing to be completed (e.g., constructing a parsing tree of the sentence to that point in the text) before sending a signal to the eye-movement system to move on to the next word.



Instead, we think it is a reasonable working hypothesis that linguistic processing affects eye movements in two different ways. First, there is a relatively low-level of linguistic processing that keeps the eyes moving forward – we have tentatively associated this with lexical processing although this might be too restrictive. Second, higher-level processing is occurring in parallel with this lexical processing system, and when it becomes clear that the higher-level processing is having difficulty, either because the higher-order processing system is falling behind the encoding of words or because an “error” has been detected (such as occurs when readers read something that is anomalous or when they misparse a sentence), this second system intervenes to tell the first system to either stay in place until the second system catches up or to go back and make an attempt to repair the damage. We think that such a hypothesis makes sense, as reading would proceed far more slowly than the normal 300 words per minute if the reader had to wait on each word until its significance within the text was ascertained. Whether this is in fact how reading goes on is, of course, an open question. However, we think it is a reasonable starting point for thinking about the reading process. In addition, modeling of the first stage then no longer seems like an insurmountable task; moreover, it allows one to define a well-defined set of eye-movement data to be modeled: all eye movements in reading other than regressions back to prior words. Let us see how this might be done.

In the E–Z Reader model, we have posited two more-or-less modular systems: the cognitive system and the motor system. Thus, there are sets of assumptions relating to the events in the cognitive system that trigger eye movements and sets of assumptions about how the commands to execute eye movements actually get carried out. In E–Z Reader, the first basic assumption is that a stage of lexical access causes a program for an eye-movement to the next word in the text to be initiated, and the second basic assumption is that this eye-movement program is executed within about 150 ms after it is initiated – in the normal state of affairs. Obviously, the above two assumptions can not be the whole story as they would predict that each word is fixated exactly one time, and we know that some words are skipped and that others are fixated more than once. Thus, other assumptions need to be made as well. In addition, one needs to include some model of covert attention in a model; that is, one has to make assumptions about what is being processed at any moment in time. It is the type of attentional assumption that is perhaps the major distinguishing feature of the various reading models. In E–Z Reader, it is assumed that low-level visual processing goes on in parallel over the whole visual field – such low-level processing, among other things, allows the eye-movement system to be able to target saccades. In contrast, E–Z Reader assumes that lexical processing is serial in the sense that only one word is being processed at any moment in time. However, we want to emphasize that this does not mean that only one word is processed on a fixation; on the contrary, the usual state of affairs in the E–Z Reader model is that two words are processed on a fixation, and, not infrequently, at least partial processing of three words occurs on a fixation.

As just indicated, the key assumption is that words are attended to (and thus lexically processed) one at a time. The simplest possible assumption about how this attention management would occur was made by Morrison (1984); he posited that when the

reader lexically accesses a word (a) an eye-movement program is initiated and (b) attention shifts to the next word. For reasons we discuss later, we thought this was too simple, so we chose a slightly more complex model. We assumed that an earlier stage of lexical access (L1) is the trigger for an eye movement, but that a later stage of lexical access (L2) is the trigger for an attention shift and hence the start of processing the next word. One can view this system as one in which readers have developed a “cheat” and trigger an eye movement (which takes appreciable time to execute) when they are reasonably sure that the word has been comprehended (L1), but only start to process the next word after the lexical process has completed (L2). In the model, we assume that both (a) the duration of the first stage and (b) the time between the completion of the first stage and the completion of the second stage are linear functions of log frequency. (We will subsequently refer to the difference in time between when L1 and L2 are completed as the duration of the L2 stage.) We also assume that the durations of both stages are affected by the predictability of the word from the prior text. In earlier versions of the model, we assumed that the influence of predictability was multiplicatively related to the influence of word frequency, but then realized (Rayner, Ashby et al., 2004) that an additive version (i.e., that frequency and predictability made independent contributions to the speed of lexical access) was better.

There is one more assumption about how eye movements are triggered. This is an assumption related to refixating the currently attended word. In an earlier version, we assumed that a refixation on the currently attended word was programmed automatically when a fixation began. (We will discuss below why this does not imply that all words are refixated.) However, there were problems with this simple mechanism for refixations on words, and in the current version, refixations are programmed (a) with a probability  $< 1$  when a word is fixated and (b) the probability depends on how far from the center of the word the fixation is. However, as the refixation component is not a particularly well-worked out aspect of the model, we will only give a sketch how these mechanisms can explain refixations.

Now we turn to the assumptions about the programming and execution of eye movements. The key assumption (adapted from Morrison, 1984) is that later eye-movement programs *can* cancel earlier eye-movement programs. This assumption is based on the work of Becker and Jürgens (1979), who examined a much simpler situation than reading. They had participants fixate a small area of light at point 1, which then moved abruptly to point 2. When this was all that happened, people quickly fixated point 2. The key trials were when the light moved abruptly again to point 3. If the gap in time between the two movements was sufficiently large, participants fixated point 2 and then point 3. However, when the gap was short enough, they merely fixated point 3, indicating that they could cancel the eye movement to point 2.<sup>3</sup> To capture saccade cancellation in our modeling, we assume that there are two stages in a motor program: a labile stage followed by a non-labile stage. If an

---

<sup>3</sup>At intermediate times, sometimes there were compromise saccades. We have modeled such compromise saccades and there was little change in the predicted values in the simulations.

eye movement program is in the labile stage, a subsequent eye-movement program can cancel it, whereas if it is in the non-labile stage, it will be executed regardless of what other programs are initiated. (In the latest versions of our model, the former stage is assumed to be about 100 ms and the latter stage about 50 ms.)

Let us see how this relates to reading. First, let us examine skipping. What words are skipped? Mainly words that take little time to process such as frequent and/or predictable words. In *E-Z Reader*, as we shall see, a reader typically has done some processing of a word before it is fixated and then finishes processing the word when it is fixated and the signal from the completion of the L1 stage typically occurs about 100–150 ms after the word is fixated. This produces a signal to fixate the next word in the text (word  $n+1$ ). Not that much after this (the L2 duration, which is typically  $\sim 50$  ms), attention moves to word  $n+1$  and lexical access of it begins. However, if word  $n+1$  is easy to process, stage L1 will be quick, and can finish before the end of the labile stage of the eye-movement program to fixate word  $n+1$ . The completion of the L1 stage of word  $n+1$  in these cases thus produces a program to fixate word  $n+2$  which will then cancel the program to fixate word  $n+1$ , and hence word  $n+1$  will be skipped. Moreover, as we have argued above, the model predicts that the probability that this will happen will be greater for more frequent and/or more predictable words.

Before moving on, let us briefly sketch how this cancellation assumption affects refixations. This is easier to do if we use our earlier simple assumption that a program to refixate a word is made automatically upon first fixating a word. As we indicated above, this assumption would cause all words to be refixated unless the refixation program is cancelled. When will the refixation program be cancelled? Answer: when the program to fixate word  $n+1$  occurs during the labile stage of this refixation program. This will occur when word  $n$  (the fixated word) is easy to process because it is high frequency and/or predictable. Thus, the model predicts that lower-frequency words and less predictable words are more likely to be refixated. (We should emphasize that we assume that the eye-movement system knows nothing about cognition; thus all eye-movement programs are assumed to have the same properties, regardless of the triggering mechanism.)

In addition, one needs to make assumptions about the targeting of the saccades. There are two issues involved here. The first is to specify exactly what the target of a saccade is. The second is to posit the error involved in the targeting procedure. In *E-Z Reader*, for both of these issues, we basically imported the data and assumptions from work by McConkie et al. (1998). We assume that the target of a saccade is the center of a word, but that this is subject to both a constant error (i.e., short saccades, on average, will tend to overshoot the target location and long saccades, on average, will tend to undershoot the target location) and random error. These assumptions give a pretty good account of the landing positions on a word. We should also emphasize that they also imply that the target word (i.e., the attended word) is not always the fixated word (due to noise in the oculomotor system). One more processing assumption is worth mentioning in this regard. That is, that the speed of lexical processing does not only depend on the frequency of a word and its predictability, but also where its letters are with respect to the fixation

point – the further these letters are from fixation, the slower processing is (due to visual acuity concerns). This implies not only that processing a word in the parafovea before it is fixated is less efficient than when it is fixated, but that longer words will be processed more slowly than shorter words (all else being equal) because, on average, the letters of longer words will be further from fixation. Moreover, this length effect is magnified, the further the fixation point is from the center of the word.

This, in outline, summarizes the E–Z Reader model. As we hope we have indicated, it can account qualitatively for many of the major phenomena of reading related to word variables. It predicts that longer, less frequent, and/or less predictable words will be (a) fixated longer, (b) skipped less, and (c) refixated more often. Moreover, it gives a very good quantitative account of these phenomena for sentence reading (see Pollatsek et al., 2006; Rayner et al., 2003). Furthermore, it does so using quite reasonable assumptions about how long lexical processing takes and how long motor programs take. We make the latter point, because there is still some skepticism that cognitive processes in reading can possibly be fast enough to influence eye movements in an on-line fashion.

Before going on to briefly discuss the competing models, we need to return to a point that we quickly slid over before: the motivation for positing two stages of lexical processing rather than just have a single stage be the trigger for an eye-movement program and an attention shift. One reason is that, if one assumes that there is a single stage of lexical processing, there can be no delayed effects due to difficulty in lexical processing. That is, a one-stage model would predict that one continues to process (and fixate) a word until it is processed, and then, simultaneously, (a) an eye movement is programmed to fixate word  $n+1$  and (b) attention shifts to word  $n+1$ . Thus, the time that word  $n+1$  will be processed in the parafovea before it is fixated will be merely equal to the latency of the eye-movement program and will not be a function of the difficulty of word  $n$ . However, as indicated in an earlier section of this chapter, there are many findings that the difficulty of word  $n$  often “spills over” to affect the time taken to process word  $n+1$ . Our assumption that the duration of the second stage of word processing is also a function of difficulty of lexical processing explains such spillover effects. We should point out, however, that the E–Z Reader model does not predict such spillover effects when early stages of word identification are manipulated. In one such manipulation (Reingold & Rayner, 2006), a target word was made quite faint and this increased the gaze durations on this word by over 100 ms. However, in this case there were no spillover effects, consistent with the likelihood that this manipulation only affected early stages of lexical processing.

A second motivation for our assumption of two stages of lexical processing is that it nicely explains the phenomenon that there is less preview benefit when the fixated word is more difficult to process (e.g., Henderson & Ferreira, 1990; Kennison & Clifton, 1995; White, Rayner, & Liversedge, 2005). This again, is explained by our assumption that the duration of the second stage of lexical processing is a function of the difficulty of processing the fixated word. (Roughly speaking, in the E–Z Reader model, the amount of time a word is processed in the parafovea is equal to the eye movement latency minus the

duration of this second lexical processing stage.) A third reason for our positing two processing stages is that it also gives a significantly better account of the general data. That is, a model that assumes only one stage can actually produce a decent global fit to the reading data (i.e., predict mean gaze durations, mean first-fixation durations, mean skipping rates, and other indices of reading as a function of word frequency, length, and predictability). However, in order to do so, a one-stage model needs to posit that processing of parafoveal material is much more efficient than it is, which, in turn, leads it to predict far too big effects of *preview benefit* (see Section 4 above).

As a result, we feel reasonably comfortable in the claim that our model gives a good overall account of eye movements in reading as long as there are no “higher-order” difficulties in the text. However, it is admittedly vague on some points, such as what the two stages of lexical access are, and there are some conceptual problems that we need to address in the future. However, we will discuss those after briefly discussing some of the other quantitative models of reading, perhaps not from the most unbiased perspective.

## 7.2. Other Models

One model we will spend little time on is the Competition/Interaction model of Yang and McConkie (2001, 2004). The reason for this is that the major way that their model differs from E–Z Reader is that they posit that lexical processing plays only a minor role in reading. This inference is mainly drawn from a rather unnatural paradigm in which text keeps on disappearing rapidly and which may have little to do with normal reading. Clearly, such a model is at variance with much of the data reviewed earlier, which shows that many linguistic variables play a role in how long words are fixated. They admit that such linguistic variables can play a role, but only in some cases when processing is lengthened appreciably. However, this is at odds with the data that show, for example, that the distribution of fixation durations on less frequent words is essentially the same shape as that for more frequent words, except that the whole distribution is shifted to the right (see Rayner, 1995; Rayner, Liversedge et al., 2003). The variance for the less frequent word, admittedly, is slightly bigger; however, the pattern is not what would be predicted if a large majority of fixations durations were determined by low-level processing and a few long fixations were due to difficulty in processing the low-frequency words.

In contrast, lexical processing is at the core of two of the other competing models (SWIFT and Glenmore). A major difference between them and E–Z Reader is that lexical processing is assumed to go on in parallel in both models. In SWIFT, it is assumed that lexical processing is simultaneously occurring on four words (from the one to the left of fixation up to the two words to the right), whereas in Glenmore, processing only occurs on the fixated word and the one to the right. Moreover, in SWIFT, the assumptions about the relation between lexical processing and eye-movement control are complex; for example, lexical excitation is assumed to have an inverted U-shaped function as far as how the word attracts attention and thus an eye movement. Both models appear to make these parallel assumptions about lexical processing because of so-called parafoveal-on-foveal effects (Kennedy, 2000; Kennedy & Pynte, 2005), in which the duration of

fixations on word  $n$  are influenced by the properties of word  $n+1$ . That is, such effects seem incompatible with a serial processing model such as E-Z Reader. Space does not permit a full discussion of this issue, so we will briefly address three points.

First, lexical parafoveal-on-foveal effects are far from an established empirical phenomenon (see Rayner & Juhasz, 2004; Rayner, White, Kambe, Miller, & Liversedge, 2003). That is, although there appear to be reliable parafoveal-on-foveal effects when word  $n+1$  is visually or orthographically unusual, parafoveal-on-foveal effects caused by lexical properties are far from established. Thus, at present, there is nothing in this phenomenon that is particularly suggestive that lexical processing is occurring in parallel. Second, E-Z Reader in fact can explain some parafoveal-on-foveal effects, as it predicts that words are occasionally mistargeted; hence, with some reasonable probability the reader intended to fixated word  $n+1$  (and is attending to it) but the saccade fell short and the reader is fixating word  $n$  (see Rayner, Warren et al., 2004). However, because of the uncertainty of many of the phenomena, we have not tried to model parafoveal-on-foveal effects, so we can not be sure that the model does in fact adequately explain such phenomena. Third, we hesitated to posit parallel processing of words because it does not seem all that psychologically plausible. That is, as current modeling of word identification is having a very hard time in explaining how a single word is identified, we think it is almost like positing magic to say that more than one word is processed in parallel. (There might be exceptions for frequent combinations of short words like “to the”.) More generally, from a modeling standpoint, we think that the serial processing assumption makes the model more transparent, and thus makes it a more valuable heuristic device for understanding reading, and that one should abandon it only if there are compelling reasons to do so.

### 7.3. Models' Summary

Let us briefly close this section by commenting on the problems with the model and what we see as what needs to be accomplished in the next 3–5 years in modeling. First, E-Z Reader, like the other models, does not contain a serious model of word recognition. We have merely posited processes that have finishing times influenced by variables such as frequency. This needs addressing. Second, we have been quite vague about what L1 and L2 are, and have merely used vague terms such as “lexical access”, without specifying what codes are accessed (e.g., orthographic, phonological, semantic, syntactic). One arena that this lack is particularly apparent is that, at present, we have no complete satisfactory way of explaining the effects of lexical ambiguity discussed earlier. One can possibly push these effects off to the “higher-order processing system”, but this seems unsatisfactory. We are currently working on this (Reichle et al., 2006a). Second, as indicated earlier, there are certain syntactic and semantic anomalies that have quite immediate effects on eye movements. This suggests that, if one wants to model all immediate eye-movement effects (presumably the scope of the E-Z Reader model and competing models), one has to come up with a satisfactory explanation of why these effects (but not others) are immediate. We think the solution to this problem will be non-trivial. However, let us close on a more positive note. That is, our

modeling was previously also fairly vague about its assumptions about early (pre-lexical) visual processing and its relation to lexical processing. We have addressed most of these issues in a recent article (Pollatsek et al., 2006).

## 8. GENERAL SUMMARY

In this chapter, we have reviewed data on eye movements in reading. It is clearly the case that the ease or difficulty associated with the fixated word strongly influences how long the reader will look at that word. Current models of eye-movement control have seized heavily on this fact and have been very successful in accounting for a large amount of the variance in how readers move their eyes through text. But, as we have seen, there are also clearly other influences on eye movements besides low-level lexical factors. Nevertheless, it does appear that a great deal can be accounted for via the assumption that lexical processing (or the ease or difficulty associated with the fixated word) is the engine driving the eyes through the text and that higher-order information primarily serves to intervene when something does not compute well.

Finally, we hope it is obvious that eye movements continue to be an excellent way to study the moment-to-moment processes inherent in the reading process. As we have seen in our discussion of higher-order factors, there are many difficulties in interpreting the eye-movement record when the manipulations involve syntax or higher-order semantic processing. However, in terms of inferring moment-to-moment mental processing during reading, it is not clear that there are any other measures that provide the temporal precision that eye movements provide.

## ACKNOWLEDGMENT

Preparation of this chapter was made possible by grants HD17246 and HD26765 from the National Institute of Health.

## REFERENCES

- Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish–English bilinguals. *Perception & Psychophysics*, *63*, 875–890.
- Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed language sentences: Evidence from eye fixation and naming times. *Memory & Cognition*, *24*, 477–492.
- Altmann, G. T. M. (1994). Regression-contingent analyses of eye movements during sentence processing: A reply to Rayner and Sereno. *Memory & Cognition*, *22*, 286–290.
- Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, *31*, 685–712.

- Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, *16*, 285–311.
- Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*, B89–B100.
- Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*, *19*, 391–426.
- Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology*, *58A*, 1065–1086.
- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364–388.
- Becker, W. & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*, 967–983.
- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, *48*, 615–634.
- Binder, K., Duffy, S., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, *44*, 297–324.
- Binder, K. S. (2003). Sentential and discourse topic effects on lexical ambiguity processing: An eye-movement examination. *Memory & Cognition*, *31*, 690–702.
- Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1186–1196.
- Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1162–1172.
- Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review*, *5*, 271–276.
- Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*, 653–660.
- Blanchard, H. E. (1987). Pronoun processing during fixations: Effects on the time course of information utilization. *Bulletin of the Psychonomic Society*, *25*, 171–174.
- Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, *31*, 25–44.



- Briihl, D., & Inhoff, A.W. (1995). Integrating information across fixations in reading: The use of orthographic bodies and of exterior letters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 55–67.
- Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In: G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–148). Oxford: Elsevier.
- Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical, and contextual factors. *Spanish Journal of Psychology*, *5*, 66–77.
- Carpenter, P. A., & Daneman, M. (1981). Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior*, *28*, 138–160.
- Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In: K. Rayner (Ed), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). New York: Academic Press.
- Carroll, P. J. & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition*, *14*, 509–522.
- Chace, H. E., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview benefit: Effects of reading skill. *Canadian Journal of Experimental Psychology*, *59*, 209–217.
- Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*, 225–235.
- Clifton, C. Jr., Speer, S., & Abney, S. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, *30*, 251–271.
- Clifton, C. Jr., Staub, A., & Rayner, K. (2006). Eye movements in reading words and sentences. In: R. P. G. von Gompel, M. H. Fischer, W. S. Murray, R. C. Hill (Eds), *Eye movements: A window on mind and brain*. Oxford: Elsevier.
- Clifton, C., Traxler, M., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317–334.
- Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, *50*, 268–288.
- Daneman, M., & Reingold, E. (1993). What eye fixations tell us about phonological recoding during reading. *Canadian Journal of Experimental Psychology*, *47*, 153–178.
- Daneman, M., Reingold, E., & Davidson, M. (1995). Time course of phonological activation during reading: Evidence from eye fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 884–898.

- Deutsch, A., Frost, R., Peleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review*, *10*, 415–422.
- Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2000). Early morphological effects in word recognition: Evidence from parafoveal preview benefit. *Language and Cognitive Processes*, *15*, 487–506.
- Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2005). Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Language and Cognitive Processes*, *20*, 341–371.
- Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes*, *14*, 393–421.
- Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language*, *31*, 461–477.
- Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 954–969.
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*, 551–559.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.
- Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, *33*, 103–119.
- Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, *22*, 75–87.
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655.
- Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*, 621–636.
- Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory & Language*, *25*, 348–368.
- Folk, J. R. (1999). Phonological codes are used to access the lexicon during silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 892–906.

- Folk, J. R., & Morris, R. K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1412–1429.
- Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition*, *31*, 87–99.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Doctoral dissertation, University of Connecticut, Storrs.
- Frazier, L. (1987). Sentence processing: A tutorial review. In: M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.
- Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, *26*, 505–526.
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*, 181–200.
- Frisson, S., & Frazier, L. (2005). Carving up word meaning: Portioning and grinding. *Journal of Memory and Language*, *53*, 277–291.
- Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2006). The role of semantic transparency in the processing of English compound words, submitted.
- Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1366–1383.
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 862–877.
- Garrod, S., Freudenthal, S., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, *33*, 39–68.
- Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 250–257.
- Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, *42*, 526–544.
- Gautier, V., O'Regan, J. K., & LaGargasson, J.F. (2000). 'The skipping' revisited in French: Programming saccades to skip the article 'les'. *Vision Research*, *40*, 2517–2531.

- Haberlandt, K. (1994). Methods in reading research. In: M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp.1–32). San Diego: Academic Press.
- Heller, D., & Radach, R. (1999). Eye movements in reading: Are two eyes better than one? In: W. Becker, H. Deubel, & T. Mergner, (Eds) *Current oculomotor research: Physiological and psychological aspects* New York: Plenum Press.
- Henderson, J. M., Dixon, P., Petersen, A., Twiley, L. C., & Ferreira, F. (1995). Evidence for the use of phonological representations during transsaccadic word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 82–97.
- Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 417–429.
- Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology*, 47, 201–221.
- Hill, R. L., & Murray, W. S. (2000). Commas and space: Effects of punctuation on eye movements and sentence parsing. In: A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds), *Reading as a perceptual process* (pp. 565–590). Oxford: Elsevier.
- Hirotnani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54, 425–443.
- Hyönä, J., Bertram, R., & Pollatsek, A. (2004). Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory & Cognition*, 32, 523–532.
- Hyönä, J., & Olson, R. K. (1995). Eye movement patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1430–1440.
- Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1612–1627.
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23, 612–624.
- Inhoff, A. W. (1989). Lexical access during eye fixations in reading: Are word access codes used to integrate lexical information across interword fixations? *Journal of Memory and Language*, 28, 444–461.
- Inhoff, A. W., & Liu, W. (1997). The perceptual span during the reading of Chinese text. In: H. -C. Chen (Ed.), *The cognitive processing of Chinese and related Asian languages*. Hong Kong: The Chinese University Press.

- Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 20–34.
- Inhoff, A. W., Pollatsek, A., Posner, M. I., & Rayner, K. (1987). Covert attention and eye movements during reading. *Quarterly Journal of Experimental Psychology*, 41A, 63–89.
- Inhoff, A. W., Radach, R., & Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, 42, 23–50.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40, 431–439.
- Ishida, T., & Ikeda, M. (1989). Temporal properties of information extraction in reading studied by a text–mask replacement technique. *Journal of the Optical Society A: Optics and Image Science*, 6, 1624–1632.
- Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General*, 128, 219–264.
- Johnson, R. E. (2006). The flexibility of letter coding: Nonadjacent letter transposition effects in the parafovea. In: R. P. G. von Gompel, M. H. Fischer, W. S. Murray, R. C. Hill (Eds), *Eye movements: A window on mind and brain*, Oxford: Elsevier.
- Johnson, R. E., Perea, M., & Rayner, K. (2006). Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance* (in press).
- Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, 131, 684–712.
- Juhasz, B. J. (2006). The Influence of semantic transparency on eye Movements during English compound word recognition. In: R. P. G. von Gompel, M. H. Fischer, W. S. Murray, R. C. Hill (Eds), *Eye movements: A window on mind and brain*, Oxford: Elsevier.
- Juhasz, B. J., Inhoff, A. W., & Rayner, K. (2005). The role of interword spaces in the processing of English compound words. *Language and Cognitive Processes*, 20, 291–316.
- Juhasz, B. J., Liversedge, S. P., White, S. J., & Rayner, K. (2006). Binocular coordination of the eyes during reading: Word frequency and case alternation affect fixation duration but not binocular disparity. *Quarterly Journal of Experimental Psychology* (in press).
- Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1312–1318.
- Juhasz, B. J., & Rayner, K. (2005). Age of acquisition effects on eye movements in reading. *Visual Cognition*, 13, 846–863.

- Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, *94*, 223–244.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.
- Kambe, G. (2004). Parafoveal processing of prefixed words during eye fixations in reading: Evidence against morphological influences on parafoveal preprocessing. *Perception & Psychophysics*, *66*, 279–292.
- Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words. *Memory & Cognition*, *29*, 363–372.
- Kennedy, A. (2000). Parafoveal processing in word recognition. *Quarterly Journal of Experimental Psychology*, *53A*, 429–456.
- Kennedy, A., & Pynte J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153–168.
- Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review*, *8*, 132–138.
- Kennison, S. M. (2002). Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research*, *31*, 65–81.
- Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 68–81.
- Kliegl, R., & Engbert, R. (2003). SWIFT explorations. In: J. Hyönä, R. Radach, & H. Deubel (Eds) *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 391–412). Amsterdam: Elsevier.
- Kliegl, R., & Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychonomic Bulletin & Review*, *12*, 132–138.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262–284.
- Legge, G. E., Hooven, T. A., Klitz, T. S., Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research*, *42*, 2219–2234.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review*, *104*, 524–553.
- Lima, S. D. (1987). Morphological analysis in reading. *Journal of Memory and Language*, *26*, 84–99.

- Lima, S. D., & Inhoff, A. W. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 272–285.
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science*, *4*, 6–14.
- Liversedge, S. P., Pickering, M. J., Branigan, H. P., & van Gompel, R. P. G. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 461–475.
- Liversedge, S. P., Pickering, M., Clayes, E. L., & Branigan, H. P. (2003). Thematic processing of adjuncts: Evidence from an eye-tracking experiment. *Psychonomic Bulletin & Review*, *10*, 667–675.
- Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements when reading disappearing text: is there a gap effect in reading? *Vision Research*, *44*, 1013–1024.
- Liversedge, S., White, S.J., Findlay, J.M., & Rayner, K. (2006). Binocular coordination of eye movements during reading. *Vision Research*, *46*, 2363–2374.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations in words. *Vision Research*, *28*, 1107–1118.
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578–586.
- McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & Psychophysics*, *25*, 221–224.
- McDonald, S. A., Carpenter, R. H. S., & Shillcock, R. C. (2005). An anatomically-constrained, stochastic model of eye movement control in reading. *Psychological Review*, *112*, 814–840.
- McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*, 648–652.
- McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*, 1735–1751.
- Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*, 551–561.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.
- Mielliet, S., & Sparrow, L. (2004). Phonological codes are assembled before word fixation: Evidence from boundary paradigm in sentence reading. *Brain and Language*, *90*, 299–310.

- Miller, B., Juhasz, B. J., & Rayner, K. (2006). The orthographic uniqueness point and eye movements during reading. *British Journal of Psychology*, *97*, 191–216.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 92–103.
- Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition*, *26*, 1313–1322.
- Morris, R. K., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 268–281.
- Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 667–682.
- Morrison, R. E., & Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. *Perception & Psychophysics*, *30*, 395–396.
- Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, *27*, 1–22.
- Myers, J. L., Cook, A. E., Kambe, G., Mason, R. A., & O'Brien, E. J. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes*, *29*, 179–200.
- Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Accessing the contribution of syntax, semantics, and plausibility in resolving ambiguities. *Language and Cognitive Processes*, *11*, 283–334.
- Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, *27*, 515–539.
- Niswander-Klement, E., & Pollatsek, A. (2006). The effects of root-frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory & Cognition*, *34*, 685–702.
- Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, *45*, 2201–2217.
- O'Brien, E. J., Raney, G. E., Albrecht, J., & Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Processes*, *23*, 1–24.
- O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 410–420.
- O'Regan, J. K. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, *25*, 501–509.



- O'Regan, J. K. (1980). The control of saccade size and fixation duration in reading: The limits of linguistic control. *Perception & Psychophysics*, 28, 112–117.
- Paterson, K. B., Liversedge, S. P., & Underwood, G. (1999). The influence of focus operators on syntactic processing of short relative clause sentences. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 52A, 717–737.
- Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 556–573.
- Pickering, M. J., & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: Lexical information and verb-argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1401–1410.
- Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequencies in the processing of unbounded dependencies. *Language and Cognitive Processes*, 18, 469–503.
- Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20, 261–290.
- Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 820–833.
- Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 148–162.
- Pollatsek, A., Raney, G. E., LaGasse, L., & Rayner, K. (1993). The use of information below fixation in reading and in visual search. *Canadian Journal of Experimental Psychology*, 47, 179–200.
- Pollatsek, A., & Rayner, K. (1982). Eye movement control in reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 817–833.
- Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception & Psychophysics*, 40, 123–130.
- Pollatsek, A., Reichle, E., & Rayner, K. (2003). Modeling eye movements in reading. In: J. Hyönä, R. Radach, H. Deubel (Eds). *The mind's eyes: Cognitive and applied aspects of eye movement research* (pp. 361–390). Amsterdam: Elsevier.
- Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, 52, 1–56.
- Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typographical errors on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 178–202.
- Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology*, 49, 151–172.

- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7, 65–81.
- Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, 4, 443–448.
- Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618–660.
- Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, 8, 21–30.
- Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In: J. M. Findlay, R. Walker, & R. W. Kentridge (Eds), *Eye movement research: Mechanisms, processes and applications* (pp. 3–12). Amsterdam: North-Holland.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E–Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 720–732.
- Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic processing during eye fixations in reading. *Canadian Journal of Psychology*, 40, 473–483.
- Rayner, K., & Bertera, J.H. (1979). Reading without a fovea. *Science*, 206, 468–469.
- Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, 41, 943–954.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–374.
- Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements, *British Journal of Psychology* (in press).
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201.
- Rayner, K., & Duffy, S. A. (1988). On-line comprehension processes and eye movements in reading. In: M. Daneman, G. E. MacKinnon, & T. G. Waller (Eds), *Reading research: Advances in theory and practice* (pp. 13–66). New York: Academic Press.
- Rayner, K., & Fischer, M. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, 58, 734–747.
- Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, 38, 1129–1144.

- Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *The Quarterly Journal of Experimental Psychology*, 39A, 657–673.
- Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 779–790.
- Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 167–179.
- Rayner, K., & Juhasz, B. J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology*, 16, 340–352.
- Rayner, K., Juhasz, B. J., Ashby, J., & Clifton, C. (2003). Inhibition of saccade return in reading. *Vision Research*, 43, 1027–1034.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). Clause wrap-up effects on eye movements during reading. *Quarterly Journal of Experimental Psychology*, 53A, 1061–1080.
- Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review*, 12, 1089–1093.
- Rayner, K., & Liversedge, S. P. (2004). Visual and linguistic processing during eye fixations in reading. In J. M. Henderson, & F. Ferreira (Eds), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 59–104). Hove, UK: Psychology Press.
- Rayner, K., Liversedge, S. P., & White, S. J. (2006). Eye movements when reading disappearing text: The importance of the word to the right of fixation. *Vision Research*, 46, 310–323.
- Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science*, 14, 385–389.
- Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements. *Vision Research*, 16, 829–837.
- Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, 12, 206–226.
- Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language*, 33, 527–544.
- Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology*, 33A, 351–373.
- Rayner, K., Pollatsek, A., & Binder, K. S. (1998). Phonological codes and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 476–497.

- Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processes. In: R. F. Lorch, & E. J. O'Brien (Eds), *Sources of coherence in reading* (pp. 9–36). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Rayner, K., & Sereno, S. C. (1994a). Eye movements in reading: Psycholinguistic studies. In: M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57–82). New York: Academic Press.
- Rayner, K., & Sereno, S. C. (1994b). Regressive eye movements and sentence parsing: On the use of regression-contingent analyses. *Memory & Cognition*, 22, 281–285.
- Rayner, K., & Sereno, S. C. (1994c). Regression-contingent analyses: A reply to Altmann. *Memory & Cognition*, 22, 291–292.
- Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line comprehension processes. *Language and Cognitive Processes*, 4, SI21–SI49
- Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1188–1200.
- Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and elderly readers. *Psychology and Aging*, (in press).
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30, 1290–1301.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509.
- Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, 31, 537–550.
- Rayner, K., White, S. J., Kambe, G., Miller, B., & Liversedge, S.P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In: J. Hyona, R. Radach, & H. Deubel (Eds), *The mind's eye: Cognitive and applied aspects of eye movements* (pp. 213–234). Oxford: Elsevier.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157.
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2006a). Modeling the effects of lexical ambiguity on eye movements during reading. In: R. P. G. von Gompel, M. H. Fischer, W. S. Murray, & R. C. Hill, (Eds), *Eye movements: A window on mind and brain*. Oxford: Elsevier.
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2006b). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7, 34–55.

- Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z reader model. *Vision Research*, 39, 4403–4411.
- Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparison to other models. *Brain and Behavioral Sciences*, 26, 445–476.
- Reilly, R., & Radach, R. (2003). Foundations of an interactive activation model of eye movement control in reading. In: J. Hyönä, R. Radach, & H. Deubel (Eds), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 429–455). Oxford: Elsevier.
- Reingold, E., & Rayner, K. (2006). Examining the word identification stages identified by the E-Z Reader model. *Psychological Science* (in press).
- Schmauder, A. R., & Egan, M. C. (1998). The influence of semantic fit on on-line sentence processing. *Memory & Cognition*, 26, 1304–1312.
- Schustack, M. W., Ehrlich, S. F., & Rayner, K. (1987). Local and global sources of contextual facilitation in reading. *Journal of Memory and Language*, 26, 322–340.
- Sereno, S. C. (1995). Resolution of lexical ambiguity: Evidence from an eye movement priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 582–595.
- Sereno, S. C., O'Donnell, P., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 335–350.
- Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science*, 3, 296–300.
- Sereno, S. C., & Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Perception & Psychophysics*, 62, 402–409.
- Sparrow, L., & Mielliet, S. (2002). Activation of phonological codes during reading: Evidence from errors detection and eye movements. *Brain and Language*, 81, 509–516.
- Speer, S. R., & Clifton, C. Jr. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition*, 26, 965–978.
- Starr, M. S., & Rayner, K. (2001). Eye movements during reading: Some current controversies. *Trends in Cognitive Science*, 5, 156–163.
- Staub, A., & Rayner, K. (2006). Eye movements and comprehension processes. In: G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.
- Sturt, P. (2003). The time course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
- Tinker, M. A. (1939). Reliability and validity of eye-movement measures of reading. *Journal of Experimental Psychology*, 19, 732–746.

- Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language*, *53*, 1–25.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*, 454–475.
- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558–592.
- Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye movements and self-paced reading. *Journal of Memory and Language*, *47*, 530–547.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, *33*, 285–318.
- van Gompel, R. P. G., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, *90*, 255–264.
- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S.P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284–307.
- van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two stage models. *Journal of Memory and Language*, *45*, 225–258.
- Vitu, F. (1991). The influence of parafoveal processing and linguistic context on the optimal landing position effect. *Perception & Psychophysics*, *50*, 58–75.
- Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vision Research*, *41*, 3513–3533.
- White, S. J., Rayner, K., & Liversedge, S.P. (2005). The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic Bulletin & Review*, *12*, 466–471.
- Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition*, *28*, 1011–1021.
- Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339.
- Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, *97*, 259–268.
- Yang, S., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision Research*, *41*, 3567–3585.
- Yang, S., & McConkie, G. W. (2004). Saccade generation during reading: Are words necessary? *European Journal of Cognitive Psychology*, *16*, 226–261.

This page intentionally left blank

## Chapter 17

### **Psycholinguistics Electrified II (1994–2005)**

*Marta Kutas, Cyma K. Van Petten, and Robert Kluender*

#### **1. INTRODUCTION**

Since the first edition of the *Handbook of Psycholinguistics* in 1994, investigations of language processing via electromagnetic recordings have proliferated beyond the possibility of coverage in any single chapter. Our aim here is to offer a sampling of the more seminal, influential, and controversial event-related brain potential (ERP) studies within the psychology of language, focusing on the last decade. Out of necessity, we restrict the review to studies of healthy young adults, as this segment of the population is the typical baseline against which to assess results from infants, children, middle-aged and older adults, and individuals with neurological or psychiatric disorders. Length limitations also forced us to skip studies of speech perception and production, and those bearing on the automaticity of semantic processing, topics we plan to address in some future venue.

In 1994, there were only two dominant noninvasive techniques to offer insight about the functional organization of language from its brain bases: the behavior of brain-damaged patients (neuropsychology), and ERPs. Positron emission tomographic and magnetoencephalographic (MEG) measures were just beginning to contribute to our understanding. Over the ensuing decade plus, these have been joined by functional magnetic resonance imaging, transcranial magnetic stimulation, event related spectral changes in the electroencephalogram (EEG), and noninvasive optical imaging (see Gratton & Fabiani, 2001; Gratton, Fabiani, Elbert, & Rockstroh, 2003, for review of the last and new technique). As outlined below, three of these methods are closely related in their neural and physical bases: ERPs, event-related frequency changes in the EEG and MEG. After that brief review of the neural bases of these methods, we devote a modicum of attention to the latter two methods and chiefly focus on ERP studies of language processing. The remainder of the review is then devoted to four major domains of language processing: visual word recognition, basic semantic processing, higher-level semantic processing, and syntax and morphology.



## 2. ELECTROMAGNETIC MEASURES OF BRAIN ACTIVITY

### 2.1. Neural Activity and the Electroencephalogram

Interactions between neurons are the essence of brain activity. These interactions consist of current flow – the movement of charged ions – across cell membranes, such that the direction and magnitude of current flow in one neuron depends on the neurons it communicates with. A recording electrode close to a neuron can detect one sort of rapid change in voltage (or potential) caused by rapid changes in current flow: the action potential that causes neurotransmitter release in the vicinity of another neuron. Placing an electrode close to a single neuron is too invasive for use in healthy humans. After neurotransmitter is released and bound by other neurons, the result is a change in current flow across the membranes of those other post-synaptic cells. These small changes in current flow can sum in different ways depending upon the number, location, and timing of active synapses on the target neuron, as well as on the inward or outward directions of current flow. The target neuron may fire its own action potentials, reduce its firing rate, or show no change in firing rate but become more or less responsive to future inputs. An immediate change in activity due to synaptic input may also be accompanied by changes in gene transcription resulting in long-lasting structural modifications of the neuron. Whatever the short or long-term outcome, current flow around synapses *is* the currency of neural communication.

Like action potentials, the small changes in voltage around active synapses can be recorded by nearby electrodes. The summed activity of many synapses on many neighboring neurons (called a *field* potential) can also be recorded by a pair of electrodes – one placed directly in neural tissue and one some distance away. Perhaps surprisingly, summated synaptic potentials can also be recorded outside the head, noninvasively from electrodes placed on the scalp; this record of fluctuating voltage across time is the *EEG*. The amplitude of the EEG is considerably smaller than invasively recorded field potentials because the skull is a strong electrical insulator. Like field potentials, the amplitude and polarity of the EEG depend on the number and amplitude of the contributing synaptic potentials, on whether current is flowing into or out of cells (i.e., movement of positive or negative ions, excitatory or inhibitory synaptic potentials), and on the geometric relationship between the synapses and electrode (i.e., current flow toward versus away from the electrode, or both toward and away, which will lead to cancellation of the opposing signals; Nunez, 1981). Finally, any record of electrical potential (voltage) consists of the difference between two locations (like the positive and negative poles of a battery), so that the polarity and spatial distribution of the EEG across the head depend on what pairs of sites are chosen. Most typically, a single location or pair of locations that are somewhat more insulated from brain activity – such as that provided by the thick bones behind the ears for mastoid or earlobe sites, or the air-filled sinuses for the nosetip are used for *reference* for each scalp site, although other references are possible. Given the low electrical conductivity of the skull, electrical potentials recorded from the scalp must reflect the activity of large numbers of neurons, estimated 1000 to 10,000 for the smallest signals recorded. Cortical pyramidal cells are likely to dominate the EEG signal, because they are the largest and most numerous cell type, and because their dendritic

processes are spatially parallel to their neighbors; such an organization leads to summation of the small electrical fields generated by each active synapse.

## 2.2. Event-Related Brain Potentials

The main emphasis of this chapter, however, is on electrical or magnetic brain activity that is synchronized to some external event (i.e., an ERP). At the scalp an ERP (5–10  $\mu\text{V}$ ) is substantially smaller in amplitude than the background EEG (50–100  $\mu\text{V}$ ) and is, therefore, generally extracted by computer averaging. This involves recording ERPs to repeated presentations of conceptually, if not physically, similar stimuli. Voltage fluctuations generated by neurons, unrelated (or at least not phase locked) to the processing of the stimuli of interest will be random with respect to stimulus onset time and thus cancel each other, leaving a record of event-related activity. The number of stimuli needed for a reliable average is a function of the amplitude of the ERP component and the question under study: the smaller the component, the more trials that are needed to extract it from the spontaneous EEG (“noise”).

The major statistical assumption in averaging is that the signal is indeed time locked to the averaging trigger whereas the “noise” is not. For the early “sensory” portion of the ERP, the time-locking assumption is well supported. In the case of later portions of the ERP which are instead elicited by higher-level “cognitive” analyses of the stimulus, the latency of the signal may not be invariant with regard to stimulus onset on a trial by trial basis, but there are techniques to correct for misalignment (see, e.g., Handy, 2005).

### 2.2.1. Peaks and components

The ERP waveform of voltage plotted against post-stimulus time consists of a series of positive and negative peaks; these are typically compared to a pre-stimulus baseline, which is a short (100–200 ms) record of activity (or preferably inactivity) immediately preceding each experimental stimulus, although other baselines are possible. Voltages are thus only negative or positive with respect to the baseline.

The ERP peaks are typically labeled according to their polarity (negative [N] or positive [P]) and latency in milliseconds relative to stimulus onset (e.g., N100, P230, P300). Occasionally, peaks are designated by their polarity and ordinal position in the waveform (e.g., N1, P1, N2). Sometimes, the labels denote a functional description (e.g., mismatch negativity or MMN) or refer to its presumed neural generator (e.g., auditory brainstem response) or its most reliable scalp location (e.g., LAN or left anterior negativity). The mix of descriptive and functional labels brings us to the distinction between an ERP peak, readily observed by the eye, and the more abstract concept of a “component” (see Allison, Wood, & McCarthy, 1986; Donchin, Ritter, & McCallum, 1978).

The underlying notion of a “component” is clear: the processing of any external stimulus occurs over time, so that different parts of the nervous system with different functions

are likely involved at different time points. The ERP is a record of this neural processing, so that different temporal intervals of the waveform are likely to reflect different anatomical locations and different functional processes, although any particular interval may reflect more than one brain region/functional process. One set of factors, visible in a single ERP waveform, bears some, usually unknown, relationship to the anatomy of the underlying neural generators: polarity, latency from stimulus onset, and relative amplitude across scalp locations (i.e., scalp distribution). Most commonly, the ERP waveform is reduced to a series of peak or mean amplitude measurements relative to a pre-stimulus baseline. There also exists a number of algorithms for decomposing the ERP waveform into some weighted average of subcomponents, which may provide truer reflections of the neural components responsible for particular component psychological processes. Two such decomposition techniques are independent component analysis (ICA) (Delorme & Makeig, 2004; Makeig, Bell, Jung, & Sejnowski, 1996) and spatial or spatio-temporal principal component analysis (PCA) (Beauducel & Debener, 2003; Spencer, Dien, & Donchin, 2001).

The second set of factors critical for identifying some portion of the ERP as a unitary component involves comparisons between two or more experimental conditions to determine which manipulations influenced a particular temporal region of the waveform. Susceptibility to some experimental manipulation is essential for component identification, making “peak” or for that matter ICA or PCA component non-synonymous with “component”. The functional characterization offered by psychologists and the neural characterization that might be offered by a physiologist are thus all part of the definition of an ERP component, under ideal circumstances. However, circumstances are rarely ideal. A functional characterization is most easily carried out via experiments involving large numbers of healthy human subjects, whereas a neural characterization typically requires converging evidence from animal models, neurological patients undergoing invasive clinical procedures, and scalp recordings from patients with defined brain damage (Arezzo, Vaughan, Kraut, Steinschneider, & Legatt, 1986; Buchwald & Squires, 1982; Halgren, 1990; Knight, Scabini, Woods, & Clayworth, 1989; McCarthy, Wood, Williamson, & Spencer, 1989; Pineda, Swick, & Foote, 1991). In this review, we focus on functional characterizations, and discuss psycholinguistic manipulations implemented by varying the stimuli and/or the instructions to participants, referring to anatomical generators when known.

As a general rule, the amplitudes, latencies, and scalp distributions of the earlier ERP components (with latencies <100 ms) are highly reproducible across sessions within an individual (Halliday, 1982). Moreover, systematic variations in the physical parameters of the evoking stimulus (e.g., intensity, frequency, duration) lead to predictable changes in these early components reflecting the altered activation of sensory pathways. Hence, the earlier evoked components are considered to be “exogenous” or stimulus bound; they are generally relatively impervious to an individual’s state of alertness or attentiveness. This invariance in the face of changing psychological states makes them an excellent diagnostic tool for certain sensory and neurological disorders (Chiappa, 1983).

For psycholinguistic purposes, the more informative brain waves are the so-called endogenous components, which may precede or follow a triggering event by hundreds of milliseconds. An “event” in this case refers to a stimulus, a response, a voluntary movement, or a cognitive operation for which an external timing marker can be specified. The relative (although not total) insensitivity of endogenous components to variations in the physical stimulus parameters contrasts with their exquisite responsivity to task demands, instructions, and subjects’ intentions, decisions, expectancies, strategies, mental set, and so on. In other words, endogenous ERP components are not “evoked” by a stimulus but are elicited by the perceptual and cognitive operations that are engendered by that stimulus. The same physical stimulus may or may not be followed by a particular endogenous component depending on how the subject chooses to process it. The term “late” component is often used interchangeably with “endogenous” component because most of these potentials occur with a latency beyond 100 ms, although some earlier potentials can be modulated by cognitive processes.

### 2.3. Magnetoencephalography and Event-Related Magnetic Fields

Current flow in the brain produces small magnetic fields in addition to the voltage fields recorded as EEG. Epochs of the MEG following stimulus presentation can be averaged to derive the event-related magnetic field. Although both the raw MEG and the event-related fields resemble their electrical counterparts in many ways, some physical differences make the anatomical origins of the magnetic signals easier to localize (while preserving the same temporal resolution as electrical signals; for review see Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). One reason is that although the skull is a very good electrical insulator and thus imposes a spatial blurring between the brain and the scalp, bone is magnetically transparent. The magnetic fields recorded just outside the head are also more strongly influenced by the geometrical orientation of intracranial current flow. The latter fact is a mixed blessing. On the one hand, the convoluted shape of the cortex ensures that current flow in different sulci and gyri will have distinctive orientations; this is useful in modeling the location of tissue responsible for a magnetic field (particularly when combined with structural magnetic resonance scans showing the gyral/sulcal pattern of each subject). On the other hand, only current flow that is at least somewhat tangential to the surface of the head will produce a detectable magnetic field. Thus it is primarily cortical activity in sulci, rather than in gyri (where the pyramidal cells are oriented perpendicular to the skull) that can be detected. This is only a minor limitation as it is estimated that two-thirds of the cortical sheet lies in sulci (Armstrong, Schleicher, Omran, Curtis, & Zilles, 1995). Finally, MEG is less sensitive to cortical sources located far away from the scalp, because the magnetic signal shows a steeper decline with increasing distance between neural source and external sensor. Overall, MEG presents perhaps the best combination of spatial and temporal resolution of noninvasive methods in common use. However, MEG studies are not very common, because the recording devices (*SQUID*, superconducting quantum inference device) are expensive and, to date, not as widely supported by routine clinical applications as magnetic resonance scanners are.

### 3. VISUAL WORD RECOGNITION

Any reader must first classify visual inputs as linguistic (e.g., letters, words) rather than as non-linguistic objects and then further categorize letter strings as pronounceable or not, meaningful or not, etc. The spatiotemporal dynamics of visual processing have been investigated via scalp and intracranial ERPs and MEG. While the exact timing details vary somewhat because of methodological and analytic differences, results overall suggest that visual responses become increasingly selective for classes of visual stimuli over time, with an especially critical role for left inferior occipito-temporal areas in visual word processing.

#### 3.1. Intracranial Data: Orthographic and Non-orthographic Stimuli

Allison and colleagues recorded evoked potentials directly from the cortical surface to a variety of visual stimuli — sinusoidal gratings, pictures of faces, word and nonword letter strings, number strings, and animate (butterflies) and inanimate (cars) objects, etc. — in a large number of patients with intractable seizures (Allison, McCarthy, Nobre, Puce, & Belger, 1994; Allison, Puce, Spencer, & McCarthy, 1999). The earliest activity in the visual cortex (V1 and V2) was evidenced by N100 and P100 components which are sensitive to luminance, luminance contrast, and stimulus size, but not to stimulus category. Other relatively early responses in posterior visual cortex were sensitive to sinusoidal gratings (P120-N180-P260). Approximately 20–30 ms later, category-specific activations were observed in more ventral areas (e.g., P150-N200-N290-N700). For instance, several different cortical patches within extrastriate cortex generated surface negativities with peak latencies around 200 ms (N200), but different patches were specific to faces, objects, or letter strings (Allison et al., 1994; Nobre, Allison, & McCarthy, 1994). Letter-string specific areas in the posterior fusiform gyrus responded equally to words and nonwords, whereas the anterior fusiform gyrus was sensitive to properties of letter strings (Nobre et al., 1994). On the surface of the anterior fusiform gyrus, bilaterally, a P400 component was specific to real words; a potential of the opposite polarity (N400) was observed just superior to this region, indicating locally generated activity (McCarthy et al., 1995). Overall, depth recordings reflect the segregation of the ventral object recognition system into functionally discrete regions.

#### 3.2. Scalp-Recorded and MEG Data: Orthographic and Non-orthographic Stimuli

Schendan, Ganis, and Kutas (1998) compared ERPs to object- and word-like stimuli as shown in Figure 1. Regardless of assigned task, a negative peak at around 95 ms (N100) over midline occipital sites was smaller for single object-like stimuli than for any variety of “string” stimuli. This distinction was quickly followed (~10 ms later) by a differentiation between strings of real letters (words and pseudowords) versus those of non-letter characters (icon strings, pseudo-font strings). The first sign of specialized processing of “linguistic” stimuli in the scalp record thus appeared around 105 ms, perhaps reflecting the experience-based tuning of the visual system to rapidly

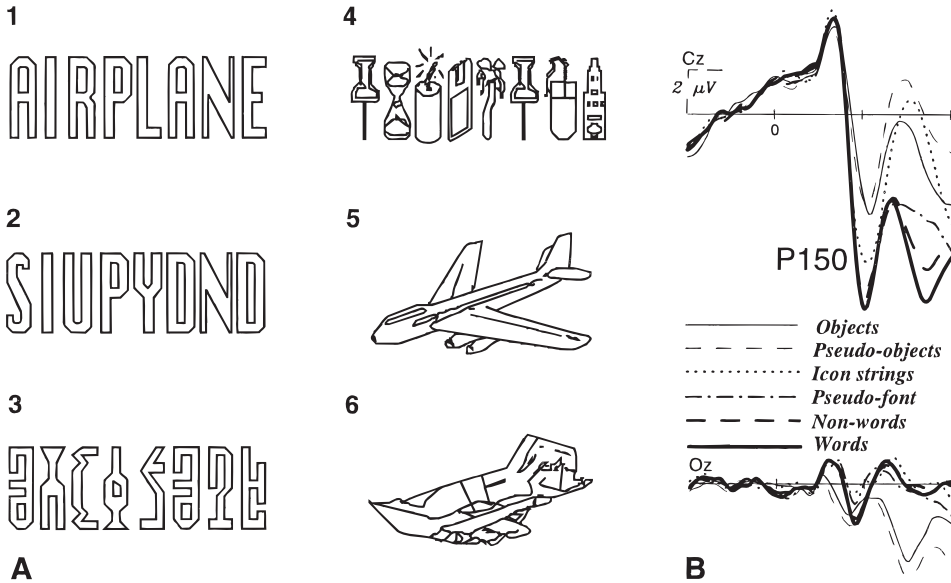


Figure 1. (A) Examples of the six different image types presented to subjects in a random sequence: (1) words, (2) nonwords, (3) pseudofont strings, (4) icon strings, (5) real objects, and (6) pseudo-objects. (B) First 350 ms of the grand average ERP to these stimuli from midline central (Cz) and occipital (Oz) sites. Note that P150 amplitude differentiates word-like (or letter stringlike) images from non-linguistic object-like stimuli. Reprinted with permission from Schendan, Ganis, and Kutas, (1998).

detect physical stimuli with real letter properties. Around 200 ms, word ERPs were distinguishable from those to random letter strings. This ERP difference alone, however, does not warrant the conclusion that the brain has identified one type of stimulus as a word and the other as not a word, given that the stimuli also differ in amount of prior exposure (recency, or frequency of constituent letters, bigrams, entire strings, etc.).

Bentin, Mouchetant-Rostaing, Giard, Echallier, and Pernier (1999) examined ERPs to orthographic (words, pseudowords, consonant letter strings) and non-orthographic stimuli (alphanumeric strings, strings of forms) in various oddball tasks. Comparisons were between the frequent, non-target stimuli across different tasks, designed to induce different levels of analysis (visual, phonological/phonetic, phonological/lexical, and semantic). The earliest electrophysiological distinctions arose in the visual size-judgment task in which the occipito-temporal N170 distinguished orthographic from non-orthographic stimuli: the N170 was reliably larger over the left than the right hemisphere for orthographic stimuli and marginally reversed for non-orthographic stimuli. This scalp potential resembles the intracranial N200 component elicited by all sorts of visual stimuli, albeit with different non-overlapping distributions within posterior fusiform gyrus for orthographic (words, pseudowords, nonwords) versus

non-orthographic (faces) stimuli (Allison et al., 1994); the N200 for words is more left lateralized while that for faces is bilateral or right lateralized.

The next reliable ERP difference – an N320 larger over left hemisphere sites – distinguished pronounceable (words, pseudowords) letter strings from unpronounceable consonant strings. Simon, Bernard, Largy, Lalonde, and Rebai (2004) found that this component was modulated by word frequency and eliminated by massive repetition. Soon thereafter, an N350, similar to the N320 albeit with a wider scalp distribution including temporo-parietal areas, distinguished phonologically legal from phonologically illegal letter strings. Finally, an N450 similar to the N350 but extending to fronto-central areas, distinguished words from pseudowords and pseudowords from consonant strings in a task aimed at inducing semantic processing by asking participants to respond to all abstract (as opposed to concrete) words and pseudowords.

Similar results have been obtained in MEG studies comparing letter strings of various lengths to symbol strings of equivalent lengths to letter-like symbols (rotated letters) embedded in varying amounts of Gaussian noise. The earliest responses around 100 ms in midline occipital cortex are modulated by visual noise, vary with string length, and more generally increase with visual complexity; this so-called Type I response has been linked to low-level visual analyses such as extraction of nonspecific image properties – perhaps contrast borders (Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999; Tarkiainen, Cornelissen, & Salmelin, 2002). The first distinction between the processing of letter or letter-like strings versus symbol strings occurs around 150 ms over left inferior occipito-temporal regions (M170), with greater activity for letter strings (Tarkiainen et al., 1999). Like the electrical N170 and intracranially recorded N200, the M170 does not distinguish among words, pseudowords, and consonant strings (Salmelin, Service, Kiesila, Uutela, & Salonen, 1996). Coincident activity over right occipital-temporal regions is modulated by string length but shows no preference for letter strings. Subsequent to the M170, various word stimuli elicit an M250, sensitive to phonotactic probability (Pykkänen, Stringfellow, & Marantz, 2002); and an M350, sensitive to lexical frequency (Embick, Hackl, Schaeffer, Kelepir, & Marantz, 2001). The latencies of both the M250 and M350 vary with phonotactic probability but not neighborhood density (Stockall, Stringfellow, & Marantz 2004), although neighborhood effects interact with probability and density around the region of the M350.

In scalp recordings, orthographically legal, pronounceable pseudowords elicit ERPs that are qualitatively similar to words for several hundred milliseconds (up to ~450 ms), although ERP amplitudes typically differ. Unlike orthographically illegal, unpronounceable nonwords, but similar to real words, pronounceable pseudowords elicit an N400. With no surrounding context, the amplitude of the N400 to pseudowords may be about the same as that to real words with low usage frequency (this has not been carefully investigated). The N400 is not believed to be identical to the M350. The brain thus seems to deal with pseudowords, which although potentially meaningful, have no particular, learned meaning, no differently than real words for a considerable period after words have been differentiated from nonwords, suggesting that this early distinction may reflect

differential amounts of prior exposure. Indeed, at about the same time that pseudowords are differentiated from real words, both frequency and repetition effects are observed for real words, suggesting that amount of prior exposure is one critical factor. All written words, for example, elicit a negativity (220–400 ms) over left anterior scalp (lexical processing negativity or LPN) whose peak latency varies with the eliciting word's frequency of usage (King & Kutas, 1998)

#### 4. SEMANTIC PROCESSING AND THE N400 COMPONENT

Of the large number of ERP components sensitive to language processes, the N400 is the best used to date. The label “N400” refers to a negative-going voltage in the ERP peaking in amplitude around 400 ms after stimulus onset. This component was first noted by Kutas and Hillyard (1980a, 1980b) in a comparison of sentence-final words that formed predictable completions and those that were semantically improbable or incongruent (left column, Figure 2). While predictable endings elicited a broad positive waveform from 200 to 600 ms, the incongruent words elicited a large negative wave in this time range. It is important to note that the N400 semantic context effect – the difference between the two conditions – extends over some period of time. Labeling experimental effects by the latencies of their peaks is conventional in ERP research, but ERP activity always has a temporal duration (as do single neuron responses). An onset latency ~200 ms is typical of semantic context effects for visual words; context effects on spoken words typically begin somewhat earlier (as early as 50 ms in natural speech, because of coarticulatory information from the previous word, or 150 ms when the eliciting words are recorded separately and spliced into the speech stream).

Across the 1980s, it became clear that the *N400 sentence congruity effect* was only one indication of a much broader sensitivity to semantic context, and there was nothing special about anomalous completions. First, even congruent sentence completions elicited N400s whose amplitude was directly (inversely) proportional to the goodness-of-fit between the sentence frame and the eliciting word, as indexed by offline cloze probability (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). Second, context effects were observed in word pairs: the second words of semantically related word pairs elicited smaller N400s than those of unrelated pairs (right column, Figure 2; Bentin, McCarthy, & Wood, 1985; Boddy, 1981). Third, examination of the ERPs elicited by the intermediate words of sentences presented one at a time in serial order showed large N400s for the first open-class words, which became progressively smaller as the sentence (specifically semantic) context built up and constrained subsequent words (Van Petten & Kutas, 1990, 1991; Van Petten, 1989, 1993). This sentence-position effect on N400 amplitude was observed only in isolated sentences for which readers had no prior inkling of the sentence topic and not for sentences in discourse, which did not introduce completely new topics (Van Petten, 1995). Also in the 1980s, it was shown that spoken words and signs in American Sign Language elicited N400s that were qualitatively similar to their visual counterparts, and also reduced in amplitude by supportive semantic context (Kutas, Neville, & Holcomb, 1987; Holcomb & Neville, 1990, 1991; Neville, Mills, & Lawson,



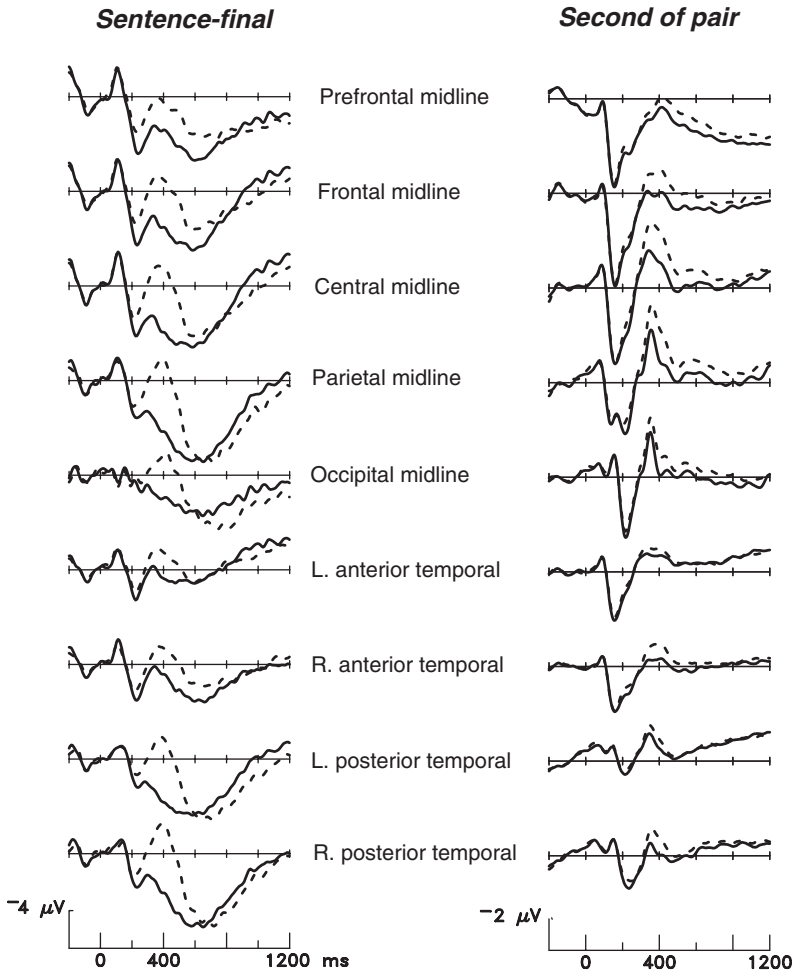


Figure 2. Left: grand average visual ERPs ( $N=16$ ) elicited by high-cloze congruent and incongruent word completions of sentences read one word at a time, followed by a delayed "sensible and grammatical" judgment. Unpublished data from Groppe and Kutas. Right column: grand average ERPs ( $N=30$ ) to the second words of semantically related and unrelated visual word pairs. Participants performed a delayed letter search task (decide if a letter presented after the second word was present in either word of the pair). Unpublished data from Luka and Van Petten. Note that for both data sets decision-related potentials were postponed beyond the epochs shown here. In contrast to the predictable sentence completions, even related words elicit substantial N400 activity. Also note the right greater than left asymmetry of both N400 context effects.

1992). The correct characterization of the N400 context effect is thus not that anomalous or unrelated words elicit unusual brain responses, but rather that a large negativity between 200 and 500 ms or so (N400) is the default response, and that its amplitude is reduced to the degree that context aids in the interpretation of a potentially meaningful

stimulus. Finally, as detailed in Section 4.2 the amplitude in the N400 region of the ERP is sensitive not only to the context surrounding a word, but also to the lexical characteristics of the eliciting words themselves.

N400-like potentials are also evident in response to other meaningful stimuli – line drawings, photos, and environmental sounds – and also reduced in amplitude when these nonverbal stimuli are preceded by conceptually related stimuli (Ganis, Kutas, & Sereno, 1996; Holcomb & McPherson, 1994; Plante, Van Petten, & Senkfor, 2000; Van Petten & Rieffers, 1995). These closely resemble the verbal N400 in wave-shape and timing but have slightly different spatial distributions across the scalp. The conceptual context effect for pictorial materials has shown a more anterior maximum than the analogous effect for printed words, and the effect for meaningful non-linguistic sounds has a small lateral asymmetry that is opposite that for spoken words (left greater than right for sounds, right greater than left for words). These data suggest that verbal and nonverbal N400s reflect similar cortical computations occurring in different, but overlapping, populations of neurons. Overall, the extant data suggest that N400 amplitude is a general index of the ease or difficulty of retrieving stored conceptual knowledge associated with a word (or another meaningful stimuli), which is dependent on both the stored representation itself, and the retrieval cues provided by the preceding context.

#### **4.1. Neural Bases of the N400**

There is no single, completely direct path from scalp-recorded ERPs to certain knowledge of their neural generators, but instead multiple methods that contribute to this knowledge. These include scalp recordings from patients with brain damage in known locations, intracranially recorded ERPs in patients with electrodes implanted prior to surgery for the relief of epilepsy, and recordings of evoked magnetic fields. Application of these three methods implicate the left temporal lobe as the largest source of the scalp N400, with a substantial but lesser contribution from the right temporal lobe (see Van Petten & Luka, 2006).

In split-brain patients, stimuli presented to hemispheres that also have productive speech capability elicit N400 context effects, whereas stimuli presented to a mute hemisphere do not. Because speech production is strongly dominated by the left hemisphere in most neurologically intact individuals, this suggests that the N400 is more dependent on left than right hemisphere processes (Kutas, Hillyard, & Gazzaniga, 1988). Large amplitude reductions and delayed latencies of the N400 semantic context effect are observed in patients after strokes in the left temporal lobe or temporo-parietal junction – broadly, the same regions leading to an aphasic syndrome marked by a semantic comprehension deficit (Friederici, Hahne, & von Cramon, 1998; Hagoort, Brown, & Swaab, 1996; Swaab, Brown, & Hagoort, 1997). Indeed, there is a close correspondence between magnitudes of N400 effects and standardized comprehension test scores in these aphasic patients (Kojima & Kaga, 2003; Marchand, D’Arcy, & Connolly, 2002). Lesions in the perisylvian region of the right hemisphere lead to smaller-than-normal N400 context effects, albeit

with more modest reductions than following similar left hemisphere lesions (Kotz & Friederici, 2003). In contrast to the severe impact of left temporal and inferior parietal damage on the scalp-recorded N400, patients with damage restricted to the frontal lobe have normal N400 effects, and may even show N400 effects to syntactic violations that elicit different components in normal controls (Friederici, von Cramon, & Kotz, 1999; Hagoort, Wassenaar, & Brown, 2003a; Swick, Kutas, & Knight, 1998; see also Swick, 2004). Thus, although the frontal lobe is critical for many aspects of language processing, it makes little direct contribution to the N400.

Electrodes placed directly on the cortical surface, or within the depths of the cortex, are used in patients being evaluated for possible surgical relief of seizures resistant to drug treatment. The potentials recorded from these electrodes have the same neurophysiological basis in synaptic activity as scalp-recorded ERPs, but can show large amplitude gradients within a distance of a few centimeters. A potential recorded in the anterior medial part of the temporal lobe (anterior to the hippocampus, in the vicinity of the collateral sulcus dividing the fusiform gyrus from the parahippocampal gyrus) has the same time course as the scalp-recorded N400, and is sensitive to the same experimental manipulations (Nobre et al., 1994; Nobre & McCarthy, 1995). Other research groups have reported that what appears to be the same ERP component, in the same location, is reduced by repetition of words and line drawings (Elger et al., 1997; Fernández et al., 2001; Guillem, N'Kaoua, Rougier, & Claverie, 1996; Smith, Stapleton, & Halgren, 1986). The anterior medial part of the temporal lobe is the core brain region affected in a neurodegenerative disease known as *semantic dementia*, in which patients suffer a progressive loss of semantic knowledge with relative preservation of phonology, syntax, and recent episodic memory (Mummery et al., 2000; Patterson & Hodges, 2000). The convergence between the intracranial recordings and the neuropsychological data are a very strong indication that this brain region is critical for access to semantic memory, and almost certainly contributes to the scalp-recorded N400 activity.

MEG studies using dipole models all suggest sources of the N400m in the left superior and/or middle temporal gyri, with a more individually variable source in the homologous right hemisphere region (Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998; Helenius et al., 2002; Kwon et al., 2005; Simos, Basile, & Papanicolaou, 1997). Halgren et al. (2002) applied a distributed source modeling method showing the spatial extent of cortical activity, and found that most of the left temporal lobe (including inferior and anterior regions) was more active for incongruent than congruent sentence completions, with additional activity in the right anterior temporal lobe.

Overall, the neuropsychological, intracranial, and MEG results converge to suggest that both temporal lobes are responsible for the scalp-recorded N400 component, but that the left hemisphere makes a larger contribution than the right. Comparisons between language modalities (spoken, written, signed), between literal and non-literal language, and between conceptual relationships expressed by words versus nonverbal stimuli await further research.

## 4.2. Lexical Factors

### 4.2.1. *Words, pseudowords, nonwords*

When letter strings are presented in lists or pairs, words that are unrepeated, semantically unrelated to previous words, and low in frequency elicit very large N400s, as do orthographically legal, pronounceable nonwords (*pseudowords*).<sup>1</sup> By contrast, unpronounceable nonwords elicit little or no N400 activity (Anderson & Holcomb, 1995; Bentin, McCarthy, & Wood, 1985; Chwilla, Brown, & Hagoort, 1995; Holcomb & Neville, 1990; Rugg & Nagy, 1987; Smith & Halgren, 1987; Ziegler, Besson, Jacobs, & Carr, 1997). Likewise, illegal nonwords do not yield reliable incidental repetition effects (Rugg & Nagy, 1987). By contrast, immediately repeated pseudowords do elicit an ERP repetition effect smaller than that observed for real words (Doyle, Rugg, & Wells, 1996), although the contribution of the N400 and a distinct ERP component related to episodic memory retrieval to the pseudoword repetition effect have not been disentangled (see Olichney et al., 2000). However, pseudowords derived from real words seem to show characteristic ERP concreteness effects (Kounios & Holcomb, 1994), suggesting that pseudowords that closely resemble real words may contact semantic memory.

Recently, Deacon et al. examined both repetition of words (TRAIN–TRAIN) and pseudowords (WOLM–WOLM), as well as semantically related pairs of words (TULIP–ROSE) and pseudowords derived from related words (PLYNT–TLEE from PLANT–TREE). The task was a delayed decision on a probe following each pair (Deacon, Dynowska, Ritter, & Grose-Fifer, 2004). ERPs to both words and pseudowords were modulated by repetition, as in previous studies. More interestingly, the “related” pseudowords elicited smaller N400s than unrelated pseudowords (derived from a pair of unrelated words). This finding is similar but different from a behavioral study appearing about the same time: Perea and Lupker (2003) reported that masked pseudowords created by letter transposition (JUGDE) led to faster lexical decision times for related words (COURT), but that letter-replaced pseudowords (more like Deacon’s derived pseudowords, e.g., JUDPE) did not. The differential results may be due to the different dependent measures, or differences between masked and unmasked contexts. Forster and Hector (2002) also reported slower reaction times to reject (unmasked) derived pseudowords (e.g., TURPLE) during a semantic categorization task. At first blush, all of these studies suggest that pseudoword processing activates semantic representations of (at least some) orthographically similar

---

<sup>1</sup> This standard result in adults fluent in a language differs from that of children acquiring their first language. Words known to 13–20-month olds elicit larger N400s than both unknown real words and pseudowords (Mills, Coffey-Corina, & Neville, 1997; Mills et al., 2004). When infants learn associations between pseudowords and novel objects, N400 amplitude increases (Mills, Plunkett, Prat, & Schafer, 2005). These developmental results indicate that the neural systems for word recognition and retrieval of meaning function somewhat differently in the immature state, before many words have been acquired and their meanings organized in an adult manner. Infant N400 effects resemble adult N400s over posterior sites, but are accompanied by large frontal negativities atypical of adult comprehension studies. Friedrich and Friederici (2004) similarly observed that unrelated picture–word pairs elicit more frontal N400 context effects in 19-month olds than in adults. Overall, frontal cortex may be more critically involved in word comprehension during first-language learning than adulthood.

words. However, Deacon et al. drew a subtly different conclusion, based on an experiment in which repetition effects were also observed for pseudowords not obviously derived from real words (nonderived pseudowords). Deacon et al. argued that N400s to pseudowords — and repetition effects in general — are unlikely to be due to semantic activation *per se*, because nonderived pseudowords are not likely to activate word representations in the mental lexicon. Instead, they concluded that “N400 appears to be generated by orthographic/phonological analysis and is attenuated by the top–down feedback of semantic information to the orthographic/phonological level” (p. 60).

We are more inclined to view N400 activity as arising from the semantic system itself, but it is nontrivial to distinguish this account from Deacon et al.’s (2004) feedback account on the basis of existing data. However, we believe that the starting assumption of their argument — that nonderived but legal pseudowords do not activate any semantic representations — is speculative. Recent behavioral work suggests that the word recognition system may be remarkably tolerant of mismatches between the actual input and real words. In her dissertation work, Guerrero (2004) observed that scrambled masked primes still produced repetition effects in lexical decision (e.g., SIEDAWLK or SDIWELAK both speeded (RTs) to target SIDEWALK). This work did not address semantic activation, but suggests caution in assuming that a letter string is too distant from a real word to contact at least some aspect of its representation in memory.

#### 4.2.2. *Vocabulary class*

Kutas and Hillyard (1983) first noted that *open-class* or “content” words (nouns, verbs, most adjectives, -ly adverbs) elicited different ERPs than *closed-class* or “function” words (pronouns, articles, conjunctions, prepositions, etc) in sentences. Subsequent experiments have attempted to determine which aspects of the vocabulary distinction — word length, frequency of usage, repetition, contextual constraint, abstractness of meaning, referentiality, syntactic role, etc. — are responsible for these differences.

Closed-class words in sentences typically elicit smaller N400s than open-class words. Van Petten and Kutas (1991) suggested that this may reflect the converging influences of higher frequency of usage, higher repetition rate, and greater predictability of closed-class items within sentences. And, indeed, each of these factors has been found to modulate the N400. When closed-class words are contextually unexpected in a sentence context, they too can elicit sizeable N400s (King and Kutas, 1995). However, the semantic content of the eliciting words may also be relevant. “Wh-words” such as “who” or “what” elicit larger N400s than do complementizers such as “that” (Kluender & Kutas, 1993a). Likewise, in the context of a sentence (“I wonder whether the candidate was annoyed THAT/WHEN ...”), WHEN was found to be associated with a larger N400 than THAT (McKinnon & Osterhout, 1996). These effects most likely reflect differences in referential specificity, the richness of the information retrieved from semantic memory. In line with this suggestion, Münte et al. (2001) found

that N400 amplitudes to (German) closed-class items in a lexical decision task were smaller than those to open-class items even after frequency matching.

Another reliable difference between open- and closed-class words is a late ramp-shaped negativity over frontal scalp sites, called the *N400–700*, which is larger for closed-class items (Van Petten & Kutas, 1991; Neville, Mills, & Lawson, 1992; Osterhout, Bersick, & McKinnon, 1997; King & Kutas, 1998; Brown, Hagoort, & Keusters, 1999; Münte et al., 2001). Van Petten and Kutas first observed variability in this component when comparing closed-class words in random word strings, syntactically legal but semantically anomalous sentences, and congruent sentences. The frontal N400–700 proved sensitive to both sentence type and word position. Early in a sentence, the N400–700 was essentially absent in all three conditions, but grew in amplitude over the course of congruent sentences. As the N400–700 also had not been observed when open- and closed-class words appeared in a lexical decision task (Garnsey, 1985), we suggested that it might be a member of a family of slow negative-going potentials – known as the contingent negative variation (CNV) – typically seen between a warning and an imperative stimulus that an individual actively anticipates or prepares for (McCallum, Curry & Hutch, 1993). In the case of sentences, closed-class items presumably serve as a syntactic signal (warning) that a new head of a constituent is imminent. Brown et al. (1999) also identified this negativity to closed-class items as a CNV but speculated that it warns the reader that the next word is likely to be meaningful. Either of these functional interpretations is consistent with the finding that the N400–700 is significantly smaller in patients with Broca's aphasia (Keusters, Brown, Hagoort, & Stegeman, 1999). Münte et al. (2001), however, observed N400–700 effects not only during a sentence reading task but also with word lists (lexical decision task), although only for very high frequency closed-class items (mostly determiners). They thus proposed a modified version of our hypothesis limited to determiners. As they noted, testing this hypothesis would require comparing ERPs to various types of closed-class items (determiners, conjunctions, prepositions), matched on critical variables such as length, frequency, and sentence position.

These ERP differences in the N400, N400–700, and LPN have been used to argue both for and against a qualitative and neural distinction for open-versus closed-class words. People generally agree that the N400 difference is quantitative rather than qualitative (although see Neville et al., 1992).

The N400–700 is a somewhat more plausible candidate for a vocabulary-class marker in that open-class words never elicit N400–700s as large as closed-class. However, this component also reflects the contextual milieu in which closed-class words appear and may reflect the typical functional role of such words in parsing, rather than pure representational differences between the word classes.

Finally, recent results indicate that the ERP is also sensitive to divisions within open-class words such as that between nouns and verbs, albeit differently in word lists and sentences (Federmeier, Segal, Lombroso, & Kutas, 2000; Rösler, Streb, & Hahn, 2001; Khader, Scherag, Streb, & Rösler, 2003).

#### 4.2.3. *Word frequency*

Word frequency refers to an individual's life history of encounters with a particular word (estimated from normative frequency counts). High-frequency words tend to elicit reliably smaller N400s than low-frequency words (Allen, Badecker, & Osterhout, 2003; Barber, Vergara, & Carreiras, 2004; Van Petten, 1993; Van Petten & Kutas, 1990, 1991a, 1991b). This frequency effect is qualified by interactions with both repetition (within an experiment) and sentence semantic constraints (Van Petten, Kutas, Kluender, Mitchener, & McIsaac, 1991). When words are repeated in lists, or when entire sentences are repeated, the N400 frequency effect disappears upon second presentation (Smith & Halgren, 1987; Rugg, 1990; Besson, Kutas, & Van Petten, 1992). In other words, low-frequency words show a disproportionate repetition effect. Within sentences, semantic factors are also capable of abolishing the N400 frequency effect. Early in a sentence, low-frequency words elicit significantly larger N400s than high-frequency; this frequency effect is progressively attenuated during the course of a sentence. This ordinal word position by frequency interaction is due to semantic constraints since it is not present in either random word strings or in syntactically legal but semantically anomalous sentences; in both, the N400 effect remains unabated throughout. The N400 word frequency effect is also unaffected by grammaticality. Allen et al. (2003) found that high-frequency verbs (WORK) elicited smaller N400s than low-frequency verbs (SWAY) whether they were grammatical or ungrammatical when they appeared within a sentence (e.g., "The man will WORK/SWAY/WORKED/SWAYED on the platform.") that participants judged for acceptability. The N400s to low-frequency ungrammatical verbs were indistinguishable from those elicited by low-frequency grammatical verbs, at least for regular verbs, whereas the subsequent P600 component (see Section 6) was larger for ungrammatical than grammatical verbs but was unaffected by lexical frequency.

#### 4.2.4. *Concrete versus abstract words*

Within the category of nouns, those depicting a tangible object (often pictureable) have often been associated with a larger negativity in the N400 region than less imageable nouns. This *concreteness effect* tends to be more pronounced over frontal than parietal scalp, unlike the more centro-parietal distribution of the N400 semantic context effect for written words. The concreteness effect is larger when word processing goes beyond surface-level features and when contextual constraints are weak.

Paller, McIsaac, and Kutas (1987) first noted greater negativity between 300 and 900 ms to concrete than abstract words during a concrete/abstract judgment task, although Smith and Halgren (1987) did not see a similar effect during lexical decision. Kounios and Holcomb (1994) demonstrated the importance of task parameters within a repetition priming paradigm. They found larger (somewhat frontal) N400s to concrete than abstract words, but larger differences during an abstract/concrete judgment than in lexical decision. Assigned task also modulates the concreteness effect in sentences, which is larger in tasks requiring semantic analysis and mental imagery (although these two effects had slightly different distributions), and absent in a letter search task (West & Holcomb, 2000). As Kounios and Holcomb (1994) had also observed different ERP repetition

effects for these two word types, they argued that their topographical results were most consistent with the dual-coding theory's structural account of concreteness effects: namely, that concrete words had a processing advantage relative to abstract words because they not only activated a verbal semantic memory store, like abstract words, but also a nonverbal image-based semantic memory store (Paivio, 1991).

The concreteness effect also has been examined in sentences designed to bias the final word to be either concrete or abstract, as participants rendered sense–nonsense judgments (Holcomb, Kounios, Anderson, & West, 1999). When congruent, abstract and concrete words elicited indistinguishable ERPs. When anomalous, concrete words elicited somewhat more negative potentials at fronto-central sites than abstract words (in addition to a centro-parietal N400 elicited by both concrete and abstract words). The results of a follow-up experiment with less predictable congruent endings (lower cloze) demonstrated that an N400 concreteness effect, although more widespread across the scalp, also could be elicited by congruent endings. These findings led to the “context-extended dual coding” hypothesis, which maintains the notion of two different memory stores – a verbal one and an imagistic one – but offers similar effects of context for concrete and abstract words in the verbal system together with greater context effects for concrete words in the image system. The hypothesis also stipulates that concreteness effects can be overridden by supportive contexts under the assumption that contextual information is available prior to concreteness information. For auditory word pairs, Swaab, Baynes and Knight (2002) showed that a single related word is not sufficient context to override the concreteness effect. Although the ERPs to highly imageable words showed greater negativity than those to less imageable words, the concreteness effect was equivalent for related and unrelated pairs.

Two features of the concreteness effect warrant further discussion: (1) although it begins coincident with the N400, the concreteness effect can last well beyond the typical N400 context effect (sometimes to 1000 ms post-stimulus onset); (2) not just the amplitude but also the scalp topography of the concreteness effect varies with task variables. West and Holcomb (2000) noted that, whereas the concreteness effect in their semantic task had a frontal extent, it was still evident at posterior sites, much like that observed for sentences with low constraint. By contrast, the concreteness effect in their mental imagery task was progressively more frontal with time and, like the concreteness effect for anomalous sentence endings, was absent posteriorly. They thus suggested that the concreteness effect may sometimes comprise an N400 reflecting semantic processing plus an N700 reflecting image-based processing.

#### 4.2.5. *Orthographic neighborhood*

In reading, words with more orthographic neighbors (other words than can be formed by changing one letter) elicit larger N400s than words with fewer neighbors, although words with more neighbors elicit faster lexical decision times (Holcomb, Grainger, & O'Rourke, 2002). Holcomb et al. attributed this latter effect to greater global semantic activation when a word from a dense neighborhood is encountered, because this includes partial activation of numerous other words that form near-matches (activation that must ultimately be suppressed in order to zero in on the meaning of the current word).



### 4.3. Using N400 Latency to Track the Timing of Semantic Processing

One of the appeals of the ERP as a dependent measure is its exquisite temporal resolution, which can be used to clarify the relative timing and sequence of distinct processes in comprehension and production. Many components of the ERP show a fairly broad range of latencies that can be readily linked to the onset or completion of different aspects of stimulus analysis (see Kutas, McCarthy, & Donchin (1977) for stimulus evaluation and decision making using the P3b, and Schmitt, Münte, & Kutas (2000) for studies of speech production using the no-go N2 and Lateralized Readiness Potential components). In contrast, N400 latency for visual words is generally quite stable in the face of experimental manipulations, particularly as compared to the lawful variation in amplitude across conditions in the same subjects. There are only a handful of cases of reliable differences in the latency of N400 context effects in the visual modality. It appears that strong semantic relationships in word pairs may elicit slightly earlier effects than weaker relationships (Kutas & Van Petten, 1994, Figure 7; Luka & Van Petten, 2000). In lexical ambiguity paradigms, probes related to a contextually inappropriate sense of the ambiguous word are associated with a delayed context effect as compared to contextually appropriate probes (Van Petten & Kutas, 1987; Van Petten, 1995).

A major reason for the stability of N400 latencies in the visual modality may be the constancy in the timing of perceptual processes across psycholinguistically distinct categories of stimuli; perceptual information about the identity of a word is likely to reach the cortical areas involved in meaning construction at much the same time regardless of contextual manipulations. Moreover, the visual information in a single word is present all at once, and likely analyzed as a single visual pattern. The physical nature of the speech signal is quite different in having an extended temporal duration, so that information about the identity of a word accrues over time. This property of speech makes it more amenable to examinations of the temporal relationships between perceptual and semantic processes.

The earliest ERP studies of auditory word pairs and sentences revealed that the onset (and usually the peak) latency of semantic context effects were well before the acoustic offset of the eliciting words (McCallum, Farmer, & Pocock, 1984; Holcomb & Neville, 1990, 1991). This was not especially surprising given that, even when presented in isolation, most English words can be identified well before their offsets (Grosjean, 1980). More recent studies have explicitly examined the information content of the auditory signal at different points in time, relative to ERPs at those same timepoints.

Connolly and Phillips (1994) were the first to use incongruent sentence completions that shared initial phonemes with the congruent completions of those sentences, as in "The gambler had a streak of bad LUGGAGE". They found that the onset of the difference between congruent and these incongruent completions was delayed, as compared to incongruities with initial phonemes that mismatched the most expected sentence completion. Van Petten et al. pursued this finding by adding a condition, so that sentence completions were either congruent, incongruent with initial overlap, incongruent with final-overlap, or wholly incongruent (e.g., "It was a pleasant surprise to find that the car

repair bill was only 17 DOLLARS/DOLPHINS/SCHOLARS/HOSPITALS”; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). The rhyming (final-overlap) condition elicited ERPs identical those in the wholly incongruous condition, lending support to the idea that (at least in English) initial phonemes are used to establish a set of lexical candidates, and that “late entry” of candidates based on subsequent phonemes is limited. The sentence congruity effect in the rhyme and wholly incongruous conditions began ~150 ms after word onset, although the words averaged about 600 ms in duration. As in Connolly and Phillips’ (1994) study, the sentence congruity effect for words whose initial phonemes overlapped a congruent completion was delayed, beginning only ~400 ms after word onset. More critically, Van Petten et al. established the *isolation points* of the critical words prior to the sentence experiment, via the *gating* method. In this method, listeners are presented with only the first 50 ms of word, or the first 100 ms, etc. (in increments of 50 ms), and forced to guess/decide what the word might be. With brief amounts of acoustic input, the number of candidates generated may be as large as the number of participants (and all the candidates might be wrong), but at some point, the large majority of participants correctly specify the actual word. The signal duration when sufficient acoustic information is present to pick out one word and eliminate alternatives with similar onsets (i.e., to distinguish CAPTAIN from CAPTIVE and CAPSULE and CAPTION) is the isolation point, which ranged from 100 to 700 ms for the critical words in this experiment. When ERPs were time-locked to the isolation points of the critical words, the difference between congruent and incongruent words with shared initial phonemes began at the isolation point, because this was when listeners could first determine that the sentence completion was not, in fact, congruent (see Figure 3). More interestingly, however, the sentence congruity effect for words whose initial phonemes were inconsistent with a congruent completion (e.g., SCHOLARS in the sentence about a car repair bill) began some 200 ms before the isolation point. This result indicates that listeners initiate semantic processing, including integration with a sentence context, with only partial perceptual information about word identity.

Van den Brink and colleagues also compared auditory sentence completions that were congruent, wholly incongruent, or incongruent but sharing initial phonemes with a congruent completion (van den Brink, Brown, & Hagoort, 2001; van den Brink & Hagoort, 2004). They also observed substantially delayed sentence congruity effects when the acoustic onsets of the incongruent words were consistent with a potentially congruent completion. The results from three laboratories are thus remarkably consistent in showing that the timing of ERP sentence congruity effects closely track the auditory input, although somewhat different descriptions of these results have been offered. Van Petten et al. (1999) describe their results in terms of continuous semantic processing, with the onset of the sentence congruity effect depending on how long the acoustic input remains compatible with an acceptable sentence completion. Connolly and colleagues instead describe their results in terms of an early ERP component sensitive to the match or mismatch between expected and incoming words at a phonological level (phonological MMN or PMMN) versus a later one reflecting semantic processing (N400). Van den Brink and colleagues likewise describe their results in terms of two distinct phenomena: an early “N200” reflecting whether or not “assessment of form-based activated lexical candidates reveals the presence of a candidate that fits the semantic and syntactic constraints of the preceding sentence” versus

**The main highway was flooded so  
they had to take a long ...**

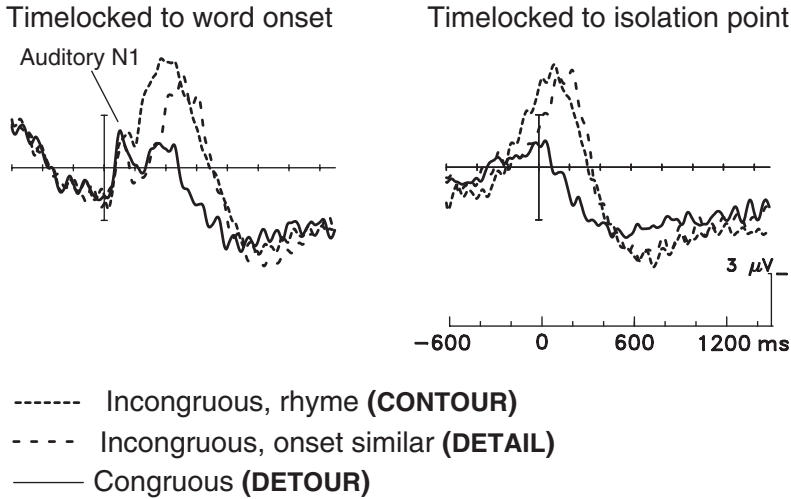


Figure 3. Grand average ( $N=21$ ), midline parietal site. Spoken sentences were completed with congruous words, words sharing onset phonemes with the congruous word, or words sharing final phonemes (rhyming) with the congruous words. ERPs to the rhyming completions were indistinguishable from ERPs to incongruous completions that were phonemically dissimilar to the congruous completion. Left panel: conventional average in which time zero corresponds to the onset of critical words. Note the auditory *N1* component peaking  $\sim 100$  after acoustic onset. *N400* onset and peak are delayed when the early portion of the presented word sounds like a congruous completion. Right panel: same data averaged with respect to the word isolation point, established in a gating experiment. *N1* is not visible here because it is elicited by acoustic onsets, and the isolation points for the words occur at variable times post-onset; the averaging process leads to a “smearing” of the *N1* across several hundred milliseconds, so that no peak is visible. Right panel shows that the *N400* sentence congruity effect begins before the presented word can be uniquely identified (at the isolation point). Data from Van Petten, Coulson, Rubin, Plante, and Parks (1999).

an *N400* that “indexes difficulty in lexical integration” (p. 1079). These disputes may or may not be of interest to the non-ERP researcher, given the consistency of the result that semantic context effects begin before word identification has run to completion.

## 5. HIGHER-LEVEL SEMANTICS: SENTENCES AND DISCOURSE

In the previous section, we described cases in which *N400* amplitude or latency reflects interactions between lexical characteristics of the eliciting word and sentence context. In this section, we describe interactions between different levels of semantic context: single word, sentence level, and discourse level, and then examine exactly what sorts of predictions readers/listeners derive from sentential or discourse contexts. The

extant results suggest that on-line comprehension involves an interplay between the current functional organization of semantic memory and more local contextual constraints.

### 5.1. Single Word versus Sentential Semantic Context

Several experiments have directly compared the ERP effects of lexical and sentential contexts. Kutas (1993) contrasted sentence-final words varying in cloze probability to strongly related, moderately related, and unrelated word pairs extracted from the same sentences. ERPs to the sentence completions were more positive overall than those elicited by the second words of the pairs (presumably due to sentence wrap-up effects), but the context effects were qualitatively quite similar, although somewhat larger and earlier in sentences. We also compared intermediate words of normal congruent sentences to “syntactic prose” (meaningless but grammatically correct sentences) as a measure of sentence-level context, and related to unrelated word pairs embedded in the syntactic prose as a measure of lexical context. These two varieties of semantic context produced qualitatively similar N400 effects, with the same onset latency, although the sentential effect lasted longer. When both lexical association and sentence-level congruity could aid in constructing a message-level interpretation of a sentence, both forms of context influenced N400 amplitude in an additive fashion (Van Petten, 1993; Van Petten, Weckerly, McIsaac, & Kutas, 1997).

More recently, we have pitted lexical and sentential context against one another by constructing sentences in which the final words were lexically associated to an intermediate word, but formed incongruent matches with the broader sentence context (e.g., “The zoo was working to breed the endangered BALD HEAD.”). These incongruent completions elicited an N400 as large as unrelated incongruent completions in healthy young adults, and much larger than that elicited by lexically unrelated but congruent completions (Coulson & Van Petten, 2000; Coulson, Federmeier, Van Petten, & Kutas, 2005; Van Petten et al., 1999). When lexical association and higher-level congruity are in conflict, the language processor thus favors contextual congruity.

#### 5.1.1. *Quantification and negation*

A possible exception to the general principle that “higher-level context overrules lower-level context” comes from studies examining simple subject–predicate statements with quantification or negation in sentence verification tasks. For instance, the final words of true class-inclusion statements (“A robin is a BIRD” or “All/Some apples are FRUITS”) elicit N400s equivalent in size to the final words of false statements (“A robin is not a BIRD” or “No apples are FRUITS”; Fischler, Bloom, Childers, Roucos, & Perry, 1983; Kounios & Holcomb, 1992). In these cases with category–exemplar relationships, noting the semantic relationship between the elements of the proposition rather than assessment of truth value seems to be reflected in the N400 (also see Section 5.3).

## 5.2. Sentential versus Discourse-Level Context

Kutas and Hillyard (1983) first observed an N400 effect of semantic violations within written prose passages, but made no systematic attempt to separate local from global context. A strong suggestion that the N400 is sensitive to semantic constraints that span sentence boundaries arises from comparing sentence-position effects in isolated sentences to those embedded in coherent text. Although the N400 to open-class words is large for the initial words of isolated sentences and then declines as the sentences progress, there is no equivalent sentence-position effect in connected prose, because even the earliest words of a given sentence continue the topic established earlier (Van Petten, 1995).

St. George and colleagues examined the impact of extra-sentential semantic cues directly, by recording participants' ERPs to all words as they read vague paragraphs that either were or were not preceded by a disambiguating title that made them easier to understand (St. George, Mannes, & Hoffman, 1994 as in Bransford & Johnson, 1972). Although the actual words as well as the local and global context were identical in the two conditions, a comparison of the ERP average of all the words in the untitled paragraphs versus those in the titled paragraphs revealed a smaller N400 for the titled stories. This is unequivocal evidence that the N400 is sensitive to context effects beyond the individual sentence and also reflects global or discourse-level context.

Similarly, van Berkum, Hagoort, and Brown (1999b; van Berkum, Zwitserlood, Hagoort, & Brown, 2003) showed that words that elicited N400s of approximately equal amplitude in an isolated (written or spoken) sentence showed differential N400 activity when they occurred in a discourse context that made one version more plausible than the other. For instance, QUICK and SLOW elicited about the same size N400s in "Jane told her brother that he was exceptionally quick/slow this morning." However, QUICK elicited a much smaller N400 when this sentence was preceded by "By five in the morning, Jane's brother had already showered and had even gotten dressed." The latency and topography of the discourse-level N400 effect are indistinguishable from those observed for various lexical semantic violations within isolated sentences.

The timing of the ERP context effect thus offers no support to models that give temporal precedence to lexical over sentential representations or processes, or to sentential level information over discourse-level information. Some form of parallel or at least cascaded processing thus must be incorporated into any viable model of language comprehension, unless the priority of word level over higher order information is so short-lived as to be empirically imperceptible, untestable, or theoretically inconsequential. Nor do the N400 data lend any support to language processing accounts that invoke distinct processing mechanisms for the recruitment and/or integration of word versus sentence level, or for intra-sentential (sentence-level) versus extra-sentential (discourse-level) information during the construction of sentence meaning. This is not to say that there are no functional differences in how different levels of context are initially computed. For instance, Van Petten et al. (1997) demonstrated that readers with smaller working memory capacities are less able to avail themselves of sentential context than high-span readers, but are equally able to utilize single-word contexts. However, after a preceding context has been

appropriately interpreted, there appears to be little difference in how single-word, sentence-level, and discourse-level contexts are applied to the immediate processing of the current word.

### 5.3. Language-Intrinsic Semantics versus Real-World Knowledge

Memory researchers use the term *semantic memory* to refer to a person's store of knowledge independent of the time, place, or manner in which that knowledge was acquired (in contrast to *episodic* memories of single events that occurred in a particular spatial and temporal context). Descriptions of the organization of semantic memory typically include no distinction between word definitions specified in a language, and thus known to any speaker with an adequate grasp of vocabulary (e.g., that GOOD would fall somewhere between EXCELLENT and FAIR in a rating scale containing all three words), and facts about the world that may or may not be known to a given native speaker (e.g., Australian but not American speakers of English may know that John Gorton was prime minister between John McEwen and William McMahon). Some linguistic and psycholinguistic theories do, however, draw a distinction between semantic knowledge that is intrinsic to a language and pragmatic knowledge that is independent of the language.

Most electrophysiological studies have used operational definitions of “semantic context”, based on what normative samples of participants consider good sentence completions (cloze probability), or related word pairs (production norms or ratings). It seems likely that a large proportion of the items used in comparisons of congruent to incongruent sentence completions utilized “semantic” rather than “pragmatic” knowledge, but most studies have included some mix thereof. Fischler and colleagues were the first to examine whether recently acquired knowledge modulated N400 amplitude, independent of linguistic knowledge of word definitions (Fischler, Childers, Achariyapaopan, & Perry, 1985). In the first phase of this experiment, participants learned a set of name/occupation pairs (“Matthew is a lawyer.”); in the second phase, correct and re-arranged pairs were presented (“Matthew is a lawyer.” versus “Matthew is a dentist.”). As compared to the true statements, the false items elicited a larger N400. More recently, definitionally incongruent sentence continuations were explicitly compared to those that were incongruent only by virtue of incidental knowledge (e.g., “The Dutch trains are ...” was continued with YELLOW (plausible and true), WHITE (plausible but false), or SOUR (incongruent by definition); Hagoort, Hald, Bastiaansen, & Peterson, 2004). The two varieties of inappropriate sentence continuations elicited statistically indistinguishable N400s as compared to the correct sentence continuations, as well as indistinguishable responses in functional magnetic resonance imaging data.<sup>2</sup> Both studies thus present a contrast to the results described above (Section 5.1.1), in which statements became false due to an inappropriate quantifier or the presence of a

---

<sup>2</sup> Hagoort et al. (2004) did, however, observe that compared to the correct sentence continuations, the world-knowledge violations elicited increased power in the gamma (30–70 Hz) frequency band, during roughly the same latency range as the N400, whereas the semantic violations did not (see Makeig, Debener, Onton, & Delorme (2004) for the value of mining information from different EEG frequency bands after stimulus presentation).

negative particle. Altogether, the data suggest that the N400 primarily indexes access to semantic memory, largely independent of the semantic/pragmatic distinction. However, just how general these results are remains to be seen. Negation has not been investigated with materials and tasks that are very natural and no work has yet examined the ERP consequences of systematically quantifying different sentence constituents (grammatical subject or object, verb phrase).

## 5.4. What Do Contextual Constraints Specify?

### 5.4.1. *Word forms versus semantic features (with some remarks on the immediate predictive value of context)*

Psycholinguistic descriptions of semantic context effects include several hypothetical mechanisms by which prior context can influence the processing of a current word. One way of categorizing such descriptions is by the timecourse of the proposed mechanism: true *priming* in which the representation of a word is preactivated during the processing of the context and before that word is actually presented (Collins & Loftus, 1975; Morton, 1969), versus various other *integrative* mechanisms that involve interactive processing of context and current word after both have been presented (Neely & Keefe, 1989; Norris, 1986; Ratcliff & McKoon, 1988). Characterizations of semantic context effects can also be divided according to their assumptions about the underlying representation of word meaning, and what is actually constrained by prior context. Some discussions of mechanisms assume that a predictive or priming mechanism necessarily involves anticipation of specific lexical items that might occur next, but other models stipulate that word meanings are comprised of bundles of semantic features shared with other words. In the latter sort of model with distributed representation of word meanings, context effects can be readily simulated by connectionist networks in which encountering one word produces partial activation of other words with shared semantic features (McRae, de Sa, & Seidenberg, 1997; Sharkey, 1989). Modeling semantic context effects that are not based on similarity (thematic relations) – which will include nearly all sentence and discourse effects – are likely to be a more difficult enterprise, although likely feasible (Elman, 2004; Ferretti, McRae, & Hatherell, 2001). A full discussion of these topics is outside the scope of the current review, but we note that questions about when semantic context acts are logically orthogonal to questions about exactly what becomes easier to process with supportive prior context. In empirical work, however, the same results often provide information about both “when” and “what”, so that the two issues are interlaced in the following review.

Both older and newer ERP studies of sentence processing argue very strongly that context can facilitate the processing of words with some appropriate semantic features, even when the specific lexical items could not have been expected or predicted. This was first demonstrated by comparisons between congruent completions of high cloze probability, anomalous completions, and anomalous completions that were semantically related to the congruent words (e.g., “The pizza was too hot to EAT / CRY / DRINK”; Kutas, Lindamood, & Hillyard., 1984; Kutas & Hillyard, 1984). The related anomalies elicited

a larger N400 than the congruent endings, but substantially smaller than the unrelated anomalies. Critically, the difference between the related and unrelated anomalies showed no latency delay, suggesting that words like DRINK were facilitated directly by the sentence context itself, rather than by some sort of secondary priming between the (unpresented, but predictable) congruent ending and the related anomaly. This result thus argues for a featural semantic representation, and suggests that a sentence context facilitates the processing of words containing at least some features that can be matched to the specifications of the preceding sentence fragment. Moreover, the results suggest that context specifies something about the meaning of upcoming words, but not necessarily a list of candidate word forms *per se*.

More recently, Federmeier and Kutas (1999a, 1999b) refined the “related anomaly” design by constructing contexts that more narrowly constrained the semantic features that would form a good fit. For instance, although both EARRING and NECKLACE are types of jewelry and thus share many semantic features, they also differ in multiple properties such that EARRING is a better completion for the context “I guess his girlfriend really encouraged him to get it pierced. But his father sure blew up when he came home wearing that ...” In contrast, NECKLACE is a better completion for the context “She keeps twirling it around and around under her collar. Stephanie seems really happy that Dan gave her that ...” The ERPs thus showed much larger N400s for the wrong variety of jewelry (or the wrong team sport, wrong hand tool, etc.) than for the congruent word. More critically however, the incongruent words that had high featural overlap with the congruent word elicited smaller N400s than words with low featural overlap. For the “twirling” sentence above, N400 amplitudes would follow a gradient of smallest for NECKLACE (congruent), intermediate for EARRING (incongruent but high-overlap), and largest for LIPSTICK (incongruent and low-overlap). Visual half-field studies revealed that this effect was present only for initial presentation to the left hemisphere (Federmeier & Kutas, 1999b).

Although the “related anomaly” experiments suggest that sentence and discourse contexts act to specify the meanings of plausible continuations, and not their physical forms, other results are most compatible with the idea that context also can be used to predict particular words. Some recent experiments have capitalized on situations in which semantic plausibility is linked with a non-semantic lexical feature. For instance, in both Spanish and Dutch, nouns have grammatical gender. The gender of a noun is largely unpredictable from its meaning, but in grammatically correct sentences, the genders of articles and adjectives must match their nouns. Thus, if readers actively anticipate that “Little Red Riding Hood carried the food for her grandmother in ...” A BASKET, specifically, rather than some sort of container generically, the Spanish reader will also predict the feminine article UNA — to agree with the feminine noun CANASTA (i.e., “Caperucita Roja cargaba la comida para su abuela en una canasta”).

Wicha et al. examined ERPs elicited by articles whose gender agreed or disagreed with the most plausible sentence continuation in written Spanish sentences, and in spoken and written sentences that continued with line drawings of objects (Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004). Van Berkum and colleagues used a very similar design with spoken Dutch materials, except that the



critical words were gender-marked adjectives (e.g., GROOT versus GROTE) that preceded their nouns by several words (van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). In all cases, articles and adjectives whose gender was inappropriate for the most plausible sentence continuation elicited different ERPs than words whose gender matched. These results are striking because the “inappropriate” articles/adjectives did not introduce any sort of error into the sentences; the observation of ERP differences at the early processing point can only indicate that participants were anticipating specific nouns of specific genders. The nature of the ERP effect differed across experiments, however. In the Spanish experiment using nouns depicted by line drawings, the “wrong gender” articles elicited larger negativities (N400-like activity) than “correct gender” articles (Wicha et al., 2003). In the language-only experiments (both Spanish and Dutch), the gender mismatching items elicited positive potentials similar to those observed to overt agreement errors (“P600-like” effects, see Section 6). The differential effects across experiments suggest that participants may have perceived the gender mismatch as an agreement error in the language-only experiments, but a violation of a more meaning-based violation in the mixed-media case.

Although English does not include grammatical gender, DeLong, Urbach and Kutas (2005) exploited a conceptually similar agreement phenomenon to examine the specificity of sentence-based predictions, namely the A/AN alternation based on noun phonology. Sentence fragments were constructed such that the cloze probability of possible completions (all nouns) ranged from 10 to 90%, and half of those nouns began with a vowel sound (calling for AN) and half began with a consonant sound (calling for A). Sentences were visually presented one word at a time, so that ERPs elicited by the articles could be examined contingent on whether they matched a very predictable sentence completion, or a less predictable (but congruent) completion. The articles always agreed with the subsequent nouns, so that no phonological mismatches were encountered during the experiment. The articles elicited N400 activity, whose amplitude was strongly (inversely) correlated with the cloze probability of the article. In other words, if KITE was a favored sentence continuation, then the word AN (as in AN AIRPLANE) elicited a larger N400 than the word A immediately preceding the expected noun. Because the effects were graded according to cloze probability, rather than showing a dichotomous split between more- versus less-favored sentence continuations, DeLong et al. concluded that although readers do predict word forms, they entertain a range of possibilities that are graded in strength. Like the experiments with grammatical gender, the results strongly suggest that predictions from context may include specific word forms, and that these predictions are made in real time during reading. Much work still needs to reveal exactly what informs these predictions, what neural substrates support predictive processing, and how what people know and how quickly they can access that knowledge influence the routine use and/or efficacy of prediction.

We began this section with a dichotomy: that sentences and discourse may be used to derive expectations for specific lexical items, or for the semantic content (features) of upcoming words. The data, however, seem not to respect this dichotomy. The “related-anomaly” experiments (Federmeier & Kutas, 1999a, 1999b; Kutas & Hillyard, 1984) indicate that words which would never be predicted on the basis of the preceding

context nonetheless undergo more fluent processing if they share semantic features with the most plausible completion. The experiments varying the gender and phonological form of articles clearly indicate that readers and listeners form expectations for specific words, because neither gender markings nor the A/AN alternation carry much, if any, semantic information, but are specific to lexical items. We might be tempted to conclude that general expectations about meaning are the norm, and that these become focused onto specific word forms only when the contextual constraint is extremely strong, such that only a single word exemplifies exactly the meaning specified by the context. Indeed, in the Federmeier and Kutas (1999a) study, the reduction in N400 amplitude to implausible items with high featural overlap was more pronounced in highly constraining sentences. However the graded effects observed in the “A versus AN” experiment would at first glance appear to argue against this (DeLong et al., 2005). Overall, the uneasy match between the initial theoretical dichotomy and the empirical results suggest that a division between word forms and their meanings may not be the correct description of how the brain processes words, and that some revision of our thinking may be in order.

#### 5.4.2. *Immediacy, incrementality, and prediction*

All of the context effects reviewed in this section emerge rather quickly after presentation of information that fits or does not fit with the prior context. In natural speech, van Berkum et al. (2005) showed that the differential response to adjectives whose gender was inconsistent with expectations about a subsequent noun began ~50 ms after presentation of the relevant acoustic input (the inflection at the end of the adjective). This early onset is about the same as when N400 differences between congruent and incongruent sentence completions begin in unedited natural speech (Holcomb & Neville, 1991). With visual presentation, differential activity to articles that do not match anticipated nouns begins somewhat later, around 200–300 ms (DeLong et al., 2005; Wicha et al., 2003), although sooner when time locked to the distinguishing information, (van Berkum et al., 2005). Congruent and incongruent open-class words show differential ERPs beginning around 200 ms after stimulus onset in the visual modality, as do incongruent words differing in their relationship with the congruent words. Combined with the results discussed in Section 4.3 showing contextual influences on spoken words before the acoustic information is sufficient to uniquely identify the words (Van Petten et al., 1999), the data point to a major role for top-down contextual influences in reading and listening.

While ERP data are no longer alone in suggesting this, we note that they have been offered as evidence for the immediacy and incrementality of (at least) semantic processing during on-line language comprehension since the early 1980s. To date, the ERP experiments have not addressed the full extent of the argument about the nature or completeness of the representations computed immediately on a word-by-word incremental basis (full or partial, and underspecified). Until quite recently ERP data, however, have stood as part of a small minority arguing not only for immediate semantic analysis and integration into an evolving sentence representation, but for contextual neural pre-activation of upcoming semantic, syntactic, and lexical information.

## 5.5. Non-literal Language

People use language in different ways for different purposes because it serves various communicative and social functions that go well beyond conveying facts. People do not always mean what they say, or say what they mean directly – and yet typically a reader/listener from the same culture as the speaker has no difficulty understanding that what s/he read or heard was a promise, a threat, a command, or an indirect request or that a statement is dripping with irony, funny, or intended to be metaphorical. The psycholinguistic and linguistic literatures are rife with discussions about the extent to which there is a basic distinction between literal and non-literal language representations and processes. Views span the range from those that argue that the dichotomy between literal and figurative thought or language is a psychological illusion, and that a single set of processes is responsible for the processing of both, to the strong claim that figurative language is unusual and special, and as such engages different comprehension processes (Katz, Cacciari, Gibbs, & Turner, 1998).

To date, there are only a few electrophysiological investigations of non-literal language processing, specifically of jokes and metaphors. One recurrent question in these studies is whether the right hemisphere makes a special contribution to the comprehension of non-literal language. This question has been of interest since early reports that one subtle communicative deficit in patients with damage to the right hemisphere is difficulty understanding non-literal language (Brownell, Simpson, Bihle, & Potter, 1990; but see Gagnon, Goulet, Giroux, & Joannette, 2003 for a recent claim that right- and left-hemisphere patients are more similar than different). None of the studies described below included neurological patients, relying instead on less direct means of assessing hemispheric asymmetry: examining the lateral distribution of scalp ERP effects, comparing right- and left-handed participants (on the hypothesis that left-handers have a somewhat more bilateral neural substrate for language), and visual half-field presentations.

### 5.5.1. Jokes

Coulson and Kutas (2001) compared the processing of one-line jokes versus non-joke sentences, with final words matched on cloze probability. Their primary aim was to test a two-stage model of joke comprehension wherein an initial stage of “surprise” registration is followed by a stage of coherence reestablishment. They were also able to assess the psychological reality of *frame shifting* – a process of activating a new frame from long-term memory in order to reinterpret information already in working memory (Coulson, 2001). Although not specific to jokes, frame shifting is necessary to reestablish coherence when encountering the punch word or line. As in many recent language studies, the specific pattern of results differed depending on contextual constraint (final word cloze above and below 40%) and whether or not individuals “got” the joke. Better joke comprehenders responded to jokes with larger late positivities (500–900 ms), a sustained negativity over left frontal sites, and — for those in constraining contexts — a slightly larger N400 as well. By contrast, in the poorer joke comprehenders, the punch-words elicited an enhanced frontal negativity (300–700 ms). Coulson and Lovett (2004)

likewise observed larger late positivities to jokes relative to cloze-equated straight endings, with a laterality influenced by participant handedness and gender. A frontal negativity was seen only in right handers, and a slightly enhanced N400 only in left handers with low verbal skills. The results were not simply explicable in terms of any two-stage theory. However, as the enhanced late positivity to jokes is not dissimilar to those reported for syntactic violations in nonhumorous sentences (see Section 6), it is worth considering the possible commonalities between the two in terms of sentence reanalysis, retrieval, and integration of information in working memory, etc.

Coulson and Williams (2005) examined ERPs to similar materials when punch words or straight endings were presented to one or the other hemifield to ensure that visual information reached one hemisphere slightly before the other. Jokes elicited larger N400s than straight endings only when the sentence-final words went into the right visual field (RVF) (left hemisphere). With left visual field (LVF) presentation, both jokes and low-cloze straight endings elicited larger N400s than high-cloze non-joke endings, but did not differ from each another. A sustained frontal negativity and a late fronto-central positivity to jokes did not differ with visual field of presentation. Overall, the right hemisphere seems no more stymied by processing a joke than by any other unexpected noun, suggesting that it may be better able to use sentential context to facilitate processing and integration of a punch word. This conclusion is supported by the studies in Coulson and Wu (2005), showing greater N400 reduction to a single word (in central vision) relevant than irrelevant to an immediately preceding one-line joke as well as a greater N400 reduction when the probe word was presented to the LVF than RVF (right hemisphere).

We can now reconsider whether joke processing differs from that of non-joke sentences. Certainly the data patterns indicate substantial overlap in processing, with the reading of both accompanied by modulations in N400 amplitude. At the same time, there appears to be a difference in the contributions of the two hemispheres to joke and non-joke processing; some aspect (unknown) of joke comprehension appears to be easier for the right hemisphere, as reflected in reduced N400s associated with lateralized presentation of either punch words or joke-relevant probe words following one-liners. Whether the ephemeral sustained negativity over left frontal sites also will prove to distinguish jokes from non-jokes remains to be seen. A similar uncertainty colors the specificity of the late positivities (frontal and/or parietal) that occasionally characterize the ERPs to jokes. What is most clear from these studies is the need to track more than just whether a sentence is a joke or not, including whether participants get it, and stable characteristics of participants such as verbal ability, handedness, familial handedness, and gender. Indeed, this is undoubtedly a valuable lesson for all language studies.

### 5.5.2. *Metaphors*

Most current processing models of metaphor comprehension assume that the same operations are involved in literal and metaphorical language comprehension, but that metaphorical language especially taxes certain operations (see Katz et al., 1998).

Several sources of behavioral evidence indicate that metaphorical meanings are sometimes available at the same time as literal meanings and may even compete with each other. Researchers have examined these issues with ERPs as equivalent reaction times do not necessarily translate into equivalent processing demands. No electrophysiological study has yet offered any strong evidence for a qualitative difference in the way literal and metaphorical language is processed. The final words of metaphors typically elicit slightly larger N400 amplitudes than equally unexpected (low cloze) words completing literal statements. This suggests that people invoke the same operations, but also do experience more difficulty integrating words with a metaphoric than literal context.

Pynte and colleagues initially established that final words of short metaphoric sentences elicited larger N400s than categorical statements, despite being matched on cloze probability (Pynte, Besson, Robichon, & Poli 1996). Subsequent experiments showed that the ease of processing metaphoric statement, like literal statements, could be modulated by prior context. When presented in isolation, relatively familiar and unfamiliar metaphors elicited equivalent ERPs (e.g., “Those fighters are LIONS.” versus “Those apprentices are LIONS.”). However, both sets of metaphors benefited from preceding context so that an unfamiliar metaphor with a useful context (“They are not cowardly. Those apprentices are LIONS.”) elicited a smaller N400 than a familiar metaphor preceded by an irrelevant context (“They are not naïve. Those fighters are LIONS.”), and similarly the familiar metaphors with a useful context were easier to process than unfamiliar metaphors with an irrelevant context. The metaphors-in-context were not compared to a literal condition to determine if the enhanced N400 observed for isolated metaphors disappeared with appropriate context. However, across the multiple experiments, there was no hint of distinct processing stages during metaphor comprehension.

Tartter and colleagues raise the possibility that, while processing a metaphorical expression, comprehenders nonetheless do take note of the anomalous nature of the expression’s literal meaning (Tartter, Gomes, Dubrovsky, Molholm, & Stewart, 2002). They suggest this realization may underlie the phenomenological sense of satisfaction experienced when confronting a metaphorical statement. They compared the ERPs to final words completing the same sentence frame either literally, metaphorically, or anomalously (e.g., “The flowers were watered by nature’s RAIN / TEARS / LAUGHTER”, respectively). Cloze probabilities were higher for the literal endings than the other two conditions (both near zero). They argue that if context is used to construct a meaningful interpretation of a metaphorical expression without any accompanying appreciation that the expression’s literal meaning is anomalous, then a metaphorical but literally incongruous ending should not elicit an N400. This construal of the N400 as an anomaly detector is problematic given that words that fit but are less expected also elicit sizable N400s; semantic anomalies are neither necessary nor sufficient to elicit N400s. Tartter et al. obtained a three-way amplitude difference in the peak latency range of the

N400: anomalous > metaphorical > literal; however, the ERPs to literal completions pulled away from the other two conditions earlier than the differentiation between metaphoric and anomalous completions. This pattern of results suggests (to us) that semantically anomalous sentence endings were more difficult to process (as reflected in larger and longer N400 congruity effect) than the metaphorical endings, which were in turn more difficult to fit with the prior context (as reflected in greater N400 activity) than the literal, congruent endings. The data pattern is also consistent with the view that metaphors are initially processed much the same as semantic anomalies, although they are meaningfully resolved in a shorter duration. However, this latter conclusion is somewhat complicated by the difference in cloze probability and frequency between the literal and metaphoric completions.

A significant analytic and empirical step in this area was taken by Coulson and Van Petten (2002), who hypothesized that the same conceptual operations important for understanding metaphors are often also engaged during the comprehension of literal statements. These include establishing mappings and recruiting background information, or, more specifically, looking for correspondences in attributes and relations between the target and source domains, setting up the mappings, aligning them, selecting some, and suppressing others. By using sentences describing situations where one object was substituted, mistaken for, or used to represent another (the *literal mapping* condition, e.g., “He used cough syrup as an INTOXICANT.”), they created sentences requiring mappings between two objects and the domains in which they commonly occur, albeit with less effort than for a metaphor (e.g., “He knows that power is a strong INTOXICANT.”), but more than for a simple literal statement with fewer or no mappings (e.g., “He knows that whiskey is a strong INTOXICANT.”). ERPs elicited by sentence-final words showed graded N400 activity, with metaphor > literal mapping > literal, although the three conditions were matched in cloze probability. These data indicate that although literal and figurative language may engage qualitatively similar processes, increasing the burdens on mapping and conceptual integration can make metaphors more difficult to process.

Finally, Kazmerski and colleagues examined individual differences in metaphor comprehension, and found that both vocabulary and working memory capacity were important factors as individuals determined whether a metaphoric statement was literally untrue (as compared to false statements without metaphoric interpretations, e.g., “The beaver is a LUMBERJACK.” versus “The rumor was a LUMBERJACK.”). High IQ participants showed greater interference, presumably because the figurative meaning was extracted without voluntary effort (Kazmerski, Blasko, & Dessalegn-Banchiamlack, 2003). Lower IQ participants had equivalent N400s for the metaphoric and anomalous statements, suggesting that they had no additional trouble rejecting metaphorical sentences as untrue. Thus, although individuals with lower IQs clearly understood the metaphors in an off-line task, the on-line evidence provided by the ERP seems to indicate that metaphorical processing is not always obligatory or automatic.

## 6. MORPHOSYNTACTIC PROCESSING AND RELATED COMPONENTS

This section surveys a number of issues concerning morphological and syntactic processing that have been addressed using ERPs: (1) the encapsulation and/or interaction of semantic and syntactic processes, (2) the influence of other, non-linguistic cognitive variables (such as working memory) on syntactic processing, and (3) the fractionation of syntactic processing into discrete stages. While the jury is still out on most of these issues, a body of evidence has begun to accumulate that allows us to reflect on just how much is known at this point. Invariably, predictions of Fodor's (1983) modularity hypothesis with regard to linguistic representations and processes provide much of the framework for this inquiry.

Before evaluating the evidence, however, it may be useful to invoke a caveat. While it is relatively easy, via experimental manipulation of linguistic materials, to obtain differences in the polarity, latency, amplitude, and scalp distribution of brain responses, it is often difficult to ascertain exactly what such differences might reflect functionally.

### 6.1. Background

As Sections 4 and 5 make clear, the N400 has become well established as a brain index of semantic and pragmatic processing. More recently discovered components related to syntactic and morphological processing have both complicated this picture and raised questions about the extent to which the N400 should be considered an all-purpose index of semantic processing. As early as 1983, Kutas and Hillyard demonstrated that while violations of semantic well-formedness reliably elicited an N400 (but see Section 6.2.2), violations of morphosyntactic well-formedness elicited different ERP components. In addition to semantic violations, the study included number agreement discrepancies (e.g. "*she dig*"; "*a balloons*"), as well as both finite and non-finite verb forms in inappropriate sentence contexts ("*to stayed*", "*are consider*"). In contrast to the centroparietal N400 between 300 and 500 ms elicited by semantic anomalies, the responses to all these morphosyntactic violations showed *frontocentral* negativities between 300 and 400 ms and marginally significant parietal *positivities* at 300 ms post onset of words immediately following the violations (Kutas & Hillyard, 1983, Figure 4). Although the import of these differences was not entirely clear at the time, Kutas and Hillyard observed that the elicitation of N400s by semantic but not morphosyntactic anomalies pointed to potentially separate underlying neural processing systems.

This state of affairs has largely persisted to the present day: morphosyntactic anomalies of various sorts have typically been associated with either anterior negative or late positive responses, or with both, but typically not with an N400, or not just an N400. Such morphosyntactically triggered responses exhibit a certain degree of variability with regard to both latency and scalp distribution.

Anterior negative responses are usually either left lateralized or bilaterally distributed (although see Ueno & Kluender, 2003b for right lateralized anterior negativities in

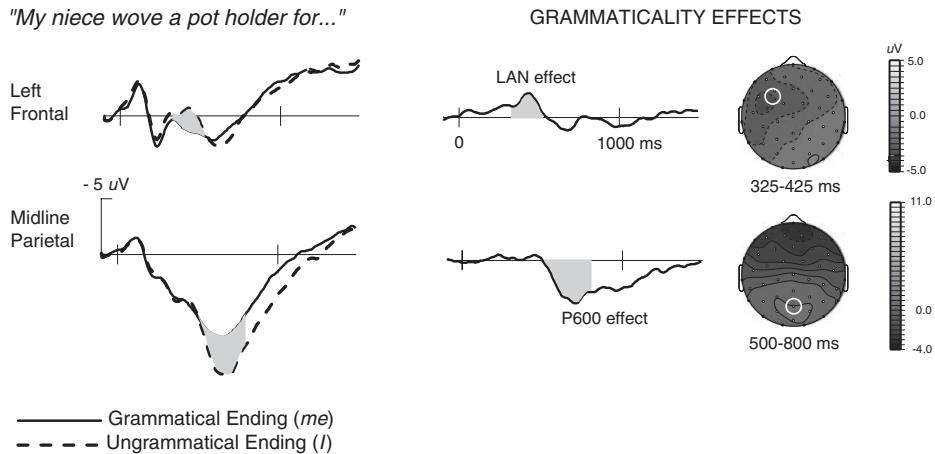


Figure 4. Representative data to grammatical violations (incorrect pronoun case markings, verb number mismatches) completing written sentences. Far left: grand average ERPs ( $N=16$ ) to final words of grammatical and ungrammatical sentences presented one word at a time (2 wds/s), intermingled with semantically congruent and anomalous sentences (see Figure 3). Middle: difference ERP created by point by point subtraction of ungrammatical minus grammatical ERPs; LAN effect is evident over the left frontal site and the P600 over a midline parietal site. Far right: topographical maps of the spatial distribution of the mean amplitudes (shaded area) for the LAN and P600 effects. Unpublished data from Grope and Kutas.

Japanese). There are, however, enough reports that this enhanced negativity is most pronounced at left anterior sites that it is commonly referred to as LAN. While this response often does not exhibit a clear peak, its latency usually falls between 300 and 500 ms post-stimulus onset, and it has also been reported as early as 100 ms. Although this early effect frequently persists into the 300–500 ms latency window and has the same scalp distribution as the later latency LAN, some researchers have proposed a functional distinction between negativities that occur between 100 and 300 ms post word onset – a so-called early LAN or “ELAN” – and those that occur between 300 and 500 ms (reserved for the LAN).

The experimental paradigms that elicit an ELAN violate the parser’s expectation that the incoming word will be of a particular grammatical category (e.g., a verb rather than a noun following a preposition plus article). Since the early LAN is impervious to the proportion of ill-formed experimental sentences (Hahne & Friederici, 1999), does not appear until 13 years of age (Hahne, Eckstein, & Friederici, 2004), is suppressed under degraded visual presentation (Gunter, Friederici, & Hahne, 1999), and is compromised by damage to left anterior regions, as in Broca’s aphasia (Friederici, von Cramon, & Kotz, 1999; Kotz, Frisch, von Cramon, & Friederici, 2003), it has been taken by some to index an early automatic process of local phrase structure building, during which word category information is used to assign initial syntactic structure. However, this conclusion remains controversial, as the ELAN has to date been reliably elicited under only a narrow set of conditions involving word category violations of this type.



The functional significance of the LAN (300–500 ms) has been similarly difficult to pin down, as it not only appears (like the ELAN) to word category violations, but also often to agreement violations in word (Hagoort & Brown, 2000) and pseudoword sentences (Münte, Matzke, & Johannes, 1997), and almost always accompanies fully grammatical long-distance dependency (filler-gap) constructions containing no violations (Kluender & Kutas, 1993a; King & Kutas, 1995). The main conundrum here is whether the LAN is a specific response to morphosyntactic illformedness, and/or whether it can (also) be explained in terms of general working memory processes. On the first view, inspired by serial parsing models, the LAN is hypothesized to reflect difficulties in the use of grammatical (as opposed to semantic) information like inflectional morphology (person, number, gender, and case features) used in thematic role assignment. Thus, while the ELAN is hypothesized to index an initial stage of phrase structure building, the LAN itself is hypothesized to index a subsequent processing stage devoted to thematic role assignment. On the second view, inspired by models of verbal working memory, the LAN is hypothesized to index both a “look forward” function triggered by displaced sentence constituents (e.g., fillers seeking subsequent gaps; Kluender & Kutas, 1993a, 1993b; Kluender & Münte, 1998), as well as a kind of a “look back” function triggered when current, unexpected syntactic information must be reconciled and aligned with preceding information occurring earlier in the sentence, including gaps seeking appropriate fillers (Kluender & Kutas, 1993a, 1993b; King & Kutas, 1995; Ueno & Kluender, 2003a), verbs seeking appropriate subjects (Osterhout & Holcomb, 1992; King & Kutas, 1995; Vos, Gunter, Kolk, & Mulder, 2001), anaphora seeking appropriate antecedents (Coulson, King, & Kutas, 1998; van Berkum, Brown, & Hagoort, 1999a; van Berkum, Brown, Hagoort, & Zwisserlood, 2003), and negative polarity items seeking appropriate licensors (Shao & Neville, 1996).

Note that these two views of the LAN are not mutually exclusive. Attempts to demarcate these two views have suggested that the LAN elicited by (morpho)syntactic illformedness is more reliably left lateralized than the LAN elicited by long-distance dependencies, but even LANs in response to morphosyntactic violations exhibit bilateral distribution on occasion, with both auditory and visual presentation. What remains consistent across the entire family of LAN and ELAN components, however, is its anterior scalp distribution (see Figure 4).

Late (i.e., later than the ELAN if not LAN and N400) positive ERP components to morphological and syntactic anomalies are typically largest over centroparietal sites, but can exhibit anterior maxima. This potential (measured between 500–800 ms) is now routinely referred to as the P600, as it often displays maximum amplitude at this latency (see Figure 4), although it can onset as early as 200 ms (following another positive component - the P200) and often appears as a long-lasting positive shift with no clear peak. While the P600 has been observed in response to a wide variety of violation types, including subject–verb agreement, verb inflection, case inflection, phrase structure, and higher-level syntactic constraints, it is not specific to violations *per se*. Enhanced late positivities have also been observed in syntactically well-formed sentences with a non-preferred structure (e.g., garden path sentences) or with relatively complex syntactic structures, such as those containing long-distance dependencies.

As with the LAN component, there are basically two schools of thought with regard to the P600. One is that the P600 is a general purpose response to low-probability target events often associated with some form of categorization and/or binary decision (P3b component). Alternatively, among the proposals for limiting the functional significance of the P600 to language contexts are suggestions that it indexes (1) the inability of the parser to assign the preferred structure to the input (Hagoort, Brown, & Groothusen, 1993), (2) a late, controlled (as opposed to automatic) process of syntactic reanalysis or repair once a syntactic error has been detected in a multi-stage parsing model (Friederici, Hahne, & Mecklinger, 1996), (3) syntactic integration difficulty (Kaan, Harris, Gibson & Holcomb, 2000), and (4) any kind of linguistic parsing difficulty (semantic, morphosyntactic, or orthographic; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998). The P600, unlike the early LAN, is sensitive to the proportion of experimental sentences that are syntactically ill-formed (Gunter, Stowe, & Mulder, 1997; Coulson et al., 1998; Hahne & Friederici, 1999), not present for morphosyntactic violations on pseudowords (Münte et al., 1997), and visible in Broca's aphasics (Friederici et al., 1999), but smaller or absent in individuals with damage to the basal ganglia (Frisch, Kotz, von Cramon, & Friederici, 2003; Kotz et al., 2003).

## 6.2. The Encapsulation versus Interaction of Syntax and Semantics

### 6.2.1. Interactions among ERP components

A number of the earliest studies of syntactic processing contained both standard semantic anomalies, which elicited an N400, as well as various types of morphosyntactic anomalies, which elicited ELANs, LANs, and P600s, although not in the same combinations: Hagoort et al. (1993) reported only a P600, Osterhout and Holcomb (1992) reported both a P600 and a LAN, while Neville, Nicol, Bars, Forster, and Garrett, (1991) and Friederici, Pfeifer, and Hahne (1993) reported an ELAN in addition to the LAN and P600.

A fair amount of research has since been devoted to determining the extent to which the N400 does or does not interact online with the purported markers of syntax-related processing (ELAN, LAN, P600). The results of studies crossing semantic and syntactic violations to this end have been rather mixed. Consequently, the answer to this question has been somewhat perplexing, suggesting that the inquiry may need reframing.

Gunter et al. (1997) reported that purely morphosyntactic violations elicited P600s of equivalent amplitude to morphosyntactically erroneous words containing an additional semantic anomaly, just as there was no difference in amplitude between the N400s elicited by semantic violations with or without additional morphosyntactic deformations. However, when complexity was added (in the form of an adverbial adjunct clause separating the main clause verb from its arguments), morphosyntactic violations with and without semantic incongruity elicited ELAN and LAN components in addition to the P600. Moreover, while the P600 varied in amplitude as a function of both semantic congruity and complexity (i.e., it was smaller either when the verb was morphosyntactically and semantically incorrect, or when the sentence containing it had a complex structure), the ELAN and LAN did not.

Osterhout and Nicol (1999) used stimulus materials of the type “One kangaroo at the San Diego Zoo would sometimes SIT / SITTING / WRITE / WRITING all day” in a similar manipulation. They reported that double violations (e.g., WRITING) elicited both an N400 and a P600. Both components showed marginally significant interactions between morphosyntax and semantics, and were not as large as would be predicted by linear summation of the N400 to semantic anomaly and the P600 to morphosyntactic ill-formedness — both facts suggesting some measure of interaction between the two types of processes. Purely morphosyntactic violations (SITTING) elicited a LAN in one experiment and an N400 in the other.

Using similar manipulations of subject–verb agreement in Finnish, Palolahti, Leino, Jokela, Kopra & Paavilainen (2005) reported that purely morphosyntactic violations elicited a LAN + P600 complex, semantic violations an N400, and the combined violation a LAN/N400 + P600 complex, which, however, was not a mere linear summation of the independent responses to morphosyntactic and semantic anomaly. The combined response showed greater negativity between 300 and 500 ms over the left hemisphere, again indicating an interaction between morphosyntactic and semantic processes. Given the slow presentation rate (one word every 800 ms), however, the results cannot be taken as conclusive.

Instead of manipulating verbal forms, Hagoort (2003a) placed morphosyntactic violations of gender and number marking (on definite articles) and semantic violations of plausibility (on following adjectives) in definite noun phrases in both sentence-initial and sentence-final positions. Compared to the control condition (roughly, “these<sub>COMMON</sub> noisy YOUNGSTERS<sub>COMMON</sub>”), in sentence-initial position the semantic violation (“these<sub>COMMON</sub> cloudy YOUNGSTERS<sub>COMMON</sub>”) elicited a larger N400, the morphosyntactic violation (“this<sub>NEUTER</sub> noisy YOUNGSTERS<sub>COMMON</sub>”) a larger LAN+P600 complex, and the combined violation (“this<sub>NEUTER</sub> cloudy YOUNGSTERS<sub>COMMON</sub>”) an even larger N400 than the semantic violation alone, followed by a P600 equivalent to that for morphosyntactic violations. Thus again, the double violation was not a linear sum of the independent responses to morphosyntactic and semantic violations alone, indicating some level of interaction between morphosyntactic and semantic processes. Since Hagoort reported a LAN to morphosyntactic violations, but a more anterior than usual (p. 892) equipotential negativity across the scalp (p. 887) to combined violations in sentence-initial positions, this may have been the same LAN/N400+P600 reported by Palolahti et al. (2005). Difference ERPs would help to resolve this issue.

Ainsworth-Darnell, Shulman, & Boland (1998) contrasted subcategorization violations with semantic violations using quartets of sentences like:

Jill entrusted the recipe to FRIENDS/PLATFORMS before she suddenly disappeared.

Jill entrusted the recipe FRIENDS/PLATFORMS before she suddenly disappeared.

The semantic violation (“to PLATFORMS”) elicited an N400, while the syntactic violation (“FRIENDS”) elicited both a small N400 and the expected P600. However, the combined violation (“PLATFORMS”) elicited an N400 that did not differ from that to the semantic violation, and a P600 that did not differ from that to the syntactic violation, suggesting no interaction of syntax with semantics.

In contrast to the above studies, which all relied on visual (sometimes rather slow) presentation, Friederici and colleagues have used spoken materials such as:

Das Brot wurde GEGESSEN  
*the bread was eaten*

Der Vulkan wurde GEGESSEN  
*the volcano was eaten*

Das Eis wurde im GEGESSEN  
*the ice [cream] was in-the eaten*  
 ‘The ice cream was eaten in the’

Das Türschloß wurde im GEGESSEN  
*the door-lock was in-the eaten*  
 ‘The door lock was eaten in the’

The manipulation in the last two examples is a word category violation because, instead of the expected head noun, the verb (“GEGESSEN”, ‘eaten’) immediately follows the co-alesced preposition plus definite article (“im”, ‘in-the’).

Since, on a modular view, syntactic processing should precede semantic processing, Hahne and Friederici (2002) predicted that combined violations of syntax and semantics should suppress the N400 elicited by semantic violations alone. Using a 100 ms post-stimulus onset baseline to compensate for the fact that the critical word was preceded by different lexical items across conditions, the semantic violation elicited an N400, the word category violation an ELAN + P600 complex, and the combined violation a broad anterior negativity plus a broad posterior positivity, both starting early and continuing throughout the epoch, but no N400, as predicted. When participants were instructed to determine whether the sentence made sense or not and to ignore the structural violations, the semantic violation again elicited an N400, the word category violation an ELAN and a very small, unreliable late positivity, and the combined condition a temporally extended anterior negativity plus a phasic posterior N400 — i.e., no P600 whatsoever. Hahne and Friederici concluded that task-dependent variables may suppress or enhance late controlled processes indexed by the N400 and the P600 but do not affect early automatic processes indexed by the ELAN.

An alternative interpretation is that the structure of the word category violation conditions, with the head noun of the aborted prepositional phrase (PP) missing, may have discouraged readers from attempting to associate the clause-final verb with the subject

noun phrase altogether, effectively blocking N400 modulation — unless participants were explicitly instructed to relate them for the sense judgment. Nevertheless, the task manipulation did not affect the appearance of the ELAN, even though its morphology and distribution in the word category violation and combined violation conditions differed.

Friederici, Gunter, Hahne, and Mauth (2004) used essentially the same paradigm with certain crucial modifications: (1) verbs were used with prefixes that were ambiguous between verbal and nominal readings, with disambiguation in the suffix, (2) nominals beginning with the same prefixes were used in four of the filler conditions to prevent early prediction of word category, and (3) stimulus sentences were extended such that the main clause verb was not sentence-final. They reported an N400 to the semantic violation and a LAN + P600 complex to both the word category violation and the combined violation. In other words, the N400 was again suppressed, although possibly for the same reason as above and/or because of overlap with the subsequent P600, which was larger in response to the double violation than to the word category violation alone, indicating some degree of interaction between syntax and semantics. The elicitation of a late LAN rather than an ELAN (early negativity) in this study was attributed to the fact that the word category information was contained in the disambiguating suffix rather than the prefix of the critical words.

An independent line of research investigating morphosyntactic parsing preferences rather than outright violations *per se* also bears on the question of syntax/semantics interactions: van Berkum and colleagues have shown immediate ERP responses to dispreferred parses triggered solely by preceding referential discourse context with auditory (Brown, van Berkum & Hagoort, 2000; van Berkum et al., 2003) and visual presentation (van Berkum, Brown, & Hagoort, 1999a). The experimental paradigm biased preference for or against a relative clause interpretation of a morphosyntactically ambiguous complementizer/relative pronoun (“dat”, “that”) by prior introduction of either two unique (the boy and the girl) or two non-unique (the two girls) referents. The use of a definite noun phrase (the girl) to refer to one of two non-unique referents in the previous discourse elicited a LAN to the head noun. Thereafter, continuations consistent with a complement clause interpretation elicited a P600 when two non-unique referents (the two girls) had been introduced in the preceding discourse context (crucially, even when these were morphosyntactically incompatible with an interpretation of “dat” as a relative pronoun), while continuations consistent with a relative clause interpretation elicited a P600 when two independent and unique discourse referents (the boy and the girl) had been introduced. In other words, discourse context influenced the initial parse of a structural ambiguity.

To sum up, simultaneous violations of syntax and semantics have resulted in the following reported ERP effects: no apparent differences (Ainsworth-Darnell et al., 1998), a larger N400 (Hagoort, 2003), suppression of the N400 (Hahne & Friederici, 2002; Friederici et al., 2004), a slightly smaller N400 and P600 (Osterhout & Nicol, 1999), a smaller P600 (Gunter et al., 1997), a larger P600 (Friederici, Gunter, Hahne, & Mauth, 2004), and increased anterior negativity of various types (Palolahti et al., 2005; Friederici et al., 2004). Bear in mind that there were crucial differences across these studies: some manipulated morphosyntax while others manipulated subcategorization or word category information,

some used visual and others auditory presentation. At a minimum, however, there do appear to be interactions of syntax with semantics in online processing, both in double violation studies and in manipulations of discourse context to influence morphosyntactic parsing. Moreover, these reported interactions appear to involve the N400, P600 and LAN equally, ERP components that do not emerge any earlier in the ERP record than 200 ms post-stimulus onset. Whether or not such interactions reliably involve or exclude the ELAN is a question that will have to await further research for a definitive answer.

### 6.2.2. *Elicitation of unexpected ERP components*

The above investigations of combined (morpho)syntactic and semantic violations were originally designed to test the exact nature of the relationship between the N400 and P600 components, although potential interactions involving other ERP components have also appeared. Beyond this, however, a spate of studies have reported manipulations that surprisingly elicited the opposite component from what one might have otherwise predicted, given the conception of the N400 as an index of semantic processing, and of the LAN, ELAN, and P600 as indices of morphosyntactic processing. These include studies that reported N400s in cases where one might have expected only P600s, and studies that reported P600 effects in contexts where one might have reasonably expected to see N400 amplitude modulations instead.

In the former category are German studies investigating case relations. German, with a relatively free ordering of arguments, relies heavily on case marking rather than on word order to determine thematic relations. Hopf, Bayer, Bader, & Meng (1998) placed case-ambiguous plural nouns at the beginning of German sentences, and verbs that assign either accusative (default) or dative (marked) case to their objects in clause-final position.

Dirigenten ... kann ein Kritiker ruhig UMJUBELN<sub>ACC</sub>/APPLAUDIEREN<sub>DAT</sub> ...  
*Conductors can a critic safely celebrate applaud*  
 ‘Critics can safely celebrate/applaud conductors ...’

Dative verbs with case-ambiguous objects (“Dirigenten”) elicited a right-lateralized, posterior N400 relative to accusative verbs, and relative to dative verbs with objects in first position clearly marked for dative case (“Musikern”, ‘musicians-DAT’ versus “Musiker”, ‘musicians-ACC’).

Frisch and Schlesewsky (2001) showed that when two adjacent animate (i.e., semantically reversible) noun phrases were both marked nominative, the second one elicited an N400 + P600 complex, even before the clause-final verb was reached.

Paul fragt sich, welchen Angler DER JÄGER gelobt hat  
*Paul asks himself which-ACC fisherman the-NOM hunter praised has*  
 ‘Paul wonders which fisherman the hunter praised’

\*Paul fragt sich, welcher Angler DER JÄGER gelobt hat  
*Paul asks himself which-NOM fisherman the-NOM hunter praised has*  
 (uninterpretable with both noun phrases (NPs) marked nominative)

This N400 was widely distributed across the scalp, raising the possibility that part of the response may have been a LAN. When these two nominative NPs were semantically asymmetrical—one animate and the other inanimate—the second NP elicited a P600 but no N400.

Paul fragt sich, welchen Förster DER ZWEIG gestreift hat  
*Paul asks himself which-ACC forester the-NOM branch brushed has*  
 ‘Paul wonders which branch brushed the forester’

\*Paul fragt sich, welcher Förster DER ZWEIG gestreift hat  
*Paul asks himself which-NOM forester the-NOM branch brushed has*  
 (uninterpretable with both NPs marked nominative)

Frisch and Schlesewsky attributed this to the fact that, with the thematic roles of the arguments clearly marked by animacy, the erroneous case marking affected only structure building operations (P600) rather than thematic interpretive processes (N400). Frisch & Schlesewsky (2005) replicated the N400 + P600 complex in response to the second of two animate arguments marked nominative, and showed the same response for two animate arguments in either the dative or the accusative case. Interestingly, the N400 to the second of two accusative NPs was larger than that to the second of two nominative NPs, while the P600 to the second of two dative NPs was larger than that to the second of two nominative NPs.

Bornkessel, McElree, Schlesewsky, and Friederici (2004) capitalized on the flexible ordering of arguments and verb-final word order in embedded clauses in German to create materials in which two case-ambiguous nouns, one singular and one plural, appeared adjacent to each other in an embedded clause, followed by a singular or plural verb. If the verb agreed in number with the first NP, then this NP had to be the subject and the second NP had to be the object (subject–object order); this is the default ordering of arguments in German. If the verb instead agreed with the second NP, then it had to be the subject of the clause and the first NP had to be the object (object–subject order). While fully grammatical, this is a marked order in German. Neither the thematic roles nor grammatical functions of these two NPs were apparent until the verb. Moreover, since both nouns were case ambiguous, it was possible to use verbs that assign accusative case to their theme objects as well as verbs that assign dative case. Accusative verbs that forced an object-first ordering of arguments elicited a P600 relative to the verbs of subject-first clauses (Friederici & Mecklinger, 1996), while dative verbs that forced an object-first ordering produced an N400. At this time, it is not entirely clear why reanalysis of case relations should elicit an N400 and reanalysis of phrase structural relations a P600. Bornkessel et al. (2004) concluded that case relations must be processed earlier than phrase structural relations.

A series of studies by various researchers during the same time period reported P600s to the verb in experimental sentences such as “For breakfast the eggs would only EAT

toast and jam” compared to the same verb in control sentences such as “For breakfast the boys would only EAT toast and jam” (Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). This finding could be attributed to a mere animacy violation, as in Frisch and Schlesewsky (2001), which would entail that this P600 be interpreted as indexing difficulty with structure-building operations. Kuperberg et al. (2003) essentially adopted this position, and Hoeks, Stowe, and Doedens (2004) likewise interpreted the P600 they observed to sentences containing animacy reversals like “The javelin has THROWN the athletes” as indexing thematic processing difficulty, specifically “effortful syntactic processing ... to obtain a semantically coherent and plausible sentence.” (p. 72)

However, Kolk, Chwilla, van Herten, and Oor (2003) and van Herten, Kolk, and Chwilla (2005) also reported P600s to Dutch sentences that reversed expected pragmatic plausibility relations rather than animacy/thematic relations *per se* (“The fox that HUNTED the poachers stalked through the woods”; note that both NP arguments preceded the critical verb in Dutch). Similarly, Kim, and Osterhout (2005) provided evidence suggesting that inanimate subject nouns do not lead to a P600 when associated with verbs that fail to encourage a pragmatically plausible combined interpretation (as in “The hungry tabletops were DEVOURING ...”). On the other hand, Kuperberg, Kreher, Sitnikova, Caplan, and Holcomb (in press) demonstrated that even when the inanimate subject noun and verb were semantically unrelated (“At breakfast the eggs would PLANT ...”), the verb still elicited a broadly distributed P600 equivalent and at some sites larger than that to semantically related verbs (EAT). Moreover, van Herten et al. (2005) provided evidence that the P600 was not due to a conflict between the expected and the actual inflectional ending on the verb. Finally, Kuperberg, Caplan, Sitnikova, Eddy, and Holcomb (2006) showed that the P600 to animacy violations in their stimuli was larger when the inanimate subject was a plausible argument of a transitive verb (“When John arrived at the restaurant, the food would ORDER ...”) than an implausible argument of an intransitive verb (“When they greeted the Queen of England, the trumpets would CURTSEY ...”). While the P600 elicited by morphosyntactic violations (“For breakfast the boys would only EATS ...”) was larger than that to thematic role animacy violations (“For breakfast the eggs would only EAT ...”), the two effects had an otherwise similar latency, morphology, and scalp distribution. As this line of research investigating unexpected N400 and P600 effects is currently very much a work in progress, one can reasonably anticipate additional clarity as more studies emerge in ensuing years.

### 6.3. Influence of Non-linguistic Cognitive Variables on Syntactic Processing

#### 6.3.1. Relationship of the P600 to the P3b

The current state of affairs regarding the relationship of the P600 (as a marker of morphosyntactic processing) to the N400 (as a marker of semantic processing) — as outlined in the preceding section — raises a number of general questions about the purported domain-specific nature of the P600 itself. N400s are known to be elicited by non-linguistic sources of semantic or pragmatic information such as line drawings and environmental sounds



(see Section 4). Likewise, P600-like potentials have been elicited by spelling errors that leave pronunciation intact (Münte, et al., 1998), harmonic and melodic violations in music (Besson & Macar, 1987; Janata, 1995; Patel, Gibson, Ratner, Besson & Holcomb, 1998), and violations of geometry (Besson & Macar, 1987), arithmetic rules (Nieddeggen & Rösler, 1999; Nuñez-Peña & Honrubia-Serrano, 2004), and abstract sequences (Lelekov, Dominey, & Garcia-Larrea, 2000). In addition to the P600s to thematic violations discussed at the end of the previous section, P600s have also been observed following N400s elicited by standard semantic anomalies (Münte, et al., 1998), and following a LAN to semantic violations of hyponymy and negative polarity (Shao & Neville, 1996).

In view of these and other complicating factors, the question has arisen whether the P600 is more judiciously viewed as a manifestation of a domain-general brain response elicited by rare, informative events. More specifically, several studies (Gunter et al., 1997; Coulson et al., 1998; Hahne & Friederici, 1999, but see Osterhout & Hagoort, 1999 for a counterargument) have proposed that the P600 should be considered a member of the P300 family known as the P3b (Donchin & Coles, 1988). Since P3b amplitude is known to be sensitive to the probability of occurrence of relevant stimulus types, each of these studies contrasted the presentation of blocks containing relatively rare (20–25%) versus relatively frequent (75–80%) syntactic violations. In response to frequent syntactic violations, the P600 effect was either reduced or eliminated entirely. By way of contrast, these probability manipulations did not affect preceding LAN or ELAN responses in any of these studies (see also Section 6.4).

To further test the relationship between the P600 and the P3b, two studies directly compared ERPs to infrequent auditory oddball tones (a standard paradigm for eliciting a P3b) versus syntactic violations in patients with brain lesions. Frisch et al. (2003) found that in contrast to brain-damaged controls, patients with lesions to the basal ganglia showed no P600s to sentences containing syntactic violations that were equiprobable (50%) with correct sentences, despite a preserved P3b to improbable (20%) high-pitched oddball tones. In Wassenaar, Brown, and Hagoort. (2004), 11 out of 12 normal controls, and all five patients with right hemisphere damage, but only 6 of 10 Broca's aphasics with frontal lesions (extent of basal ganglia damage unreported) showed a P600 to equiprobable subject–verb agreement violations, while all but two Broca's aphasics showed a P3b in an auditory tone discrimination task (20% low tones). Both studies thus claimed that the P600 and P3b were independent, that the P600 cannot be a general cognitive response of surprise, context updating, task relevance, salience, or probability, and that it must have a different neural substrate from the P3b.

### 6.3.2. *Role of working memory in syntactic processing*

In an early ERP study of English *wh*-questions, Kluender and Kutas (1993a, 1993b) showed in word-by-word comparisons of sentence positions intervening between filler and gap that, relative to *yes/no* questions, object *wh*-questions consistently elicited greater negativity over left anterior scalp between 300 and 500 ms. Subject *wh*-questions did not show this difference, presumably because the working memory load was no

higher than that incurred in *yes/no* questions. Direct comparisons of subject- versus object-relative clauses in the visual (King & Kutas, 1995) and auditory (Müller, King, & Kutas, 1997) modalities in English revealed that these transient effects were likely time slices of longer-lasting, slow anterior negative potentials, left-lateralized with visual (but not auditory) presentation.

In a related paradigm comparing biclausal structures differing only in the first word (“AFTER/BEFORE the scientist submitted the paper, the journal changed its policy”), Münte, Schiltz, and Kutas (1998) showed that BEFORE sentences with reversed chronological order likewise elicited slow negative potentials over left anterior sites relative to AFTER sentences. These differences were larger in participants with higher reading spans, a measure of verbal working memory capacity.

Relative to grammatical subject *wh*-questions, Kluender and Münte (1998) showed anterior negativities to both long-distance object *wh*-questions and to *that*-trace violations. As the negativity to object *wh*-questions consisted of slow frontal potentials (as in King & Kutas, 1995) that could also be seen in individual word responses, while that to *that*-trace violations consisted solely of left-lateralized phasic responses, Kluender and Münte hypothesized that there may be a more local LAN related to morphosyntactic processing, and a more global LAN related to processes of verbal working memory.

However, Vos et al. (2001) crossed morphosyntactic (subject–verb agreement) violations with syntactic complexity (a conjoined versus subject relative clause intervening between subject and verb), additional working memory load (an extraneous monitoring task for one versus three words in the stimulus sentences), and verbal working memory span (based on a listening span test) to show that LAN amplitude was influenced by all of these factors. In other words, these ERP responses suggested that working memory processes do interact with syntactic processing. This evidence was taken as support for the Just and Carpenter (1992) single resource (for both storage and processing) model of verbal working memory, rather than the Waters and Caplan (1996) model of two separate verbal working memory systems for automatic (e.g., syntactic) versus controlled processes.

Subsequent studies of long-distance object dependencies of various sorts both replicated and filled out this picture. Kaan et al. (2000) failed to replicate slow left-lateralized anterior negative potentials between filler and gap in object dependencies, but did demonstrate P600s to gap filling at the subcategorizing verb that were partly independent of and yet also partially overlapping with P600s to morphosyntactic anomalies. This P600 was interpreted as an index of syntactic integration costs at the subcategorizing position.

Fiebach, Schlesewsky, and Friederici (2001, 2002) compared indirect *wh*-questions that differed in how many adverbial adjunct PPs occurred between the clause-initial *wh*-filler and the second NP argument, followed by the clause-final verb — one PP (short) or two (long). Only long object *wh*-questions elicited the slow left-lateralized anterior

negative potentials reported in earlier studies. In contrast to Münte et al. (1998), however, it was the low (not the high) span readers who produced the larger slow potential effect. Fiebach et al. also replicated in German the P600 to gap filling in object questions reported by Kaan et al. (2000) for English; however, the P600 was seen already at the second NP (i.e., the subject NP with nominative case marking in object *wh*-questions) and not at the clause-final verb, indicating that the parser did not wait until the subcategorizer was reached before attempting to fill gaps in this verb-final language. Moreover, unlike the slow left-lateralized anterior negative potential between filler and gap, P600 amplitude to gap-filling did not differ across reading span groups, buttressing Kaan et al.'s (2000) claim that the P600 to gap-filling indexes syntactic integration rather than memory storage costs. While only long object *wh*-questions elicited sustained anterior negativity related to holding a filler in working memory, both long and short object *wh*-questions exhibited P600 effects related to gap-filling, albeit of larger amplitude for long object *wh*-questions.

Felser, Clahsen, and Münte (2003) reported slightly different results for long-distance object dependencies in German. Their stimuli differed from those in Fiebach et al. (2001, 2002) in the following ways. In addition to a *wh*-question condition, they included other types of filler-gap dependencies: a raising condition and two topicalization conditions. However, only object dependencies were included in the design and compared to each other; even the control condition involved short- (rather than long-) distance topicalization of a direct object. This effectively eliminated the possibility of monitoring for slow negativity between filler and gap. Also, the length of the dependency was increased by inserting an adjunct adverbial clause between filler and gap.

A transient LAN was elicited by all three long-distance filler-gap dependencies at the subject NP immediately following the intervening adverbial clause; this response was taken to index retrieval of the filler from working memory following processing of the adverbial clause. However, this LAN did not persist into the following indirect object NP immediately preceding the final verb. Contra Fiebach et al. (2002), there was no P600 at either NP prior to the final verb to indicate early syntactic integration of the filler. Instead, the long-distance topicalization and *wh*-question conditions elicited a phasic LAN relative to the raising construction at the final verb, and the *wh*-question condition also elicited a P600. What remains unclear is (1) why the purported syntactic integration of the filler in the *wh*-question condition would be delayed until the final subcategorizing verb position, and (2) why the filler in the long-distance topicalization condition would not undergo a similar process of integration at this point in the sentence.

Phillips, Kazanina, and Abada (2005) also reported a somewhat divergent set of findings from English. They contrasted long- (“The lieutenant knew WHICH ACCOMPLICE the detective hoped that the shrewd witness would RECOGNIZE in the line-up”) versus short- (“The detective hoped that the lieutenant knew WHICH ACCOMPLICE the shrewd witness would RECOGNIZE in the lineup”) embedded *wh*-object dependencies; the control condition consisted of a series of complement declarative clauses (“The lieutenant knew that the detective hoped that the shrewd witness would RECOGNIZE the

accomplice in the lineup”). The long *wh*-question condition elicited a sustained anterior negativity beginning with the *wh*-phrase and continuing throughout the dependency, similar to King and Kutas (1995), Kluender and Münte (1998), and Fiebach et al. (2001, 2002). Single-word averages in the long *wh*-condition likewise revealed phasic LAN effects at the two words immediately following the *wh*-phrase, similar to Kluender and Kutas (1993a, 1993b). Finally, there were P600s at the final verb (RECOGNIZE) in both *wh*-conditions, albeit between 300 and 500 ms in the short *wh*-condition and between 500 and 700 ms in the long *wh*-condition. The occurrence of the P600 at the subcategorizing verb was consistent with Kaan et al. (2000) for English, but the fact that the P600 difference between the long and short *wh*-conditions manifested as a latency rather than an amplitude difference was at odds with the results of Fiebach et al. (2002) for German.

By subjecting the slow negative potential to the long *wh*-condition to both a cumulative (cf. Fiebach et al., 2002) and a non-cumulative analysis (cf. King & Kutas, 1995), Phillips et al. (2005) showed that it did not increase in amplitude across the course of the sentence. Rather, it increased only across the intermediate clause containing the bridge verb (“the detective hoped that”); there was no further growth of the slow potential in the most deeply embedded clause (“the shrewd witness would”). The latter was true of the short *wh*-condition as well. Consequently, Phillips et al. claimed that while the sustained negativity likely indexed holding the *wh*-phrase in working memory, it was technically not sensitive to length, and therefore should not be viewed as an index of memory storage costs increasing over the course of a dependency. They attributed this discrepancy in interpretation with that proposed in Fiebach et al. (2002) to the greater length (in number of intervening words) and complexity (in terms of the intervening clause boundary) in their own study. As the length manipulation affected P600 latency rather than amplitude, Phillips et al. characterized P600 latency as an index of filler reactivation, a length-sensitive subprocess of syntactic integration (Gibson, 2000), and P600 amplitude as an index of thematic role assignment and compositional semantic interpretation, integration processes insensitive to length.

ERPs to scrambling in the German *Mittelfeld* (“middle field”) — i.e., ordering permutations of subject, indirect object, and direct object, the canonical word order in German — were reported in Rösler, Pechmann, Streb, Röder, & Hennighausen (1998). These were generally consistent with effects reported for *wh*-movement manipulations: LAN effects to NPs scrambled leftward — i.e., occurring in non-canonical positions — and broad P600-like effects beginning at the final argument of verb-final sentences with non-canonical argument ordering. Ueno and Kluender (2003a) showed related effects in Japanese when object NPs were scrambled leftward into non-canonical sentence positions (OSV rather than canonical SOV order): slow anterior negative potentials between filler and gap, and both phasic LANs and P600s to gap-filling, which, as in *wh*-movement in German (Fiebach et al., 2002), occurred before the subcategorizing clause-final verb was reached. Both scrambling studies pointed to the same conclusion, namely that the parser actively tries to restore constituents to their canonical underlying positions when faced with non-canonical permutations of word order. This finding may not be unrelated to the sensitivity of the parser to perturbations of the canonical chronological ordering of clausal events (Münte et al., 1998).

Two published studies of verb gapping (“Ron took the planks for the bookcase, and Bill\_the HAMMER with the big head”; Kaan, Wijnen, & Swaab, 2004) reported ELAN-like effects at the word immediately following the gap, a fronto-central negativity (120–200 ms) in German (Streb, Hennighausen, & Rösler, 2004), and a central-posterior negativity (100–200 ms) in English (Kaan, et al., 2004). Both studies also reported subsequent positivities: a fronto-central positivity (300–500 ms) in English (Kaan et al., 2004), and an apparent (although unmeasured and unreported) widespread midline positivity (400–600 ms) in German (Streb et al., 2004).

Follow-up work will be required to make sense of these commonalities, but based on the ERP evidence available thus far, it can be said with some measure of confidence that the mental processes involved in leftward scrambling appear to be quite similar to those involved in *wh*-movement, while the processes involved in verb gapping may be qualitatively different from (if not completely orthogonal to) those active in other types of filler–gap dependencies.

#### 6.4. The Fractionation of Syntactic Processing

The modularity hypothesis makes strong claims with regard to the encapsulation of language from other cognitive modules, and of syntax from semantics within the language module. As we have seen, ERP studies on aggregate point to a certain amount of cross talk among these domains. There appears to be some influence of non-linguistic factors (working memory and statistical probability) on ERP measures of human sentence processing, and purported ERP indices of syntactic and semantic processes also seem to interact. Within the syntactic processing module proper, very strong claims have likewise been made about the dissociation of a variety of syntax-related ERP effects. These are for the most part motivated by an ideological commitment to serial parsing models. This section reviews and evaluates some of these claims.

##### 6.4.1. ELAN

Much discussion has focused on the existence of the early LAN or ELAN, first reported by Neville et al. (1991) using visual presentation and by Friederici et al. (1993) using auditory presentation. Hagoort, Wassenaar, and Brown (2003b) attempted to replicate the word category violation effect in Dutch using visual presentation with minimal pair sentences (“The lumberjack dodged the vain PROPELLER/PROPELLED on Tuesday”). Note that since both conditions contained a semantic violation, this was in some sense a semantic versus semantic plus morphosyntactic violation manipulation. The end result was that an anomalous noun was presented in one condition and an anomalous verb in the other — but, as the authors pointed out, with zero cloze probability in both cases, eliminating another potential confound. The avoidance of the missing head noun confound in the Neville et al. (1991; “What did the scientist criticize Max’s OF ...?”) and in the Friederici et al. (1993; ‘The ice cream was in-the EATEN’) word category violation stimulus paradigms was another advantage. The word category violations in this study (“The lumberjack dodged the vain PROPELLED ...”) elicited a LAN + P600 complex, but no ELAN, perhaps because the stimuli were presented visually.

Rossi, Gugler, Hahne, and Friederici (2005) borrowed the double violation paradigm used in studies of syntax–semantics interactions to attempt a dissociation of early ERP effects related to word category information (ELAN) and to morphosyntax (LAN) using auditory materials. Because the stimuli were constructed in active rather than passive voice, they translate more or less directly into English (results indicated in parentheses): “The boy in kindergarten SINGS/SANG a song”, “The boy in SINGS/SANG a song” (ELAN, LAN, P600), “The boy in kindergarten SING/SANGS a song” (LAN, P600), and “The boy in SING/SANGS a song” (ELAN, LAN, P600). Owing to the differences in the words immediately preceding the verb across conditions, a post-stimulus baseline of 100 ms was used. Both the word category violation and the combined violation produced continuous negativity between 100 and 600 ms, whereas the subject–verb agreement violation by itself showed a significant LAN difference with the control condition only between 450 and 650 ms. Rossi et al. (2005) interpreted this as an indication that the processing of word category information takes primacy over the processing of other types of syntactic information like morphosyntax.

#### 6.4.2. *Is there more than one P600?*

In Mecklinger, Schriefers, Steinhauer, and Friederici (1995), fast comprehenders (i.e., participants who responded to comprehension questions with RTs < 800 ms) produced an enhanced early positivity beginning in the P200 peak in response to the sentence-final auxiliary in object relative clauses containing only case-ambiguous feminine nouns. The use of case-ambiguous nouns delayed resolution of the subject versus object relative clause reading until the clause-final verb complex. Mecklinger et al. suggested that this early positivity (P345) indexed a rapid revision of the parse — while leaving hierarchical phrase structure intact — toward the dispreferred object relative clause interpretation. The same comparison of relative clauses was contrasted with an analogous comparison of SOV vs. OSV complement clauses in Friederici, Mecklinger, Spencer, Steinhauer, and Donchin (2001). Object complement clauses elicited only a late positivity (500–900 ms), whereas object relative clauses elicited both an early (300–500 ms) and a late positivity (contra Mecklinger et al., 1995). This was attributed to the wider variety of materials used.

Hagoort, Brown, and Osterhout (1999) claimed that the revision of syntactic ambiguity toward dispreferred continuations elicits more frontally distributed P600 effects, while the repair of ungrammatical sequences elicits more posterior P600 effects. To test this, Friederici, Hahne, and Saddy (2002) crossed grammaticality (subject–verb agreement violations) with complexity (topicalization of a simple noun phrase versus a more complex verb plus noun phrase complex). The grammaticality manipulation yielded an N400 + posterior P600 (500–1100 ms), while the complexity manipulation produced an earlier frontal positivity (500–700 ms) as well as a widespread later positivity (800–1100 ms). As the earlier frontal positivity in the complexity manipulation involved neither ambiguity nor a need for revision, it was interpreted as an index of structural complexity more broadly construed.

This picture was complicated by the results of Kaan et al. (2000), who reported a posterior (rather than anterior) P600, not in response to either an outright violation or a dispreferred parse, but to a more complex structure. Kaan and Swaab (2003) investigated relative clause attachment ambiguity in order to shed further light on this issue. They compared preferred (“the cake beside the pizzas that WERE brought”), dispreferred (“the cakes beside the pizza that WERE brought”), and ungrammatical (“the cake beside the pizza that WERE brought”) continuations of sentences. Relative to preferred continuations, both dispreferred and ungrammatical continuations elicited positivity with the same posterior distribution. When these stimulus materials with two attachment sites for the relative clause were compared to a simpler grammatical structure with only one attachment site for the relative clause, all of the more complex continuations elicited greater positivity at frontal electrodes, regardless of whether they were preferred, dispreferred, or ungrammatical. Taken together, these results cast doubt on the claim that dispreferred continuations elicit frontal positive differences in the ERP record, while ungrammatical continuations elicit positivities with posterior maxima (Hagoort et al., 1999). However, a discrepancy remains between the frontal positivity elicited by structural complexity in Friederici et al. (2002) and Kaan and Swaab (2003) on the one hand, and the posterior positivity elicited by structural complexity in Kaan et al. (2000) on the other.

Further studies have raised additional questions with regard to the proper functional characterization of the P600. For example, Frisch, Schlesewsky, Saddy, and Alpermann (2002) showed a P600 not only at the point of syntactic disambiguation (in a sentence ambiguous between SOV and OSV word order), but already at the introduction of syntactic ambiguity into the parse itself (a case-ambiguous feminine noun occurring as the first argument in a sentence). This suggested that more than one word order alternative was under consideration during the ambiguous region, in line with predictions of parallel rather than serial parsing models, where the simplest structural alternative is always preferred initially. Bornkessel, Schlesewsky, and Friederici (2002) showed that the clause-final verbs of sentences in which phrase structure and case marking were held constant, but thematic relations varied based on the choice of predicate (e.g., a psychological predicate assigning the roles of theme and experiencer to the subject and object vs. a standard transitive predicate assigning the roles of agent and patient) elicited a late positivity between 200 and 600 ms when the case marking was unambiguous (i.e., using masculine nouns in German). NPs ambiguously marked for case (i.e., proper names and feminine nouns in German) elicited no such difference. Bornkessel et al. argued that late positivities should therefore be redefined as indexing more general hierarchical rather than purely syntactic information.

## 6.5. Morphosyntactic Processing: Conclusions

This section started out with relatively unequivocal and self-assured statements about the dissociation of semantic and syntactic processes as indexed by the N400 and the LAN/P600, respectively, as first suggested by Kutas and Hillyard (1983). We have now in some sense come full circle, however, as studies have begun to blur the lines of this once pleasantly simple picture. As outlined in Section 6.2.2, it now seems incontestable

that at least certain manipulations of case relations in German elicit N400s, and that certain manipulations of pragmatic plausibility at the lexical level elicit P600s.

As for the questions with which this section started out, the outlines of some of the answers have begun to emerge. First, with regard to the encapsulation versus interaction of syntactic and semantic processes, there is a fair amount of evidence that syntactic and semantic information do interact to some degree, although not with 100% predictability, and not always in the same way. Second, as to the influence of other, non-linguistic cognitive variables on syntactic processing, there seems to be a consensus that working memory does play a major role in syntactic processes (albeit to varying degrees in different individuals), but is not itself syntax specific. Likewise, it seems undeniable that P600-like effects have been elicited across a number of related cognitive domains. While lesion studies have demonstrated that the P3b to auditory tone discrimination tasks can be preserved when the P600 to syntactic violations is compromised, it seems unrealistic to expect complete overlap of brain representation for auditory tone discrimination and syntactic processing of language when even the auditory and visual P3b generators are not identical.

Perhaps the least amount of consensus is found regarding the fractionation of syntactic processing into discrete stages. While several neural models of language processing are currently available that may ultimately prove useful in having helped to shape and clarify our thinking about language processes in the brain (Friederici, 2002; Hagoort, 2003b), it is at present not clear whether our ability to theorize may have already outstripped our existing empirical base. Clearly, we still do not understand completely what the N400 indexes, let alone the exact nature of the more recently discovered language-related ERP components. By continuing to rely on the same types of known experimental paradigms, we may thus be putting the cart before the horse, and consequently not make as much progress as one might wish for. Because if we rush to load up the cart by assigning premature functional significance to differences that are relatively easy to come by, we may never get around to feeding the horse that is supposed to be pulling the wagon in the first place.

One possible way of getting around to this is addressed in Frisch, Hahne, and Friederici (2004):

... one has to keep in mind that our paradigm is somewhat ‘artificial’ in that sentences with violations like the ones we used rarely occur in normal processing. This argument ... applies to all paradigms testing ungrammatical structures ... seeing that violations seem to produce especially clear changes in the electrical activity of the brain. (p. 215)

This is an important acknowledgement that should not be completely or complacently ignored. As a research strategy, continuing to pursue the study of violation types may not necessarily answer as many questions as it raises. Reversing this trend requires a commitment to taking the road less traveled and sticking to it in small, systematic steps — an



approach to which the results and ensuing inferences of language ERP studies sans violations can already attest.

## ACKNOWLEDGEMENTS

M. Kutas (a Lady Davis fellow at Hebrew University) was supported from grants HD22614 and AG08313. We thank E. DeOchoa and T. Urbach for assistance.

## REFERENCES

- Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language*, 38, 112–130.
- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, 18, 405–430.
- Allison, T., McCarthy, G., Nobre, A. C., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, 5, 544–554.
- Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, 9, 415–430.
- Allison, T., Wood, C. C., & McCarthy, G. (1986). The central nervous system. In: M. G. H. Coles, E. Donchin, & S. W. Porges, *psychophysiology: Systems, Processes, and applications* (pp. 5–25). New York, N Y: Guilford Press.
- Anderson, J., & Holcomb, P. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: An event related brain potential study. *Psychophysiology*, 32, 177–190.
- Arezzo, J. C., Vaughan, H. G., Kraut, M. A., Steinschneider, M., & Legatt, A. D. (1986). Intercranial generators of event-related potentials in the monkey. In: R. Q. Cracco, & I. Bodis-Wollner (Eds), *Evoked potentials: Frontiers of clinical neuroscience* (pp. 174–189). New York: Alan R. Liss, Inc.
- Armstrong, E., Schleicher, A., Omran, H., Curtis, M., & Zilles, K. (1995). The ontogeny of human gyrification. *Cerebral Cortex*, 5, 56–63.
- Barber, H., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: Evidence from ERPs. *Neuroreport*, 15, 545–548.
- Beauducel, A., & Debener, S. (2003). Misallocation of variance in event-related potentials: Simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *Journal of Neuroscience Methods*, 124, 103–112.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343–355.

- Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*, 235–260.
- Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*, 14–25.
- Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience*, *4*, 132–149.
- Besson, M., Fischler, I., Boaz, T., & Raney, G. (1992). Effects of automatic associative activation on explicit and implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 89–105.
- Boddy, J. (1981). Evoked potentials and the dynamics of language processing. *Biological Psychology*, *13*, 125–140.
- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multidimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*, 495–522.
- Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Beyond syntax: Language related positivities reflect the revision of hierarchies. *NeuroReport*, *13*, 361–364.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.
- Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience*, *11*, 261–281.
- Brown C. M., van Berkum J. J. A., & Hagoort, P. (2000). Discourse before gender: An event-related brain potential study on the interplay of semantic and syntactic information during spoken language understanding. *Journal of Psycholinguistic Research*, *29*, 53–68.
- Brownell, H., Simpson, T., Bihrlle, A., & Potter, H. (1990). Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia*, *28*, 375–383.
- Buchwald, J. S., & Squires, N. S. (1982). Endogenous auditory potentials in the cat: A P300 model. In: C. Woody (Ed.), *Conditioning: Representation of involved neural function* (pp. 503–515). New York: Raven Press.
- Chiappa, K. H. (1983). *Evoked potentials in clinical medicine*. New York: Raven Press.
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274–285.
- Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

- Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal words of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256–266.
- Coulson, S. (2001). *Semantic leaps. Frame-shifting and conceptual blending in meaning construction*. Cambridge, MA: Cambridge University Press.
- Coulson, S., Federmeier, K., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word and sentence level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 129–147.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.
- Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*, 71–74.
- Coulson, S., & Lovett, C. (2004). Handedness, hemispheric asymmetries, and joke comprehension. *Cognitive Brain Research*, *19*, 275–288.
- Coulson, S., & Van Petten, C. (2000). Lexical priming and sentence processing. Presentation at thirteenth CUNY sentence processing conference, San Diego, March 2000.
- Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory and Cognition*, *30*, 958–968.
- Coulson, S., & Williams, R. F. (2005). Hemispheric asymmetries and joke comprehension. *Neuropsychologia*, *43*, 128–141.
- Coulson, S., & Wu, Y. -C. (2005). Right hemisphere activation of joke-related information: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*, 494–506.
- Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*, 60–74.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1122.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374.
- Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In: E. Callaway, P. Tueting, & S. H. Koslow (Eds), *Brain event-related potentials in man* (pp. 349–441). New York: Academic Press.
- Doyle, M. C., Rugg, M. D., & Wells, T. (1996). A comparison of the electrophysiological effects of formal and repetition priming. *Psychophysiology*, *33*, 132–147.

- Elger, C. E., Grunwald, T., Lehnertz, K., Kutas, M., Helmstaedter, C., Brockhaus, A., Van Roost, D., & Heinze, H. J. (1997). Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia*, *35*, 657–667.
- Elman, J. (2004). An alternative view of the mental lexicon. *Trends in the Cognitive Sciences*, *8*, 301–306.
- Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, *10*, 345–348.
- Federmeier, K. D., & Kutas, M. (1999a). A rose by any other name: Long term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.
- Federmeier, K. D., & Kutas, M. (1999b). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*, 373–392.
- Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs, and class-ambiguous words in context. *Brain*, *123*, 2552–2566.
- Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and *wh*-movement in German. *Brain and Language*, *87*, 345–354.
- Fernández, G., Heitkemper, P., Grunwald, T., Van Roost, D., Urbach, H., Pezer, N., Lehnertz, K., & Elger, C. E. (2001). Inferior temporal stream for word processing with integrated mnemonic function. *Human Brain Mapping*, *14*, 251–260.
- Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516–547.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler–gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, *30*, 321–338.
- Fiebach, C. J., Schlesewsky, M. & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*, 250–272.
- Fischler, I., Childers, D. G., Achariyapaopan, T., & Perry, N. W., Jr. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. *Biological Psychology*, *21*, 83–105.
- Fischler, I. S., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*, 400–409.
- Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Forster, K. I., & Hector, J. (2002). Cascaded versus noncascaded models of lexical and semantic processing: The turple effect. *Memory and Cognition*, *30*, 1106–1117.

- Friederici, A. D. (2002). Towards a neural basis for auditory sentence processing, *Trends in Cognitive Sciences*, 6, 78–84.
- Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *NeuroReport*, 15, 165–169.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1219–1248.
- Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, 31, 45–63.
- Friederici, A. D., Hahne, A., & von Cramon, D. Y. (1998). First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: Electrophysiological evidence for a double dissociation. *Brain and Language*, 62, 311–341.
- Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, 25, 157–176.
- Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research*, 11, 305–323.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 2, 183–192.
- Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain*, 122, 1033–1047.
- Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16, 1465–1477.
- Frisch, S., Hahne, A., & Friederici, A. D. (2004). Word category and verb-argument structure information in the dynamics of parsing. *Cognition*, 91, 191–219.
- Frisch, S., Kotz, S. A., von Cramon, D., & Friederici, D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology*, 14, 336–340.
- Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport*, 12, 3391–3394.
- Frisch, S., & Schlesewsky, M. (2005). The resolution of case conflicts from a neurophysiological perspective. *Cognitive Brain Research*, 25, 484–498.
- Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, 85, B83–B92.

Gagnon, L., Goulet, P., Giroux, F., & Joannette, Y. (2003). Processing of metaphoric and non-metaphoric alternative meanings of words after right- and left-hemisphere lesions. *Brain and Language*, 87, 217–226.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for common sense: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8, 89–106.

Garnsey, S. (1985). *Function and content words: Reaction time and evoked potential measures of word recognition*. Technical Report no. URCS 29, Rochester N.Y., University of Rochester.

Gibson, E. (2000). The dependency locality theory: A distance based theory of linguistic complexity. In: A. Marantz, Y. Miyashita, & W O'Neil (Eds), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). Cambridge, MA: The MIT Press.

Gratton, G., & Fabiani, M. (2001). The event-related optical signal: A new tool for studying brain function. *International Journal of Psychophysiology*, 42, 109–121.

Gratton, G., Fabiani, M., Elbert, T., & Rockstroh, B. (2003). Seeing right through you: Applications of optical imaging to the study of the human brain. *Psychophysiology*, 40, 487–491.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267–283.

Guerrera, C. (2004). *Flexibility and constraint in lexical access: Explorations in transposed-letter priming*. Doctoral dissertation, University of Arizona.

Guillem, F., N'Kaoua, B., Rougier, A., & Claverie, B. (1996). Differential involvement of the human temporal lobe structures in short- and long-term memory processes assessed by intracranial ERPs. *Psychophysiology*, 33, 720–730.

Gunter, T. C., Friederici, A. D., & Hahne, A. (1999). Brain responses during sentence reading: Visual input affects central processes. *Neuroreport*, 10, 3175–3178.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660–676.

Hagoort, P. (2003a). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883–899.

Hagoort, P. (2003b). How the brain solves the binding problem for language: A neurocognitive model of syntactic processing, *Neuroimage*, 20, S18–S29.

Hagoort P., & Brown, C. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38, 1531–1549.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483.

- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In: C. M. Brown, & P. Hagoort (Eds), *Neurocognition of language* (pp. 273–316). Oxford, UK: Oxford University Press.
- Hagoort, P., Brown, C. M., & Swaab, T. Y. (1996). Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia and patients with right hemisphere lesions without aphasia. *Brain*, *119*, 627–649.
- Hagoort, P., Hald, L., Bastiaansen, M., & Peterson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.
- Hagoort, P., Wassenaar, M., & Brown, C. (2003a). Real-time semantic compensation in patients with agrammatic comprehension: Electrophysiological evidence for multiple-route plasticity. *Proceedings of the National Academy of Sciences of the USA*, *100*, 4340–4345.
- Hagoort, P., Wassenaar, M., & Brown, C. (2003b). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, *16*, 38–50.
- Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, *16*, 1302–1318.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*, 194–205.
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*, 339–356.
- Halgren, E. (1990). Insights from evoked potentials into the neuropsychological mechanisms of reading. In: A. B. Scheibel, & A. F. Wechsler (Eds), *Neurobiology of higher cognitive function* (pp. 103–149). New York: Guilford Press.
- Halgren, E., Dhond, R., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like MEG responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*, 1101–1116.
- Halliday, A. M. (1982). *Evoked potentials in clinical testing*. New York: Churchill- Livingstone.
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography – theory, instrumentation, and application to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*, 413–497.
- Handy, T. C. (Ed.). (2005). *Event-Related Potentials: A methods handbook*. Cambridge, MA: MIT Press.
- Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, *121*, 1133–1142.

- Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S., & Lyytinen, H. (2002). Cortical activation during spoken word segmentation in nonreading impaired and dyslexic adults. *The Journal of Neuroscience*, *22*, 2936–2944.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59–73.
- Holcomb, P. J., Grainger Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*(6), 938–950.
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 721–742.
- Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*, 259–276.
- Holcomb, P. J., & Neville, H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*, 281–312.
- Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*, 286–300.
- Hopf, J. -M., Bayer, J., Bader, M., & Meng, M. (1998). Event-related brain potentials and case information in syntactic ambiguities. *Journal of Cognitive Neuroscience*, *10*, 264–280.
- Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*, 153–164.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
- Kaan, E., Harris, T., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.
- Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*, 98–110.
- Kaan, E., Wijnen, F., & Swaab, T. Y. (2004). Gapping: Electrophysiological evidence for immediate processing of “missing” verbs in sentence comprehension. *89*(3), 584–592.
- Katz, A. N., Cacciari, C., Gibbs, R. W., Jr., & Turner, M. (1998). *Figurative language and thought*. New York: Oxford University Press.
- Kazmerski, V. A., Blasko, D. G., & Dessalegn-Banchiamlack, G. (2003). ERP and behavioral evidence of individual differences in metaphor comprehension. *Memory and Cognition*, *31*, 673–689.



- Keurster, M., Brown, C. M., Hagoort, P., & Stegeman, D. F. (1999). Electrophysiological manifestations of open- and closed-class words in patients with Broca's aphasia with agrammatic comprehension. An event-related brain potential study. *Brain*, *122*, 839–854.
- Khader, P., Scherag, A., Streb, J., & Rösler, F. (2003). Differences between noun and verb processing in a minimal phrase context: A semantic priming study using event-related brain potentials. *Cognitive Brain Research*, *17*, 293–313.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–222.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-related ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 378–397.
- King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, *244*, 61–64.
- Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196–214.
- Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*, 573–633.
- Kluender, R., & Münte, T. F. (1998). Subject/object asymmetries: ERPs to grammatical and ungrammatical *wh*-questions. *Poster presented at the 11th annual CUNY conference on human sentence processing, March, New Brunswick, NJ*.
- Knight, R. T., Scabini, D., Woods, D. L., & Clayworth, C. C. (1989). Contributions of temporal-parietal junction to the human auditory P3. *Brain Research*, *502*, 109–116.
- Kojima, T., & Kaga, K. (2003). Auditory lexical-semantic processing impairments in aphasic patients reflected in event-related potentials (N400). *Auris Nasus Larynx*, *30*, 369–378.
- Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36.
- Kotz, S. A., & Friederici, A. D. (2003). Electrophysiology of normal and pathological language processing. *Journal of Neurolinguistics*, *16*, 43–58.
- Kotz, S. A., Frisch, S., von Cramon, D. Y., & Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society*, *9*, 1053–1060.
- Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related potentials and reaction times. *Journal of Experimental Psychology*, *121*, 460–480.

- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 804–823.
- Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*, 21(5), 489–530.
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (in press). The role of animacy and thematic relations in processing activate English sentences: Evidence from event-related potentials. *Brain and Language*.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117–129.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Process*, 8, 533–572.
- Kutas, M., & Hillyard, S. A. (1980a). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Hillyard, S. A. (1980b). Reading between the lines: Event-related brain potentials during natural speech processing. *Brain and Language*, 11, 354–273.
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, 11, 539–550.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Kutas, M., Hillyard, S. A., & Gazzaniga, M. (1988). Processing of semantic anomaly by right and left hemispheres of commissurotomy patients. *Brain*, 111, 553–576.
- Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and eventrelated brain potentials during sentence processing. In: S. Kornblum, & J. Requin (Eds), *Preparatory states and processes* (pp. 217–237). Hillsdale, NJ: Erlbaum.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 197, 792–795.
- Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing. *Electroencephalography and Clinical Neurophysiology, Supplement*, 39, 325–330.
- Kutas, M., & Van Petten, C., San Diego: Academic Press, Psycholinguistics electrified: Event-related brain potential investigations. In: M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, (1994,) (pp. 83–143.)

- Kwon, H., Kuriki, S., Kim, J. M., Lee, Y. H., Kim, K., & Nam, K. (2005). MEG study on neural activities associated with syntactic and semantic violations in spoken Korean sentences. *Neuroscience Research*, *51*, 349–357.
- Lelekov, T., Dominey, P. F., & Garcia-Larrea, L. (2000). Dissociable ERP profiles for processing rules vs instances in a cognitive sequencing task. *Neuroreport*, *11*, 1129–1132.
- Luka, B. J., & Van Petten, C. (2003). Qualitative types of relatedness: ERP differences between semantic similarity and lexical contiguity. *Poster presented at Linguistic Society of America Annual Meeting, January, Atlanta GA.*
- Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In: D. Touretzky, M. Mozer & M. Hasselmo (Eds), *Advances in neural information processing systems* (Vol. 8, pp. 145–151). Cambridge: MIT Press.
- Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Science*, *8*, 204–210.
- Marchand, Y., D'Arcy, R. C. N., & Connolly, J. F. (2002). Linking neurophysiological and neuropsychological measures for aphasia assessment. *Clinical Neurophysiology*, *113*, 1715–1722.
- McCallum, W. C., Curry, S., & Hutch, S. (Eds). (1993), *Slow potential changes in the human brain* (303pp.) New York, N Y: U.S., Plenum Press.
- McCallum, W. C., Farmer, S. F., & Pockock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *59*, 477–488.
- McCarthy, G., Nobre, A. C., Bentin, S., & Spencer, D. D., (1995), Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *Journal of Neuroscience*, *15*, 1080–1089.
- McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *11*, 495–523.
- McCarthy, G., Wood, C. C., Williamson, P. D., & Spencer, D. D. (1989). Task dependent field potentials in human hippocampal formation. *Journal of Neuroscience*, *9*, 4253–4268.
- McRae, K., de Sa, V., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.
- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition*, *23*(4), 477–494.
- Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13–20 months. *Developmental Neuropsychology*, *13*, 397–446.
- Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development*, *20*, 19–31.

- Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month olds. *Journal of Cognitive Neuroscience*, *16*, 1452–1464.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
- Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials to relative clause processing in spoken sentences. *Cognitive Brain Research*, *5*, 193–203.
- Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology*, *47*, 36–45.
- Münte, T. F., Heinze, H. -J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *36*, 217–226.
- Münte, T. F., Matzke, T. F., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, *9*, 318–329.
- Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*, 71–73.
- Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: Class and frequency effects. *Neuropsychologia*, *39*, 91–102.
- Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective/retrospective processing theory. In: G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207–248). San Diego: Academic Press.
- Neville, H. J., Mills, D. L., & Lawson, D. (1992). Fractionating language: Different neural systems with different sensitive periods. *Cerebral Cortex*, *2*, 244–258.
- Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151–165.
- Niedeggen, M., & Rosler, F. (1999). N400-effects reflect activation spread during arithmetic fact retrieval. *Psychological Science*, *10*, 271–276.
- Nobre, A. C., & McCarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *Journal of Neuroscience*, *15*, 1090–1098.
- Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, *372*, 260–263.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, *22*, 93–136.

Nunez, P. L. (1981). *Electric fields of the brain*. New York: Oxford University Press.

Nunez-Pena, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, *18*, 130–141.

Olichney, J., Van Petten, C., Paller, K., Salmon, D., Irarui, V., & Kutas, M. (2000). Word repetition in amnesia: Electrophysiological evidence of spared and impaired memory. *Brain*, *123*, 1948–1963.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology*, *46*, 143–68.

Osterhout, L., & Hagoort, P. (1999) A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate, *Language and Cognitive Processes*, *14*, 1–14.

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*, 283–317.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987) Neural correlates of encoding in an incidental learning paradigm, *Electroencephalography and Clinical Neurophysiology*, *67*, 360–371.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255–287.

Palohati, M., Leino, S., Jokela, M., Kopra, K., & Paavilainen, P. (2005). Event-related potentials suggest early interaction between syntax and semantics during on-line sentence comprehension. *Neuroscience Letters*, *384*, 222–227.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717–733.

Patterson, K., & Hodges, J. R. (2000). Semantic dementia: One window on the structure and organisation of semantic memory. In: L. S. Cermak (Ed.), *Handbook of neuropsychology (2nd Ed. Vol 2)*,(pp. 313–333). *Memory and its disorders* Amsterdam: Elsevier.

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory and Cognition*, *31*, 829–841.

Phillips, C., Kazanina, N., & Abada, S.H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407–428.

Pineda, J. A., Swick, D., & Foote, S. L. (1991). Noradrenergic and cholinergic influences on the genesis of P3-like potentials. In: C. H. M. Brunia (Ed.), *Event-related brain research* (pp. 165–172). Amsterdam: Elsevier.

- Plante, E., Van Petten, C., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal processing in learning disabled adults. *Neuropsychologia*, *38*, 1669–1684.
- Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, *81*, 666–678.
- Pynte, J., Besson, M., Robichon, F. -H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, *55*, 293–316.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385–408.
- Rösler, F., Pechmann, T., Streb, J., Röder B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, *38*, 150–176.
- Rösler, F., Streb, J., & Haan, H. (2001). Event-related brain potentials evoked by verbs and nouns in a primed lexical decision task. *Psychophysiology*, *38*, 694–703.
- Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters*, *384*, 228–233.
- Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory and Cognition*, *15*, 473–481.
- Salmelin, R., Service, E., Kiesila, P., Uutela, K., & Salonen, O. (1996). Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Annals of Neurology*, *40*, 157–162.
- Schendan, H., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual organization of words and faces within 150 ms. *Psychophysiology*, *35*, 240–251.
- Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, *37*, 473–484.
- Shao, J., & Neville, H. J. (1996). ERPs elicited by semantic anomalies: Beyond the N400. *Proceedings of the third meeting of the Cognitive Neuroscience Society*, San Francisco, CA, (p. 73).
- Sharkey, N. E. (1989). The lexical distance model and word priming. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society*, Hillsdale, NJ, Erlbaum, (pp. 860–867).
- Simon, G., Bernard, C., Largy, P., Lalonde, R., & Rebai, M. (2004). Chronometry of visual word recognition during passive and lexical decision tasks: An ERP investigation. *International Journal of Neuroscience*, *114*, 1401–1432.
- Simos, P. G., Basile, L. F. H., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research*, *762*, 29–39.

- Smith, M. E., & Halgren, E. (1987). Event-related potentials during lexical decision: Effects of repetition, word frequency, pronounceability, and concreteness. In: R. Johnson Jr., J. W. Rohrbaugh, & R. Parasuraman (Eds), *Current trends in event-related potential research: Electroencephalography and clinical neurophysiology* (Suppl. 40, pp. 417–421). Amsterdam: Elsevier.
- Smith, M. E., Stapleton, J. M., & Halgren, E. (1986). Human medial temporal lobe potentials evoked in memory and language tasks. *Electroencephalography and Clinical Neurophysiology*, 63, 145–159.
- Spencer, K. D., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38, 343–358.
- St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6, 71–83.
- Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language*, 90, 88–94.
- Streb, J., Henninghausen, E., & Rosler, F. (2004). Different Anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, 33, 175–201.
- Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: An ERP study. *Cognitive Brain Research*, 15, 99–103.
- Swaab, T. Y., Brown, C., & Hagoort, P. (1997). Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit. *Journal of Cognitive Neuroscience*, 9, 39–66.
- Swick, D. (2004). ERPs in neuropsychological populations. In: T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 299–322). Cambridge, MA: MIT Press.
- Swick, D., Kutas, M., & Knight, R. T. (1998). Prefrontal lesions eliminate the LPC but do not affect the N400 during sentence reading. *Journal of Cognitive Neuroscience Supplement* (March 18), 29.
- Tarkiainen, A., Cornelissen, P. L., & Salmelin, R. (2002). Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain*, 125, 1225–1136.
- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122, 2119–2132.
- Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel metaphors appear anomalous at least momentarily: Evidence from N400. *Brain and Language*, 80, 488–509.
- Ueno, M., & Kluender, R. (2003a). Event-related brain indices of Japanese scrambling. *Brain and Language*, 86, 243–271.
- Ueno, M., & Kluender, R. (2003b). On the processing of Japanese *wh*-questions: Relating grammar and brain. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, University of California, San Diego, California, 22, 491–504.

- van Berkum, J. J. A., Brown, C. M., & Hagoort, P. M. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.
- Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension. *Psychophysiology*, *40*, 235–248.
- van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 443–467.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999b). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701–718.
- van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967–985.
- van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience*, *16*, 1068–1084.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*, 241–255.
- Van Petten, C. (1989). *Context effects in word recognition: Studies using event-related brain potentials*. Doctoral dissertation, University of California, San Diego.
- Van Petten, C. (1993). A comparison of lexical and sentence-level context effects and their temporal parameters. *Language and Cognitive Processes*, *8*, 485–532.
- Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, *32*, 511–525.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Timecourse of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 394–417.
- Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: an event-related potential analysis of the time course of meaning activation. *Journal of Memory and Language*, *26*, 188–208.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*, 380–393.



- Van Petten, C., & Kutas, M. (1991a). Influences of semantic and syntactic context on open and closed class words. *Memory and Cognition*, *19*, 95–112.
- Van Petten, C., & Kutas, M. (1991b). Electrophysiological evidence for the flexibility of lexical processing. In G. Simpson (Ed.), *Word and Sentence* (pp 129–174). Amsterdam: North Holland Press.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*, 131–150.
- Van Petten, C., & Luka, B. L. (2006). Neural localization of semantic context effects in electrophysiological and hemodynamic studies. *Brain and Language*, *97*, 279–293.
- Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*, 485–508.
- Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*, 238–242.
- Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, *38*, 41–63.
- Wassenaar, M., Brown, C. M., & Hagoort, P. (2004). ERP effects of subject–verb agreement violations in patients with Broca’s aphasia. *Journal of Cognitive Neuroscience*, *16*, 553–576.
- Waters, G. S., & Caplan, D. (1996) The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*, 761–772.
- West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*, 1024–1037.
- Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*, 165–168.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, *39*, 483–508.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288.
- Ziegler, J. C., Besson, M., Jacobs, A. M., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience*, *9*, 758–775.

## Chapter 18

### Discourse Comprehension

*Rolf A. Zwaan and David N. Rapp*

#### 1. INTRODUCTION

Consider the following children's riddle<sup>1</sup> (and please do not read ahead until coming up with an answer):

- How do you get an elephant into a refrigerator?

The answer to the riddle, quite simply, is you open the fridge, put the elephant inside, and close the door. This solved, consider another riddle:

- How do you get a giraffe into a refrigerator?

Readers might be tempted to reuse the previous answer for this second riddle, but this turns out to be too simple. The correct answer is you open the fridge, take out the elephant, put the giraffe inside, and close the door. Now, a third riddle:

- All of the animals are going to a meeting held by the king of the jungle. Only one animal does not come. Which one is it?

This answer to this riddle, of course, is the giraffe. After all, it is still in the fridge. A final riddle:

- How do you get across a river where dozens of crocodiles live?

By this time, the astute reader realizes that a solution is going to require thinking about answers from the previous riddles. With that in mind, the answer, of course, is just swim. After all, the crocodiles are at that important meeting with the king of the jungle and all of the other animals.

---

<sup>1</sup>We thank Isabel Zwaan for bringing this riddle to our attention.

These four children's riddles make up a single set, which derives its effect from the reader's (or listener's) inclination to treat each riddle as referring to a separate situation. To answer them as they were intended, the riddles must be treated as connected discourse that describes a single situation with a single set of animals in a single environment. Each subsequent riddle requires the reader to consider the described events with respect to what they already know (and have updated in memory) as a function of the previous riddle. People often fail to provide correct answers to these riddles for a variety of reasons, and these incorrect responses can provide insights into some of the cognitive processes at work as we attempt to comprehend discourse. Most notably, comprehension necessitates the application of prior knowledge in combination with the encoding of information currently in discourse focus.

The example above serves to highlight some of the critical features and processes of discourse comprehension that we will discuss in this chapter. Underlying this example is the basic notion that comprehension involves the construction and application of an *integrated mental representation of the described events*. Those events can be read, heard, seen through some presentation, or even experienced firsthand. Comprehension requires building connections between those events and existing representations in memory. For example, the situations described in the first two riddles take place in *chronological order* and are *temporally contiguous*. The elephant was put in the refrigerator first; unless the elephant is removed, the giraffe cannot be stored in the fridge. To answer the riddle, the reader must connect these two situations. Comprehenders routinely assume that consecutively described events take place in the order in which they are described, and that no unmentioned event will have occurred between them (otherwise such an event would have been described or the omission of events would have been indicated by a time shift such as "an hour later"). Thus, the two events should be connected with each other and, given expectations about chronological order, those events should be assigned a predictable temporal association. In fact, a growing body of evidence suggests that comprehenders routinely and/or strategically keep track of protagonists, objects, locations, and events to build useful associations.

*Linguistic cues* also provide critical information that can either facilitate or hinder comprehension. The second riddle in the above example makes clever use of one type of linguistic cue to wrongfoot the listener. The indefinite article "an" normally functions as a cue to introduce a new entity into the developing situation, whereas the direct article "the" is interpreted as a cue to search memory for an appropriate referent. Thus, the indefinite article suggests that the refrigerator mentioned in the second riddle is not the same as that mentioned in the first. This promotes the assumption that the described situations are separate, whereas the solution actually requires listeners to think of these riddles as referring to a single situation. Consider also the use of these cues for the fourth riddle: by way of a categorical inference, crocodiles, being animals, should be included under "all animals" and therefore inherit the feature of attending the meeting. Linguistic cues such as definite and indefinite articles can either create or reduce ambiguity, and thereby influence comprehension.

In this chapter, our focus is primarily on text processing; this is based on our overarching goal of examining how processes of memory and language influence general comprehension. We begin at the macro-level by discussing the broad impact of different types of discourse genres. Genres can differ greatly, not just in the types of information they describe, but also as a function of the expectations a comprehender has for that type of genre. We discuss this issue in more detail in Section 2. In Section 3, we consider similar issues at the linguistic level, by discussing the influence of linguistic cues on comprehension.

The example also demonstrates that neither genre-specific knowledge nor more specific linguistic cues are sufficient for describing comprehension processes. Comprehenders invoke various types of background knowledge to understand described situations. In fact, the use of background knowledge is necessary for discourse comprehension. For instance, the answer to the fourth riddle only makes sense if we activate our knowledge that crocodiles should be included in the real-world category of animals. The example also shows that comprehenders must, at times, ignore and revise their prior knowledge to make sense of discourse. Refrigerators are normally not big enough to house elephants and giraffes, and animals as a rule do not convene to attend jungle conferences. To answer these questions, we must ignore our real-world expectations about these facts. Skilled comprehension of various types of discourse, for instance fairy tales and science fiction, involves discounting our expectations about the reality of the described situations and anticipating specific violations of our normal expectations (but see Prentice, Gerrig, & Bailis, 1997, for another view of suspended disbelief). For example, in fairy tales, we might expect characters to possess magical powers or animals to speak, but not expect to read about spaceships or other futuristic technologies. Conversely, in science fiction stories, we might expect such futuristic technologies but perhaps no talking animals or wizardly wands. Riddles derive most of their effects from the fact that the listener does not know beforehand which violations to expect, and what to update in memory. Thus, the acquisition and application of background knowledge is an important issue in discourse comprehension.

For the past three decades or so, the consensus view in the literature on discourse processing has been that comprehension arises out of both information provided by language experience (e.g., linguistic cues in text or speech) and information brought to the experience by the reader (e.g., background knowledge). However, more important than these separable components is the interplay between them, which yields a mental representation of the described situation, termed as mental model or situation model (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Considerable work has outlined the contents, structure, and construction of situation models. We discuss these issues in Section 5.

In Section 6, we further discuss the nature of the mental representations that may be invoked during discourse comprehension. Until recently, most views of discourse comprehension have suggested that memory is largely abstract, consisting of propositional, amodal representations. Current work has started to question this view, by examining whether these representations should be thought of as grounded in perceptual and motor

processes and representations. This topic provides a stepping stone for considering future developments in the field. Finally, we close with a return to our original riddle set, intended as an example of how the processes and issues we describe are critical to everyday discourse comprehension.

## 2. DISCOURSE GENRES

At a general level, discourse comprehension is shaped by discourse genre. Discourse genres can be categorized as a function of discourse topic, formality, delivery system, and author or speaker goals and intentions. Three of the most well-studied genres include narrative, expository, and procedural discourse. Other categorizations have been proposed as well for texts and genre subsets (e.g., Meyer & Freedle, 1984).

Narratives have often been associated with fiction, although they can include nonfictional accounts (e.g., historical narratives such as *John Adams* by David McCollough). What differentiates narratives from other genres is that they typically describe a series of events involving a protagonist attempting to overcome obstacles and accomplish a goal (e.g., Mandler & Johnson, 1977; Propp, 1968; Stein & Glenn, 1979; Trabasso & Sperry, 1985; van den Broek, 1988). Narratives are often defined by the causal structure of their events (e.g., Trabasso & van den Broek, 1985; van den Broek, 1990); they contain sequences of events that lead, by necessity and sufficiency, to later events and can be traced back, causally, to early sequences in the plot. Causal structures in narratives are, in many cases, so familiar that readers may have expectations for how the narrative will unfold. These expectations can influence comprehension. For example, readers know that Greek tragedies are likely to end with the death of a main character, mysteries contain bits of information that may be useful for guessing the identity and intent of criminals, and biographies will describe some of the major events in a subject's life, likely in chronological order. Thus, readers can rely on their knowledge about narrative subgenres to build strategies that may facilitate comprehension.

Expository texts are usually produced with the goal of explanation or persuasion. Examples are textbooks, encyclopedias, and other materials that describe facts or principles (Britton & Black, 1985; Goldman & Bisanz, 2002). Examples also include articles in newspapers and magazines (although many of these articles have narrative elements) and the chapters in this *Handbook*. As with narratives, knowledge of the expository genre can provide readers with strategies for encoding the material. For instance, knowing that you are going to read an article in an experimental psychology journal leads you to expect an abstract, followed by an introduction, followed by a method section, followed by a results section, followed by a discussion section. On the other hand, if you read an article in *Science* or *Nature*, you might expect to find the method section at the end of the article. Specifically because expository texts are often implemented in learning settings or situations, readers can use strategies to decide which information may be critical for adequate comprehension, and hence focus their attention on that material (for example, ignoring or focusing on method sections in articles).

Procedural discourse is also associated with explanations, but unlike expository materials, procedural materials are usually structured as sets of directions for completing certain activities. Examples of procedural discourse include the online manual for Microsoft Word and your car maintenance guide. These situations involve the description of a sequence of acceptable or expected actions to be executed in order to successfully perform (and complete) a task. Unlike in narratives, the causal sequence is often explicitly detailed and is a function of particular actions to be completed by the comprehender. For example, in do-it-yourself furniture guides, assembling a desk is a matter of putting together the components one piece or step at a time. Procedural discourse not only includes directions for things to do, but often also includes descriptions of what not to do (e.g., do not put metal objects in the microwave).

These three genres have received attention in experimental research on discourse comprehension, with narratives obtaining the lion's share. Knowledge or expectations with respect to genre can guide the cognitive activities that underlie comprehension processes as well as the ways in which readers represent discourse information in memory. These expectations can be set by prior experiences, or the tasks and goals associated with a particular text (Horiba, 2000; Narvaez, van den Broek, & Ruiz, 1999; Schmalhofer & Glavanov, 1986; van den Broek, Lorch, Linderholm, & Gustafson, 2001; Zwaan, 1991, 1993, 1994). For example, when assigned a particular reading purpose (e.g., to study or for enjoyment), readers may focus on different statements as well as the processes they engage in (e.g., evaluation, rereading, inference production) or deem necessary for comprehension (Linderholm & van den Broek, 2002). Similarly, the genre of a text can modify readers' propensities for generating inferences, connecting statements across a text, and building strong memory for what has been read.

One influential view with respect to genre, the material-appropriate processing framework, suggests that particular text types can lead readers to process texts differently (Einstein, McDaniel, Owen, & Coté, 1990; McDaniel, Einstein, Dunay, & Cobb, 1986). This view has focused on how narrative and expository texts lead readers to focus on either the relations among concepts, or on individual facts, respectively. It was later expanded to include reader-initiated processes (e.g., those not directly invoked during or after reading experiences) by examining how expectations for text genre, along with explicit textual genre cues, influence the processes and products of comprehension. There is evidence that expectations about genre influence how readers process and remember texts (Zwaan, 1991, 1993, 1994). Importantly, these findings are not simply a function of deeper processing; readers approached these texts in qualitatively different ways that influenced the types of representations that they formed during reading (e.g., Wolfe, 2005). Knowledge about the goals of the author, the purported purpose of the text, and the topic to be covered, are all potentially meshed into the notion of text genre. Readers often rely on such information to help them comprehend texts.

Genre information is sometimes explicitly provided, as in the case of bookstores and libraries organizing books into sections. However, it can also be cued by other, more implicit information, such as title, potential readership, availability, price, and use of terminology.

Beyond these macro-level cues, there are also micro-level cues *within* a text that directly influence the processes and products that develop during reading experiences.

### 3. LINGUISTIC CUES

Linguistic cues directly influence the ways that readers process and comprehend discourse. For example, they guide the comprehender with respect to the activation of concepts. Some of the cues that guide such activation patterns are quite explicit (e.g., “Now I am going to describe what you should know for the test.”), but others are more subtle (e.g., Bangerter & Clark, 2003). For example, in the sentence “I saw a student enter the lab,” the indefinite article “a” is a cue to the comprehender to activate a mental representation of a new protagonist. On the other hand, in “I saw the student enter the lab,” the comprehender is prompted to reactivate or keep activated a mental representation of a protagonist that was introduced before. Finally, in “I saw this student enter the lab,” the use of the pronoun “this” suggests that the student in question will be the focus of the narrative (Gernsbacher & Shroyer, 1989).

As these examples show, linguistic cues often function to help comprehenders integrate incoming information. This is thought to promote the generation of a coherent mental representation. In this sense, language can be thought of as a set of processing instructions (Gernsbacher & Givón, 1995; Givón, 1992), indicating to comprehenders what information should be activated, what information may no longer be important (and therefore should be deactivated), and where to focus attentional resources in the immediate discourse. We discuss two types of linguistic cues – lexical cues and structural cues.

Connectives such as “therefore,” “and then,” “but,” and “however” are lexical cues that provide information as to how particular associations should be built between linguistic units (e.g., Halliday & Hasan, 1976). They serve as cues for maintaining particular concepts in various states of activation over the course of the comprehension process and are therefore critical for coherence (Britton, Glynn, Meyer, & Penland, 1982; Sanders & Noordman, 2000). Causality, for example, is often marked with terms such as “because,” because such terms detail how a sequence of events leads to a particular outcome or state. Connectives facilitate memory for textual information, particularly in cases involving causal descriptions (Caron, Micko, & Thuring, 1988; Millis & Just, 1994; Myers, Shinjo, & Duffy, 1987; also see Millis, Graesser, & Haberlandt, 1993, for a discussion of lack of facilitation in cases involving expository texts). Current work has also assessed how connectives influence reading times, with evidence suggesting that appropriately structured connectives facilitate moment-by-moment comprehension (Deaton & Gernsbacher, in press).

Pronoun referents, synonyms, and direct/indirect markers are also cues that modulate the activation of concepts during comprehension. Anaphora provide information on how incoming information should be integrated with the active memory representation. For example, in the statement, “Jay wanted to work but he kept making phone calls,” “he” might be intended to refer to Jay. Anaphor resolution is the process by which a particular referent

in the memory representation is selected for a current concept. To understand discourse, comprehenders must *resolve anaphors*, i.e., determine which referent is targeted by a specific anaphor.

In addition to lexical cues, there are syntactic and macro-structural cues. For example, the position of a sentence in a paragraph or a phrase in a sentence often serves as an integration cue. With roots in linguistics (e.g., Chafe, 1976; Halliday, 1967), this issue has been addressed by examining the pragmatic functions of first mentions. As examples of the importance of first mention, the first sentence of a paragraph traditionally conveys the main idea; introductory lectures often begin with a preparatory statement on the information to be described (e.g., Mayer, 1984). Work on discourse comprehension has assessed the ways in which first mentions are processed, and their resulting effects in memory. First mentions can attract reader attention, convey information about the topic and why it is important, and provide comprehenders with introductory material with which to consider what they might already know and how it potentially relates to the discourse (Clark & Clark, 1977; Gernsbacher, 1990; Givón, 1986). Thus, first-mentioned material receives privileged status during comprehension, and sets the stage for the encoding and retrieval of subsequent information (e.g., Lorch & Lorch, 1996).

Comprehenders rely on this knowledge in developing strategies for comprehension. One such strategy, the given–new strategy (Clark & Haviland, 1977; Haviland & Clark, 1974), suggests that listeners and readers divide linguistic units into given and new segments. The given information refers to what is already known. This can help comprehenders activate existing representations with the goal of understanding the new set of information. New information is comprehended in light of associations and relations with old, given information. Originating from the study of conversation (e.g., Grice, 1975), the view is that language producers and comprehenders set up implicit contracts that align with this given–new strategy. That is, a speaker or writer will refrain from introducing new information prior to establishing its relation to old information, unless such an introduction serves a specific purpose (e.g., emphasizing a new issue or explicitly changing the topic). Cues for indicating given–new material can range from explicit reminders to, in some cases, changes in intonation (e.g., Bock & Mazzella, 1983; Terken & Nootboom, 1987). Thus, coherence is a direct function of the degree to which comprehenders can connect information they are currently processing, with prior information either in the linguistic stimulus or in memory. The given/new strategy, then, works directly in line with notions of effective coherence.

Lexical and syntactic cues often work in conjunction with each other and in conjunction with background knowledge. For example, in the sentences “Paul and Markus are going swimming today. I hope he has his water wings,” “he” will likely be associated with “Paul” given the lack of prior context. The first–mentioned concept, in this case Paul, is most likely to fill the anaphor slot (Gernsbacher, 1989; Gernsbacher & Hargreaves, 1988). Of course, background knowledge can also influence this process. If, for example, the reader knew that Markus was a 2-year child going to a pool with his father Paul, the reader might be more likely to link “he” with “Markus.” Thus, the activation of concepts across linguistic inputs is



a function of lexical and grammatical cues grammar, contextual constraints, and background knowledge (Järvikivi, van Gompel, Hyönä, & Bertram, 2005).

#### 4. BACKGROUND KNOWLEDGE

Consider the following passage:

The procedure is actually quite simple. First you arrange items into different groups. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to a lack of facilities that is the next step; otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do a few things at once than too many. In the short run it may not seem important but complications can easily arise. A mistake can be expensive as well. At first, the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then, one never can tell. After the procedure is completed one arranges the materials into different groups again. Then they can be put in their appropriate places. Eventually, they will be used once more and the whole cycle will have to be repeated. However, that is part of life.

In a classic study, Bransford and Johnson (1972, 1973) read this passage to three groups of participants. Each group rated the passage on comprehensibility and also recalled it. The *No Topic* group performed these tasks without additional information. These participants rated the passage as incomprehensible and exhibited poor recall. The *Topic Before* group was first given the topic “Washing Clothes” before being read the passage. These participants rated the passage as much more comprehensible than the *No Topic* group and recalled more than twice as much of the presentation. Finally, the *Topic After* group was given the topic following presentation of the passage. This did little to help their understanding; their performance was similar to that of the *No Topic* group. This study makes two important points. First, background knowledge facilitates understanding. Second, for background knowledge to facilitate understanding, it must be activated. As we discussed in the previous section, linguistic cues are important with respect to activating the relevant background knowledge. In the washing clothes example, the passage title is such a cue. The importance of background knowledge has been demonstrated in a study on German school children (Schneider & Körkel, 1989; Schneider, Körkel, & Weinert, 1989). Third graders with background knowledge about soccer outperformed seventh graders without this background knowledge in recall of a text about soccer. Given that we may assume that the seventh graders had superior reading skills, this result suggests that relevant background knowledge can compensate for potential differences in reading ability (Recht & Leslie, 1988).

Knowing *that* background knowledge affects comprehension is one thing, but knowing *how* is another. In order to address this issue, we need to consider how background knowledge is organized in long-term memory (LTM). Researchers in the field of

artificial intelligence (AI) have occupied themselves with this question in their quest to build intelligent computer systems that can perform human tasks such as problem solving and language comprehension. One popular format of knowledge organization is the semantic network (Collins & Quillian, 1969; see also the chapter in this volume by Ober and Shenaut), which captures facts in the form of nodes in a network and the links between them. For example, the fact that a giraffe is a mammal is captured by the nodes for giraffe and mammal and an ISA link (literally meaning “is a”) that connects these nodes. The fact, or proposition, that a giraffe has a long neck would be captured by the nodes giraffe and long neck and a link between them labeled for property. In propositional notation, this fact can be represented as [HAS[GIRAFFE, LONG-NECK]]. If a node in the network is activated, for example, because the corresponding word is read, it will send activation to its nearest neighbors, which in turn will send activation to their nearest neighbors, and so on. During each cycle, less activation will be transmitted, such that activation gradually dissipates. Thus, LTM is not completely activated all of the time. Through this process of spreading activation, the network (whether computer or human) is able to provide an affirmative response assessing the validity of statements such as “A giraffe is a mammal.”

In order to understand discourse, more structured knowledge defined by higher-order organizations may be required. For example, stories are often set in stereotypical locations (e.g., a living room, an office) and involve stereotypical event sequences. To represent this knowledge, AI researchers have developed knowledge–representation structures such as frames (Minsky, 1975) and scripts (Schank & Abelson, 1977). A restaurant script allows us to understand a story like

Donald and his third wife entered Mario’s. When they were seated, Donald declared himself unhappy with his table, which, located near the door, was not conducive to a romantic conversation. He requested a table in the back of the restaurant. He studied the menu and ordered ossobucco for two. The food was great. Instead of a dessert, they ordered cappuccino. Donald left a big tip.

This story leaves out quite a bit of detail. For example, it does not specify who seated Donald and his wife, who took his order, who prepared the food, who delivered it to his table, and who received the tip. However, this presents little difficulty to a reader equipped with a restaurant script. Among other things, the script contains slots for roles of participants. Because of this, the comprehender already knows that people visit restaurants to enjoy a meal and drinks, that greeters lead patrons to their table, waiters take orders, waiters deliver food and drinks, waiters are the recipients of tips, and that cooks prepare the food. By invoking this script, the storyteller can rely on the comprehender to supply the missing information. Schank (1982) noted that scripts might be overly rigid, given that stories are rarely if ever told if they completely follow the script. After all, in describing scripted information, we are not telling the listener anything they did not already know. In fact, we usually feel compelled to tell a story if the events somehow violate a script; for example, if our food was not served by a waiter, but by a chimpanzee on a unicycle. Schank (1982) also proposed more flexible knowledge–organization structures, such as memory–organization packages.

However, symbolic knowledge structures, even those proposed by Schank (1982), are still not flexible enough to explain human cognition. Therefore, cognitive scientists have turned their attention to neural networks – computer models that are roughly based on our understanding of the human nervous system. In these neural network models (see also the chapter in this volume by Seidenberg and MacDonald), knowledge is not programmed by the experimenter, but is acquired by the system as it processes input and receives feedback on its performance (Rumelhart, McClelland, & The PDP Research Group, 1986). In supervised learning paradigms, the output produced by the system (for example, an affirmative response to a verification statement) is judged by the experimenter to be correct or incorrect. The system incorporates this feedback by adjusting the weights on links between the nodes in its hidden layer. This adjustment may lead to a different categorization of the same input on a subsequent trial. The system has now “learned.” Connectionist models are much more flexible knowledge–representation systems than symbolic systems, which is exemplified by their ability to handle incomplete or noisy input.

Although connectionist models acquire knowledge structures themselves, they require a programmer to provide a stimulus, with the model passively receiving it. The model itself has no direct way of interacting with the world. A relatively recent development in cognitive science is to embed neural networks in robots that have sensors and effectors and are thus able to interact with their environment and learn from these interactions (Brooks, 1992; Pfeifer & Scheier, 1999). For example, a system has been developed that learns words via sensorimotor processes (Roy & Pentland, 2002). As we will discuss in Section 7, perceptual and motor knowledge may play an important role in language comprehension.

Symbolic and neural network architectures have influenced theories of discourse comprehension. As one case, text comprehension researchers have studied the role of scripts in language comprehension, focusing on the retention of scripted versus nonscripted information in stories (e.g., Graesser, Woll, Kowalski, & Smith, 1980). A very influential model is Kintsch’s construction–integration (CI) model (Kintsch, 1988, 1998), which combines knowledge structures such as propositions with the connectionist mechanism of constraint satisfaction (see chapter by Seidenberg and MacDonald). According to the CI model, comprehension involves two phases. During the construction phase, the verbal input activates knowledge in an unconstrained fashion. For example, in the sentence “During the earthquake, the mint collapsed,” the word “mint” activates both its *candy* and its *building* meaning, even though only the latter is relevant in the context of the sentence. During the integration phase, the model settles on an interpretation of the sentence via a constraint-satisfaction mechanism. In this case, for example, the *building* meaning will receive stronger support from the context of the sentence than the *candy* meaning; after all, earthquakes do not usually cause pieces of candy to collapse. Thus, nodes representing *earthquake* and *collapse* will send more activation to the node coding for the *building* meaning than to the competing node coding for the *candy* meaning. As a result, the *building* node will remain activated, whereas the *mint* node will gradually become deactivated. The model is updated in cycles. Once the amount of change between cycles is below a certain predefined threshold, the network is deemed to have settled, and it is

considered to have “understood” the sentence. The CI model has been used to simulate the online comprehension of text, text recall, and the recognition of textual materials (see Kintsch, 1998, for an overview). In more recent work, the CI model has been interfaced with a latent-semantic analysis (LSA) system, which provides information about the associations between words based on the similarities of the contexts in which they occur in large corpora of texts (Landauer & Dumais, 1997). Kintsch (2000) has applied this combination in the context of metaphor comprehension.

How knowledge is *organized* is only part of the question. An equally important issue is how knowledge is *retrieved* during comprehension. One proposal is that knowledge is retrieved via an inferencing process. Articles on discourse comprehension often refer to inference generation when they discuss how information from the text is supplemented with background knowledge. This term is somewhat misleading in that it suggests a deliberate and slow process, akin to reasoning. However, as our review of knowledge representation shows, information is usually thought of as being activated automatically, or at least passively (O’Brien, Rizzella, Albrecht, & Halleran, 1998). In this sense, it is more appropriate to view inference generation as knowledge activation or integration (Cook, Limber, & O’Brien, 2001; Kintsch, 1993).

There are several types of inferences (van den Broek, 1994). For example, there are connective (or bridging) and elaborative inferences. Connective inferences provide a way of connecting two successive text statements. For example, consider the following two sentences: “Murray poured water on the bonfire. The fire went out.” The two statements can be integrated by activating the knowledge that water extinguishes fire (Singer, 1993; Singer & Halldorson, 1996). Thus, “water extinguishes fire” functions as a connective inference here. An elaborative inference is the activation of knowledge that augments the mental representation of the described situation, but which is not needed to integrate statements. Instrument inferences are a type of elaborative inference. For example, in “John let the tomato soup cool off for a while. Then he ate it,” the most plausible instrument inference would be that John used a spoon to eat the soup. This information is not necessary to integrate the two sentences, but it is a plausible inference to make, given that spoons are normally used when eating soup. Another example of an elaborative inference is a predictive inference. For example, in “The tired speaker finished his talk and walked over to a chair,” it is possible, based on what we know about giving lectures, about being tired, and about the function of chairs, to infer that the speaker is going to sit. There is strong evidence that comprehenders generate connective inferences (e.g., Graesser, Singer, & Trabasso, 1994). The evidence for more elaborative inferences, though, tends to be mixed. For example, whether or not predictive inferences are generated appears to depend on several factors including contextual constraints, individual differences, processing strategies, task instructions, and reader expectations (e.g., Allbritton, 2004; Calvo, 2000; Calvo & Castillo, 1996; Cook et al., 2001; Fincher-Kiefer, 1996; Klin, Guzmán, & Levine, 1999; Linderholm, 2002; McKoon & Ratcliff, 1986, 1992; Murray, Klin, & Myers, 1993; Rapp & Gerrig, 2006; van den Broek, 1990, 1994; Weingartner, Guzmán, Levine, & Klin, 2003). Even reader preferences, the wishes and desires readers build for story characters and events, influence the construction and application of inferences (Allbritton

& Gerrig, 1991; Gerrig & Rapp, 2004; Rapp & Gerrig, 2002, 2006). At least partially as a function of this large number of potential influences, there is still considerable debate about the particular circumstances that lead to automatic activation of predictive inferences. Not surprisingly then, evidence suggests that such activation occurs under a limited set of circumstances, rather than across all discourse conditions.

This debate over predictive inferences is directly related to a more general issue; the extent to which comprehension involves the active construction of a mental representation or a more passive form of knowledge activation. One extreme view would be that comprehension is a very active process, akin to reasoning, in which the comprehender makes a conscious effort to generate bridging and elaborative inferences in order to arrive at a detailed “high resolution” mental representation of the described situation. At the other extreme is the view that background knowledge is retrieved automatically as a function of the processing of incoming stimuli; in this the comprehender passively activates background knowledge, with integration being a function of such passive activation. Neither view is completely supported by the data, which is why most theories fall somewhere in the middle; some state that comprehenders do not indiscriminately generate inferences, but only those that are relevant to their “effort after meaning” (Graesser et al. 1994), whereas others focus on the activation of information that is “easily available” through passive memory processes (McKoon & Ratcliff, 1992; Myers & O’Brien, 1998). A comprehensive model of discourse comprehension likely invokes both sets of processes (see Gerrig & O’Brien, 2005; Guéraud & O’Brien, 2005; van den Broek, Rapp, & Kendeou, 2005, for a discussion).

Additionally, these processes capture two important general intuitions about discourse comprehension. The first intuition is that comprehension often seems incomplete, regardless of whether it involves automatic activation or strategic processing. For example, we often fail to notice inconsistencies in texts (e.g., Barton & Sanford, 1993), or between text and background knowledge (e.g., Kendeou & van den Broek, 2005; Otero & Kintsch, 1992). This suggests that comprehension is a “sloppy” process. The second intuition is that language comprehension, especially narrative comprehension, often produces a sense of experiential richness. Again, whether information is activated as a function of dynamic spreading activation processes driven by either (or both) a ‘dumb’ passive process or an active search for meaning, it is clear that comprehenders can become engaged in their discourse experiences. It is difficult to see how theories that view comprehension as involving a small number of abstract representations, whether automatically activated or strategically generated, can explain why children and adults become engrossed in popular stories such as the *Harry Potter* books and *One Hundred Years of Solitude* by Gabriel Garcia Marquez.

The role of background knowledge in the comprehension process cannot be overstated (Kendeou, Rapp, & van den Broek, 2004). Comprehenders rely on this knowledge, even in cases for which it is incorrect and may create problems. For example, students often hold misconceptions, and these misconceptions are resistant to change, even in the face of refuting evidence (e.g., Diakidoy & Kendeou, 2001; Diakidoy, Kendeou, & Ioannides,

2003; Guzzetti, Snyder, Glass, & Gamas, 1993; Kendeou & van den Broek, 2005). Students attempt to understand material in line with these faulty beliefs, rather than spontaneously engaging in processes of conceptual change to revise their beliefs and mental models (DiSessa, 2002; Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou, 2002; Vosniadou & Brewer, 1994). Thus, background knowledge from LTM serves as the scaffold for newly encountered information, regardless of the validity of that knowledge.

## 5. SITUATION MODELS

The most basic purpose of discourse is to convey information about a state of affairs in the real or a fictional world. Accordingly, the comprehender's usual goal is to achieve an understanding of the described situations. As we have suggested, the comprehender relies on linguistic cues in the discourse and his/her prior knowledge to achieve understanding (McNamara & Kintsch, 1996). Integration of information from the discourse and the comprehender's knowledge and cognitive activities is necessary for successful comprehension (as well as for forming a representation of the experience that could be applied at a later time point).

Because people apply their discourse comprehension skills for many different goals (e.g., browsing a magazine; listening for key facts in a lecture; evaluating the views of a political candidate), it is likely that they can build a variety of types of mental representations in the process. One influential theory has proposed that discourse comprehension involves at least three levels or types of mental representations (Schmalhofer & Glavanov, 1986; van Dijk & Kintsch, 1983).<sup>2</sup> The most basic level is the surface structure, a mental representation of the exact text read. Surface representations decay rapidly from memory (e.g., Anderson & Bower, 1973; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Sachs, 1967, 1974), although they tend to be more resistant to decay when pragmatically relevant (e.g., Keenan, MacWhinney, & Mayhew, 1977; Murphy & Shapiro, 1994). The second level, the textbase or propositional representation, contains idea units explicitly stated in the text, along with some bridging inferences. The textbase representation is also described as gist-like memory (Kintsch & van Dijk, 1978). At this level, readers represent the ideas described in the text, but not every single word or concept contained therein. The highest level of representation, often viewed as essential to comprehension, is the situation model (e.g., Hess, Foss, & Carroll, 1995; van Dijk & Kintsch, 1983; Zwaan, 1999b; Zwaan & Radvansky, 1998). At this level, readers represent information described by the text, activating knowledge that goes beyond what was explicitly stated. Readers often rely on their prior knowledge to fill in gaps in the text, as well as to run "mental simulations" of the information (Kahneman & Tversky, 1982; also see Barsalou, 1999, for a recent discussion of mental simulation), and these processes constitute activity at the level of the situation model.

---

<sup>2</sup>Other theories have suggested two levels as sufficient, although these levels likely combine elements of the three-level view we will describe, e.g., Johnson-Laird (1983)

An example of representation at each of these levels will help clarify this taxonomy. Consider the following sentence: "Sid searched for a new apartment on the North side of Chicago." A surface representation would contain every word in that sentence; essentially it would be a verbatim replication of the sentence. A textbase representation would contain the major idea units described in the sentence. This would include Sid (as agent) searching for an apartment, Sid searching Chicago, and so on. Thus, readers at this level would remember the concepts conveyed in the sentence (as a sample recall, "Sid was looking for a place to live in Northern Chicago") but not necessarily every word. At the situation level, the reader might build expectations about the type of apartment Sid is looking for as a function of prior knowledge about Chicago, inferences about the potential neighborhoods Sid might be exploring, and perhaps even elaborations with respect to why Sid is moving. Thus, it is readily apparent that readers can build a variety of representations for what they read as a function of the text proper, the concepts conveyed by the text, and relevant information from prior knowledge (and likely, some integration across these levels).

The situation model is considered central to comprehension. The role of situation models has been considered in written discourse as well as in spoken discourse, such as dialogue (e.g., Pickering & Garrod, 2004). However, most of the research has focused on (narrative) text comprehension. A large body of work has delineated some of the elements or dimensions that readers may encode into situation models. Five such dimensions of situation models have been distinguished based on the event-indexing model; space, time, entity (protagonists and objects), motivation, and causation (Zwaan & Radvansky, 1998). Evidence for these dimensions is consistent with several general assumptions (Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b). Readers attempt to integrate each incoming event with the current mental representation in working memory on each of the five dimensions. In doing so, readers assume that the incoming event shares an index with the active mental representation. A shift on a dimension should lead to an increase in processing time, because a change in the described situation has taken place. Furthermore, the resulting memory representation should reflect this change by showing weaker links between events that are separated by a change than events that are not. These assumptions are entirely consistent with more general theories of language processing, such as the CI model (Kintsch, 1988,1998) and the structure-building framework (Gernsbacher, 1990). In fact, recent work suggests the utility of integrating event-indexing and structure-building frameworks into a cohesive theory, by appealing to the former as a description of the contents of situation models, and the latter as a process model of cognitive activity during reading experiences (e.g., Rapp & Taylor, 2004). Regardless of the nature of that cognitive activity, the event-indexing model serves as a guide for determining factors that influence both the structure and content of memory for discourse.

An example from Zwaan (1996) illustrates these ideas. This study examined how readers' tracking of the passage of time, one of the five dimensions in the event-indexing model, can affect online comprehension and LTM of text. Time shifts are common events in narratives, for example, "Later that day..." or "A few weeks later..." When they occur, the speaker or writer can use the shift to omit events that are deemed irrelevant to the situation. For example, we usually do not need to know that the main character in a novel has

brushed his teeth or tied his shoes. In fact, when such mundane events are actually reported in a story, it is usually a clue that they will be relevant later (e.g., in a murder mystery). The omission of these types of events in a story can be signaled by a time shift such as *an hour later*. Zwaan created stories that came in three versions. In the *moment* version, the temporal adverbial *a moment later* was used. The assumption was that a moment does not constitute a time shift of any temporal magnitude, and thus maintains the current time frame. In the *hour* and *day* versions, the critical phrases were *an hour later* and *a day later*, respectively. In these cases, there should be a shift to a new time frame, and indeed, more activities can take place within these longer shifts than the shorter *moment* shift. Following one of these shifts, participants were asked to identify whether a probe word had appeared in the story. Probe words were selected from descriptions preceding the temporal adverbial. The findings were consistent with the previously described assumptions of the event-indexing model. Responses to probe words were significantly longer in the hour and day versions than in the moment versions, suggesting that the time shift made the previous event less accessible. A primed-recognition task performed after all of the stories were read showed that when events occurred within the same time frame, they showed more priming than when these same events were separated by a time shift (both in the hour and day cases). This suggests that events from the same time frame are more strongly connected in LTM than events from different time frames.

Considerable research has demonstrated similar results across the dimensions described in the event-indexing model, specifically with narrative texts.

*Space.* Much of the early work on situation models focused on how the spatial features of descriptions influence memory for texts. Nearly every text has a setting or set of locations in which the described events take place. For instance, texts often detail information about locations, objects in those locations, movement of characters through locations, and other spatial relationships between elements in the text. Classic work on situation models has assessed readers' tracking of spatial information by examining how shifts in space impact the accessibility of information from memory (Bower & Morrow, 1990). For example, participants were asked to memorize a map illustrating a series of rooms and their linear pathways, along with objects contained in each of the rooms (e.g., Morrow, Bower, & Greenspan, 1989; Morrow, Greenspan, & Bower, 1987). After memorizing the map, participants read a story that described a character moving through the rooms. At particular points in the text, participants were presented with pairs of objects and their task was to determine whether those objects were located in the same or in different rooms. Across multiple studies, evidence supports the spatial distance effect; participants take longer to identify objects that are further away from the currently described situation or events in the text. This result suggests that information in the same spatial frame (e.g., a particular room from the map) is strongly activated in memory, and shifts from that spatial frame lead to decrements in response accuracy and speed. This finding is robust across a variety of methodologies and testing conditions (e.g., Bower & Rinck, 2001; Rinck & Bower, 1995, 2000; Rinck, Bower, & Wolf, 1998; Rinck, Hähnel, & Becker, 2001; Rinck, Hähnel, Bower, & Glowalla, 1997; Rinck, Williams, Bower, & Becker, 1996). The prevailing view is that reader focus, guided by the activities and descriptions in currently read text, directly influences the accessibility of



information from memory. This suggests that the spatial concomitants of text descriptions are likely to be tracked and updated as texts unfold. Even in cases without explicit map memorization (e.g., de Vega, 1995; Glenberg, Meyer, & Lindem, 1987), readers monitor the locations of objects and environments.

An important question involves the degree to which spatial models are spontaneously encoded into memory. Evidence suggests that elaborate spatial representations are not formed unless (a) the spatial features of a text are particularly salient or (b) readers are specifically instructed to focus on the spatial descriptions in a story (e.g., de Vega, 1995; Hakala, 1999; Levine & Klin, 2001; O'Brien & Albrecht, 1992; Rich & Taylor, 2000; Taylor, Naylor, & Chechile, 1999; Zwaan, Radvansky, Hilliard, & Curiel, 1998; Zwaan & van Oostendorp, 1993). Thus, the spatial dimension is more or less likely to be tracked as a function of its importance in the situation, especially when it may facilitate comprehension of a particular discourse.

*Time.* Research on the temporal components of texts has grown considerably in the last 10 years. One reason for this increased interest has been the utility of assessing readers' tracking of temporal dimensions without the map-memorization tasks traditionally employed in spatial experiments (e.g., Morrow et al., 1987). Additionally, time is implicitly (if not explicitly) provided in descriptions, mainly through verb tense or aspect. This ubiquitous quality of time provides an interesting opportunity for assessing the ways in which individuals track, remember, and comprehend event-based information.

In comprehending event sequences described in text, people have been found to entertain the *iconicity assumption*; they assume that events are described in chronological order. For example, when reading Julius Caesar's famous dictum "Veni, vidi, vici," we assume that the Roman Consul first came, then saw, and then conquered, in that sequence (Jakobson, 1971). There are various techniques to describe events in ways that deviate from chronological order and temporal expectations. Not only can we omit events from descriptions as mentioned previously, but we can also describe events later (flashbacks) or earlier (flash forwards) than they have actually occurred in a plot. Events can often overlap with each other and even occur simultaneously. Despite these various potential sequences for events, speakers and writers are constrained by the linear nature of language, which forces them to describe events one at a time. In other words, a discourse producer has to map nonlinear events onto a linear structure, whereas the comprehender has to recover the nonlinear temporal relations among these events from that linear structure. This process of recovering the appropriate order from sparse linguistic input has received an extensive amount of work in narrative comprehension.

Linguistic cues help guide readers' mapping of such events into their appropriate described sequences. These cues include lexical information such as time adverbials and grammatical cues such as verb tense and verb aspect. Recent work has examined the role of verb aspect in comprehension (Bestgen & Vonk, 2000; Carreiras, Carriedo, Fernandez, & Alonso, 1997; Madden & Zwaan, 2003; Magliano & Schleich, 2000; Rinck & Bower, 2001; Zwaan, 1996; Zwaan, Madden, & Whitten, 2000). This work shows that

the temporal nature of a verb or verb clause influences the accessibility of the described events from memory. For example, completed events, as compared to ongoing activities, are deactivated in memory.

Expectations regarding the amount of time necessary for initiating and completing activities also influence the comprehension process (Rapp & Gerrig, 2002). The processes involved in generating and applying expectations about temporal activities help outline how event-based constraints influence the structure and contents of temporal situation models (de Vega, Robertson, Glenberg, Kaschak, & Rinck, 2004; Gennari, 2004; Kelter, Kaup, & Claus, 2004; Radvansky, Zwaan, Federico, & Franklin, 1998; Rapp & Taylor, 2004; Speer & Zacks, 2005).

*Entity.* Story characters are integral to unfolding plot. Often plots revolve directly around the interactions among characters and objects in the text. Pronouns are one type of linguistic cue speakers and writers use to help comprehenders track such entities in the described situation. By investigating how people resolve pronouns, researchers have gained insight into how comprehenders track characters and objects in discourse (e.g., Gernsbacher & Hargreaves, 1988; Gordon, Grosz, & Gilliom, 1993; Sanford & Garrod, 1981; Sanford, Moar, & Garrod, 1988). For example, when provided with an ambiguous pronoun, readers tend to resolve the pronoun using the main character of a story; thus, important characters in narratives often receive privileged status (and are tracked accordingly), and thus are readily available during comprehension (McDonald & Shaibe, 2002; Sanford, Clegg, & Majid, 1998).

Comprehenders also encode the attributes and characteristics of characters as they read. In everyday situations, we often generate inferences and accompanying expectations about others' emotional states and personality traits (Fiske & Talyor, 1991; Newman & Uleman, 1990; Uleman, Hon, Roman, & Moskowitz, 1996; Winter & Uleman, 1984). Similar processes occur in text comprehension. For example, participants take longer to read outcomes that are inconsistent, as compared to consistent, with emotions, behaviors, and traits suggested by preceding story contexts (e.g., de Vega, Leon, & Diaz, 1996; Gernsbacher, Goldsmith, & Robertson, 1992; Gernsbacher, Hallada, & Robertson, 1998; Gernsbacher & Robertson, 1992). In one study examining readers' propensities for tracking the traits of characters (Rapp, Gerrig, & Prentice, 2001), participants were asked to read character-driven narratives that contained behavioral evidence for a particular trait (e.g., "[Albert's clothes] were buried under old candy wrappers, crumpled magazines, and some dirty laundry," suggesting the trait *sloppy*). In a second story, the same character behaved in a manner either consistent or inconsistent with the trait. Participants tended to agree with outcomes consistent with implied traits, and also took longer to read outcomes inconsistent with those traits. Importantly, these effects did not obtain when stories failed to provide behavioral evidence for that trait (e.g., instead of the above trait sentence, participants read "Albert's friends had suggested meeting outside the pizzeria adjacent to the movie theater"). Readers, then, use character profiles to generate predictive expectations about future narrative events (e.g., Albrecht & O'Brien, 1993; Peracchi & O'Brien, 2004). Additionally, and in line with the event-indexing model, this suggests

that readers track information about characters in a manner similar to that for time and space.

At what level of detail do readers represent information about characters? Current evidence suggests that representations of characters, particularly their emotions, are relatively superficial (Gygax, Oakhill, & Garnham, 2003; Gygax, Garnham, & Oakhill, 2004). This is consistent with views contending that implicit inferences in general may not be specified at levels akin to that for explicit information (McKoon & Ratcliff, 1986; Singer & Remillard, 2004). The situations under which character-based inferences become more specific remain to be outlined, although they are likely influenced by readers' goals and strategies.

To date, there has been much more work on character-based models than object models. This makes sense, given that narratives are typically centered around characters rather than objects. Nonetheless, a common assumption is that the tracking processes involved in encoding and applying such information would be similar regardless of entity type (and, of course, influenced by their perceived importance in the text).

*Motivation.* Sometimes labeled as intentionality, the dimension of motivation has been assessed with respect to the goals of characters in discourse. In most narratives, goals set the stage for plot; characters attempt to achieve or satisfy some goal, and along the way must accomplish subgoals to complete their activity (e.g., Trabasso & Wiley, 2005). Sometimes, goals may be satisfied (e.g., the hero saves the world), while at other times they may remain unsatisfied (e.g., the villain manages to escape). Indeed, suspense is often set up by creating hurdles that reduce the likelihood of a particular goal being satisfied (e.g., Gerrig, 1989, 1993). Character goals provide justification for character behaviors and guide story plot. Thus, the extent to which readers track goals in texts is critical to theories that assess narrative comprehension.

Comprehenders appear to actively track goals (e.g., Albrecht & Myers, 1995, 1998; Fletcher & Bloom, 1988; Foss & Bower, 1986; Goldman & Varnhagen, 1986; Richards & Singer, 2001; Singer & Halldorson, 1996). Goal information is readily available during reading (e.g., Suh & Trabasso, 1993; Trabasso, van den Broek, & Suh, 1989). Thus, the goals a character has for a particular situation are actively maintained in memory over the course of a reading experience. Indeed, maintaining such goals provides coherence over the course of a text (e.g., Huitema, Dopkins, Klin, & Myers, 1993; van den Broek, Lynch, Naslund, Ievers-Landis, & Verduin, 2003).

Authors often set up circumstances for which multiple goals occur sequentially or simultaneously over the course of some text. Readers can track such multiple goals, even when those goals are hierarchically organized (Suh & Trabasso, 1993). However, in some cases subsequently mentioned goals can reduce the availability of earlier mentioned goals, although the reverse does not seem to occur (Magliano & Radvansky, 2001). Readers appear to show quite sophisticated and dynamic goal-tracking processes. For example, they deactivate goals that have been completed, as measured by accessibility of

the goal from memory, in comparison to newly presented or failed goals (Lutz & Radvansky, 1997; Radvansky & Curiel, 1998). The degree to which goals fluctuate in accessibility may be a function of suppression-based processes, as new goals can reduce the availability of previous goals (Linderholm et al., 2004). Importantly for these cases, neutral information is still generally less accessible than even completed goals, which suggests a hierarchical organization for ongoing goals, completed goals, and nongoals. These studies demonstrate that goals hold an important role in moment-by-moment comprehension, and that they are dynamically updated in memory as plots develop.

*Causation.* The situations described in texts often lead readers to expect other events, or to attribute causes to particular events as a function of background knowledge. Linguistic cues such as causal connectives (e.g., “because” and “so”) have an impact on the unfolding comprehension process and the resulting memory representation (e.g., Caron et al., 1988; Millis & Just, 1994; Traxler, Bybee, & Pickering, 1997). In addition to linguistic cues, the comprehender’s background knowledge is instrumental to forming causal representations. For example, the sentence pair “Murray poured water on the bonfire. The bonfire went out” is not simply a temporal sequence of unrelated events, but conveys the idea that the water actually caused the bonfire to go out. To understand this bridging concept, we need to activate our knowledge that *water extinguishes fire*. This is exactly what comprehenders have been shown to do (Singer, 1993; Singer & Halldorson, 1996). Thus, readers appear to track the underlying causal structure of event sequences.

This example, though, also indicates the degree to which causality is likely to be integrated with other dimensions. Causality is often instantiated through conceptual associations with space, goals, time, and so on. In the above case, causality is associated with the entities “Murray,” “bonfire,” and “water.” Also, Murray’s goal seems to involve extinguishing the fire. Embedded in larger text, we could imagine cases for which this bridging inference might be a function of space (e.g., Murray does not want to set the nearby woods ablaze) or even time (e.g., it was almost midnight and Murray wanted to go to sleep). This suggests that the dimension of causality is often complementary with other dimensions. In fact, one could take this further to make the claim that, although it may be useful for analytic purposes to consider them separately, in practice situational dimensions rarely appear separately; the five dimensions outlined in the event-indexing model are often directly connected or integrated during discourse presentations. We turn to this issue next.

*Interactive dimensions.* It is clear that situational dimensions are not completely independent. For example, if the sequence in the previous example had been “The bonfire went out. Murray poured water on the bonfire,” then a causal interpretation would not have been possible (or easily activated). Accordingly, research has begun to consider the interactivity among dimensions, for example, between causality and space, and between causality and time (Rapp & Taylor, 2004; Zwaan & Radvansky, 1998; Zwaan & van Oostendorp, 1994). For instance, comprehenders are more likely to track and remember spatial information when it is causally relevant or relevant to a character’s goals than when it is not (Jahn, 2004; Sundermeier, van den Broek, & Zwaan, 2005).

Most novels describe multiple characters visiting multiple locations, with characters having diverse goals and characteristics, and all of this occurring within a single plot. Preliminary work on the interactions between dimensions has examined the degree to which readers are more or less likely to track particular dimensions during text experiences. Some evidence points to the dimensions of causality (e.g., Keefe & McDaniel, 1993; Singer & Halldorson, 1996) and character (Rich & Taylor, 2000) as privileged during narrative comprehension. Additional research has suggested that dimensional privilege may be a function of reader goals or text cues for particular reading activities (Taylor & Tversky, 1997). One approach to examining the relative importance of situational dimensions is to investigate the extent to which specific dimensions are impervious to reading instructions. Recent evidence suggests that time and protagonist (and to a lesser extent location) are tracked, even when subjects are instructed to focus on a different dimension (Therriault, Rinck, & Zwaan, 2006). For example, participants remain sensitive to time shifts even when asked to focus on the locations described in texts.

An additional question concerning the interaction between dimensions arises when we consider that particular event dimensions may be correlated (Zwaan et al., 1995a). For instance, space and time are often conceptualized using similar metaphors (e.g., Boroditsky, 2000; McGlone & Harding, 1998). This might suggest that these dimensions can provide complementary information. Although dimensions have traditionally been studied by considering their *supplementary* activity (e.g., what either space or time contribute to comprehension), dimensions may actually have more *complementary* relationships. Information from a particular dimension may not only occur alongside another dimension, but in some cases dimensions may actually inform each other. Thus, investigations of the ways in which dimensions interact, and perhaps become integrated into a cohesive situation model, are of great interest. This has led to a burgeoning area of study examining how multiple dimensions are integrated in situation model construction and application (e.g., Jahn, 2004; Rinck & Weber, 2004; Sundermeier et al., 2005; Zwaan et al., 1998).

As suggested earlier, two dimensions that appear to hold such complementary relationships are space and time. This relationship appears to be interactive and informs the construction of situational representations for texts. In one study investigating this issue, participants read stories describing characters as moving from one spatial location to the next (Rapp & Taylor, 2004). During this movement, characters engaged in activities that could take either a short or long time to complete. This temporal information influenced how readers structured their spatial situation models: participants took longer to recognize start location probes following long activities compared to short activities. Similar work focusing on how readers structure situation models with respect to time (e.g., Zwaan, 1996) shows that readers can also structure their situation models with respect to the interplay of time *and* space. This interplay closely aligns with what we might expect given the degree to which multiple dimensions guide unfolding plot in the diverse texts we commonly read.

*Perspective effects.* The work reviewed thus far is entirely consistent with the idea that the comprehender takes a perspective with respect to the situation conveyed by the

discourse (Zwaan, 1999b). Perspective effects have a long history in research on memory for discourse. Classic work by Bartlett (1932) demonstrated that individuals tended to distort memory toward information in line with their own experiences. After listening to a story entitled “The War of The Ghosts,” British participants tended to recall descriptions including “hunting seals” and “canoes” as “fishing” and “boats.” Thus, the perspective that participants brought to the experimental situation influenced their memory for texts. Similar to this classic study, recent work on perspective effects has examined how readers’ viewpoints may change their interpretation, comprehension, and memory for texts.

Some studies have explicitly asked readers to assume a perspective within the narrative world, for example, that of a burglar or a homebuyer in a story about a home. Readers recall more information that is relevant to their adopted perspective (Anderson & Pichert, 1978) and spend more time reading perspective-relevant than perspective-irrelevant information (Kaakinen, Hyönä, & Keenan, 2002). However, perspective effects can also be brought about by subtle lexical cues, such as the verbs *come* and *go*. For example, when we read or hear “He came into the room,” we are taking the point of view of someone who is already in the room. On the other hand, when we read or hear “He went into the room,” we take the point of view of someone who is outside of the room (Black, Turner, & Bower, 1979).

Recent evidence provides support that readers may actually take the visual perspective provided in a narrative. In one such study (Horton & Rapp, 2003), participants read stories describing characters observing the environment. In the experimental condition, a critical object was blocked from the protagonist’s view by an occluding object (e.g., a curtain). In control cases, the critical object remained visible to the protagonist. Target questions asked participants about the existence of the critical object. Participants took longer to answer questions for blocked objects than unblocked objects. These results suggest that to some degree, readers assumed the visual perspective of characters in the stories. Further support for this conclusion comes from a different study in which participants were presented with a sentence and then were to make speeded responses to nouns denoting parts of the object mentioned by the last word in the sentence (Borghi, Glenberg, & Kaschak, 2004). Crucially, the sentence was used to manipulate perspective; for example, “You are fueling a car” versus “You are driving a car.” The verification of perspective-relevant words (e.g., trunk or horn) was facilitated relative to that of perspective-irrelevant words. These findings align with other, less-visually based studies that demonstrate similar perspective-biased effects. For example, just as distant objects are less accessible in the real world than close objects, words denoting objects far from the protagonist are less accessible than words denoting objects that are close to the protagonist (Glenberg et al., 1987). In this case, reader perspective is based on the location of the protagonist with respect to the rest of the narrative situation, and objects are organized around (or away from) that protagonist.

For the described perspective effects, readers may be constructing representations as if they themselves are involved in the narrative situation (Gerrig, 1993). We discuss this issue further with specific interest in the nature of underlying representations following discourse experiences.

## 6. THE IMMERSED EXPERIENCER

The research on situation models in narrative comprehension is generally consistent with the notion of the comprehender as an immersed experienter (Zwaan, 1999b, 2004) in the narrative world (Gerrig, 1993). Traditionally, such views have been espoused in areas including literary theory and linguistics (Genette, 1983; Duchan, Bruder, & Hewitt, 1995). The strong form of this view suggests that readers actually experience information in a narrative as if they were participating in the activity. Evidence for this view comes from an unexpected source: cognitive neuroscience research on humans (and monkeys) suggests that we understand the actions of others by covertly simulating these actions using the motor programs we ourselves would use to perform the action (e.g., Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004). The brain system involved in this process has been dubbed the “mirror system.” Given that the mirror system includes Broca’s area in humans, it stands to reason that language understanding, like action understanding, might involve the simulation of described actions (Rizzolatti & Arbib, 1998; Gallese & Lakoff, 2005).

The main prediction from the mirror-system theory with regard to language comprehension is that we understand action-related language by covertly activating the motor processes we use to produce these actions. Initial evidence for this prediction has been obtained (Tucker & Ellis, 2004): participants were presented with pictures or words denoting objects and then asked to judge whether the objects were natural or manmade. They made their judgments by manipulating an input device that required either a power grip or a precision grip. Power grip responses were faster to pictures and words denoting objects requiring power grips compared to pictures and words denoting objects requiring precision grips. The reverse was true for precision grip responses. The compatibility effect for words was comparable to that of pictures. This finding suggests that words make available the affordances (Gibson, 1986) of their referent objects. Additional evidence comes from tasks with specific action descriptions (Glenberg & Kaschak, 2002). Participants listened to sentences such as “He opened the drawer,” and were asked to make sensibility judgments. These judgments necessitated button press activities requiring either movement toward or movement away from the responder’s body. An action-compatibility effect was obtained such that responses were faster when the physical response was in the same direction as the movement implied by the sentence. For instance, responses made toward the body were faster after “He opened the drawer” than after “He closed the drawer” and the reverse was true for responses away from the body. Similar findings have been extensively reported prior to these studies in situations involving positive and negative lexical stimuli (e.g., Chen & Bargh, 1999; Wentura, Rothermund, & Bak, 2000).

There is additional evidence for the relationship between language and action. We note that the affordances of referent objects have an immediate influence on sentence processing (Chambers, Magnuson, & Tanenhaus, 2004). Hand shapes, for example, prime the comprehension of sentences describing the manipulation of objects (Klatzky, Pellegrino, McCloskey, & Doherty, 1989). Recent neuroimaging studies have also

produced converging findings. Motor regions of the brain are active during the comprehension of action words (Hauk, Johnsrude, & Pulvermüller, 2004) and sentences (Tettamanti et al., 2005). More specifically, both of these imaging studies found that the areas of activation in the premotor cortex were somatotopically organized, such that sentences about mouth actions, hand actions, and leg actions each activated different areas, which in other studies have been associated with movement in these effectors.

These findings imply a close coupling between language comprehension and motor programs, which is consistent with the notion of comprehension as simulation. This research is also consistent with other views of “embodied cognition,” according to which cognition is grounded in perception and action and relies on the use of perceptual and motor representations, rather than of abstract, amodal, and arbitrary mental representations such as propositional networks or feature lists (e.g., Barsalou, 1999; Glenberg, 1997). A great deal of empirical evidence has been amassed recently showing that visual (and sometimes spatial) representations are routinely activated during language comprehension. This includes visual representations of object shape (Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Yaxley, 2004), orientation (Stanfield & Zwaan, 2001), and motion direction (Kaschak et al., 2005; Zwaan, Madden, Yaxley, & Aveyard, 2004); for more extensive reviews, see Spivey, Richardson, and Gonzalez-Marquez (2005), Zwaan (2004), and Zwaan and Madden (2005). Tanenhaus and Trueswell (this volume) discuss the close link between language and visual representations in the context of the visual-world paradigm.

In order to function as parsimonious theories of mental representations, embodied views of cognition should not only provide an account of the use and understanding of concrete concepts, but also for that of abstract ones. Accordingly, perceptual and motor representations are thought to underlie not only the comprehension of sentences about concrete objects and actions, but also the comprehension of abstract concepts such as *justice* and *love* (Barsalou & Wiemer-Hastings, 2005; Gallese & Lakoff, 2005; Prinz, 2005).

The studies discussed in this section have mostly focused on the comprehension of words or sentences. However, their relevance to the study of discourse – narrative, expository, and procedural – should be clear. It stands to reason that perceptual or motor representations activated during sentence or word comprehension play a role in discourse comprehension. Earlier, in Section 4, we stated that theories of discourse comprehension need to come to terms with two important facts. The first fact is that language comprehension involves incomplete representations. This is a function of limited attentional and memory resources, although these can be expanded in skilled comprehension (Ericsson & Kintsch, 1995). As such, it is unrealistic to assume that comprehenders build elaborate propositional networks by promiscuously activating large numbers of inferences (Graesser et al., 1994; McKoon & Ratcliff, 1992). At the same time, the sales figures for the *Harry Potter* saga, *The Lord of the Rings* trilogy, and *The Da Vinci Code* attest to the phenomenological experience of narrative comprehension as a highly engrossing and immersive activity (see Nell, 1988, for empirical evidence). It is difficult to see how this can be explained by sparse networks of abstract, amodal, and arbitrary propositions.



A view that assumes perceptual and motor representations underlie cognition provides a way to address the dilemma. Consider the following sentences (from Sanford & Garrod, 1998):

He put the wallpaper on the table. Then he put his mug of coffee on the wallpaper.

These sentences are quite easy to understand. A (somewhat simplified) propositional representation of these sentences would be [[PUT[HE, WALLPAPER]][ON[TABLE, WALLPAPER]]] and [[PUT[HE, MUG]][ON[MUG, WALLPAPER]]]. But now consider the following sentence pair (also from Sanford & Garrod, 1998):

He put the wallpaper on the wall. Then he put his mug of coffee on the wallpaper.

Most people notice right away that the second sentence does not make sense, or at least would result in a broken mug and coffee spilled all over the floor. However, it is not clear how they could come to this conclusion based on a propositional representation like: [[PUT[HE, WALLPAPER]] [ON[WALL, WALLPAPER]]] and [[PUT[HE, MUG]][ON[MUG, WALLPAPER]]]. A large number of additional propositions would have to be activated, such as that (1) walls are typically vertical, (2) wallpaper can be made to adhere to walls, (3) when wallpaper is attached to a wall, it is also vertical, (4) wallpaper when attached to a wall does not support objects, (5) unless some way to attach the object, e.g., glue is used, or (6) the object is a small insect, arachnid, reptile, or amphibian. As our discussion of inferences shows, most researchers assume that elaborative inferences are made only under a narrow set of circumstances; the complex sets of inferences necessary for the above explanation, then, seem even less tenable.

How, then, are we to account for comprehender's almost instantaneous balking at the validity of the second set of sentences in Garrod and Sanford's example? The problem seems less daunting if we abandon the idea of abstract, amodal, and arbitrary representations, and instead view comprehension as the language-guided mental simulation of the described situation. The key idea is that comprehension engages perceptual and motor systems by activating previous experiences, or *experiential traces* (Zwaan, 1999a, 2004), stored in these areas. For example, our reading or hearing of "wall" may activate a visual representation, which necessarily involves its verticality; the phrase "put the wallpaper" activates the motor program that we would use to hold and move wallpaper (e.g., using both hands); the prepositional phrase "on the wall" would activate a motor representation in which the arms are extended above the head and perhaps a visual representation of wallpaper occluding our view of the wall; the phrase "put his mug" would activate a motor representation of holding a mug (presumably by its handle) and moving it; the phrase "on the wallpaper," which would activate a visual representation of a vertical surface, would quickly lead to the conclusion that the action cannot be carried out because wallpaper cannot support the mug. Thus, this might be a way in which we can describe how linguistic input facilitates the construction of rich mental representations without assuming too much in the way of processing and knowledge activation. Mental simulations involving experiential traces can allow

humans to use language in a vicarious manner to gain “lifelike” experiences without requiring too much in the way of inferential processing activity. Current work continues to address the nature of these embodied representations by contrasting them with traditional accounts that invoke propositional explanations. The accumulating evidence continues to favor an embodied view.

## 7. CONCLUSION

Discourse comprehension is an essential and complex human endeavor involving processes and mechanisms associated with general cognition (e.g., memory and attention). We opened this chapter with a set of riddles in order to demonstrate the role of some of those processes. We now close this chapter with a final, fifth riddle. Again, recall that adequate comprehension is a function of incorporating background knowledge and current text in the service of problem solving. Hopefully that serves as enough of a hint to suggest how the reader might come up with an answer for this final question.

- In a race between an elephant and a giraffe, who do you think might win?

Well, if *our* giraffe is still in that fridge ....

## ACKNOWLEDGMENTS

We are grateful to Sid Horton, Panayiota Kendeou, and Rich Yaxley for their helpful comments on earlier versions of this chapter. Rolf Zwaan was supported by grant MH-63972 from the National Institutes of Health and David Rapp by grant R305G040021 from the Institute of Education Sciences, by the Marcia Edwards Fund, and by a Grant-in-Aid of Research from the Office of the Dean of the Graduate School at the University of Minnesota.

## REFERENCES

Albrecht, J. E., & Myers, J. L. (1995). The role of context in the accessibility of goals during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1459–1468.

Albrecht, J. E., & Myers, J. L. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes*, *26*, 87–107.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1061–1069.

Allbritton, D. W. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes*, *38*, 309–322.

- Allbritton, D. W., & Gerrig, R. J. (1991). Participatory responses in text understanding. *Journal of Memory and Language*, 30, 603–626.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. New York: Wiley.
- Anderson, J. R., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 17, 1–12.
- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27, 195–225.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In: D. Pecher, & R. A. Zwaan (Eds), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New York: Cambridge University Press.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21, 477–487.
- Bestgen, Y., & Vonk, W. (2000). Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language*, 42, 74–87.
- Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior*, 18, 187–198.
- Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, 11, 64–76.
- Borghi, A. M., Glenberg, A. M., & Kaschak, M. (2004). Putting words in perspective. *Memory & Cognition*, 32, 863–873.
- Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75, 1–28.
- Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, 247, 44–48.
- Bower, G. H., & Rinck, M. (2001). Selecting one among many referents in spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 81–98.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717–726.
- Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In: W. G. Chase (Ed.), *Visual information processing* (pp. 383–438). New York: Academic Press.

- Britton, B.K., & Black, J. B. (1985). *Understanding expository text*. Hillsdale, NJ: LEA.
- Britton, B. K., Glynn, S., Meyer, B. J. F., & Penland, M. (1982). Use of cognitive capacity in reading text: Effects of variations in surface features of text with underlying meaning held constant. *Journal of Educational Psychology, 73*, 51–61.
- Brooks, R. A. (1992). Artificial life and real robots. In: F. J. Varela, & P. Bourguine (Eds), *Toward a practice of autonomous systems: Proceedings of the first European conference on artificial life* (pp. 3–10). Cambridge, MA: MIT Press/Bradford Books.
- Calvo, M.G. (2000). The time course of predictive inferences depends on contextual constraints. *Language and Cognitive Processes, 15*, 293–319.
- Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes, 22*, 57–78.
- Caron, J., Micko, H. C., & Thuring, M. (1988). Conjunctions and the recall of composite sentences. *Journal of Memory and Language, 27*, 309–323.
- Carreiras, M., Carriedo, N., Alonso, M.A., & Fernandez, A. (1997). The role of verbal tense and verbal aspect in the foregrounding of information in reading. *Memory & Cognition, 23*, 438–446
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: C. Li (Ed.), *Subject and topic* (pp. 25–56). New York: Academic Press.
- Chambers, C. G., Magnuson, J. S., & Tanenhaus, M. K. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 687–696.
- Chen, S., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavior predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*, 215–224.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. Orlando, FL: Harcourt Brace Jovanovich.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In: R. Freedle (Ed.), *Discourse production and comprehension* (pp.1–40). Hillsdale, NJ: LEA.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240–248.
- Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language, 44*, 220–234.
- de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 373–385.
- de Vega, M., Leon, I., & Diaz, J. M. (1996). The representation of changing emotions in reading comprehension. *Cognition and Emotion, 10*, 303–321.

de Vega, M., Robertston, D. A., Glenberg, A. M., Kaschak, M. P., & Rinck, M. (2004). On doing two things at once: Temporal constraints on actions in language comprehension. *Memory & Cognition*, 32, 1033–1043.

Deaton, J. A., & Gernsbacher, M. A. (in press). Causal conjunctions and implicit causality cue mapping in sentence comprehension. *Journal of Memory and Language*.

Diakidoy, I. N., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: A comparison of the effectiveness of two instructional approaches. *Learning and Instruction*, 11(1), 1–20.

Diakidoy, I. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology*, 28, 335–356.

DiSessa, A. A. (2002). Why conceptual ecology is a good idea. In: M. Limon, & L. Mason (Eds), *Reconsidering conceptual change: Issues in theory and practice* (pp. 29–60). Dordrecht: Kluwer Academic Press.

Duchan, J. F., Bruder, G. A., & Hewitt, L. E. (1995). *Deixis in narrative: A cognitive science perspective*. Hillsdale, NJ: LEA.

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Coté, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language*, 29, 566–582.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, 22, 225–246.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, 27, 235–244.

Foss, C. L., & Bower, G. H. (1986). Understanding actions in relation to goals. In: N. E. Sharkey (Ed.), *Advances in cognitive science* (pp. 95–124). Chichester, England: Ellis Harwood Ltd.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory–motor system in reason and language. *Cognitive Neuropsychology*, 22, 455–479.

Genette, G. (1983). *Narrative discourse: An essay in method*. Ithaca, NY: Cornell University Press.

Gennari, S. P. (2004). Temporal references and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 877–890.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99–156.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: LEA.

- Gernsbacher, M. A., & Givón, T. (Eds), (1995). *Coherence in spontaneous text*. Philadelphia, PA: John Benjamins.
- Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27, 699–717.
- Gernsbacher, M. A., & Robertson, R. R. W. (1992). Knowledge activation versus sentence mapping when representing fictional characters' emotional states. *Language and Cognitive Processes*, 7, 353–371.
- Gernsbacher, M. A., & Shroyer, S. (1989). The cataphoric use of the indefinite *this* in spoken narratives. *Memory & Cognition*, 17, 536–540.
- Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. W. (1992). Do readers mentally represent characters' emotional states? *Cognition and Emotion*, 6, 89–111.
- Gernsbacher, M. A., Hallada, B. M., & Robertson, R. R. W. (1998). How automatically do readers infer fictional characters' emotional states? *Scientific Studies of Reading*, 2, 271–300.
- Gerrig, R. J. (1989). Suspense in the absence of uncertainty. *Journal of Memory and Language*, 28, 633–648.
- Gerrig, R. J. (1993). *Experiencing narrative worlds*. New Haven, CT: Yale University Press.
- Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39, 225–242.
- Gerrig, R. J., & Rapp, D. N. (2004). Psychological processes underlying literary impact. *Poetics Today*, 25, 265–281.
- Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: LEA.
- Givón, T. (1986). Prototypes: Between Plato and Wittgenstein. In: C. Craig (Ed.), *Noun classes and categorization* (pp. 77–102). Amsterdam: John Benjamins.
- Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, 30, 5–55.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9, 558–565.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26, 69–83.
- Goldman, S. R., & Bisanz, G. L. (2002). Toward a functional understanding of scientific genres: Implications for understanding and learning processes. In: J. Otero, J. A. Leon, & A. C. Graesser (Eds), *The psychology of science text comprehension* (pp. 19–50). Mahwah, NJ: LEA.

Goldman, S. R., & Varnhagen, C. K. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language*, 25, 401–418.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311–348.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395.

Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 503–515.

Grice, H. P. (1975). Logic and conversation. In: P. Cole, & J. L. Morgan (Eds), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

Guéraud, S., & O'Brien, E. J. (Ed.). (2005). Components of comprehension: A convergence between memory-based processes and explanation-based processes [special issue]. *Discourse Processes*, 39, 121–316.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly*, 28, 117–159.

Gygax, P., Garnham, A., & Oakhill, J. (2004). Inferring characters' emotional states: Can readers infer specific emotions? *Language and Cognitive Processes*, 19, 613–638.

Gygax, P., Oakhill, J., & Garnham, A. (2003). The representation of characters' emotional responses: Do readers infer specific emotions? *Cognition and Emotion*, 17, 413–428.

Hakala, C. (1999). Accessibility of spatial information in a situation model. *Discourse Processes*, 27, 261–280.

Halliday, M. A. K. (1967). *Intonation and grammar in English*. The Hague: Mouton.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longmans.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512–521.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124, 62–82.

Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, 29, 223–267.

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, 10, 104–110.

Huitema, J. S., Dopkins, S., Klin, C., & Myers, J. (1993). Connecting goals and actions during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1053–1060.

Jahn, G. (2004). Three turtles in danger: Spontaneous construction of causally relevant spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 969–987.

Jakobson, R. (1971). Quest for the essence of language. In: R. Jakobson (Ed.), *Selected writings*, Vol. 2: *Word and language* (pp. 345–359). The Hague, Netherlands: Mouton.

Järvikivi, J., van Gompel, P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16, 260–264.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kaakinen, J., Hyönä, J., & Keenan, J. M. (2002). Individual differences in perspective effects on on-line text processing. *Discourse Processes*, 33, 159–173.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In: D. Kahneman, P. Slovic, & A. Tversky (Eds), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kaschak, M. P., Madden, C. J., Theriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, 92, B79–B89.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language*, 32, 446–463.

Keenan, J., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study in natural conversation. *Journal of Verbal Learning and Verbal Behavior*, 16, 549–560.

Kelter, S., Kaup, B., & Claus, B. (2004). Representing a described sequence of events: A dynamic view of narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 451–464.

Kendeou, P., Rapp, D. N., & van den Broek, P. (2004). The influence of readers' prior knowledge on text comprehension and learning from text. In: R. Nata (Ed.), *Progress in Education* (pp. 189–209). New York, NY: Nova Science Publishers, Inc.

Kendeou, P., & van den Broek, P. (2005). The role of reader's misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, 97, 235–245.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182.



Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, 16, 193–202.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7, 257–266.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, 133–159.

Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language*, 28, 56–77.

Klin, C. M., Guzmán, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, 40, 593–604.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.

Levine, W. H., & Klin, C. M. (2001). Tracking of spatial information in narratives. *Memory & Cognition*, 29, 327–335.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, 34, 259–280.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94, 778–784.

Linderholm, T., Gernsbacher, M. A., van den Broek, P., Neninde, L., Robertson, R., & Sundermeier, B. (2004). Suppression of story character goals during reading. *Discourse Processes*, 37, 67–78.

Lorch Jr., R. F., & Lorch, E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, 21, 261–278.

Lutz, M. F., & Radvansky, G. A. (1997). The fate of completed goal information in narrative comprehension. *Journal of Memory and Language*, 36, 293–310.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, 31, 663–672.

Magliano, J. P., & Radvansky, G. A. (2001). Goal coordination in narrative comprehension. *Psychonomic Bulletin & Review*, 8, 372–376.

- Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, 29, 83–112.
- Mandler, J., & Johnson, N. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111–151.
- Mayer, R. E. (1984). Aids to prose comprehension. *Educational Psychologist*, 19, 30–42.
- McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, 25, 645–656.
- McDonald, J. L., & Shaibe, D. M. (2002). The accessibility of characters in single sentences: Proper names, common nouns, and first mention. *Psychonomic Bulletin & Review*, 9, 356–361.
- McGlone, M. S., & Harding, J. L. (1998). Back (or forward?) to the future: The role of perspective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1211–1233.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 82–91.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440–466.
- McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247–287.
- Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, 21, 121–143.
- Millis, K. K., Graesser, A. C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology*, 33, 317–339.
- Millis, K. K., & Just, M. A. (1994). Connectives increase clausal activation. *Journal of Memory and Language*, 33, 128–147.
- Minsky, M. A. (1975). A framework for representing knowledge. In: P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.
- Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, 28, 292–312.
- Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, 26, 165–187.
- Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, 22, 85–94.
- Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Predictive inferences in narrative text. *Journal of Memory and Language*, 32, 464–473.

- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131–157.
- Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26, 453–465.
- Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, 91, 488–496.
- Nell, V. (1988). *Lost in a book: The psychology of reading for pleasure*. New Haven: Yale University Press.
- Newman, L. S., & Uleman, J. S. (1990). Assimilation and contrast effects in spontaneous trait inferences. *Personality and Social Psychology Bulletin*, 16, 224–240.
- O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 777–784.
- O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1200–1210.
- Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, 3, 229–235.
- Peracchi, K. A., & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory & Cognition*, 32, 1044–1052.
- Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.
- Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–226.
- Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66, 211–227.
- Prentice, D. A., Gerrig, R. J., & Bailis, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review*, 4, 416–420.
- Prinz, J. J. (2005). Passionate thoughts: the emotional embodiment of moral concepts. In: D. Pecher, & R. A. Zwaan (Eds), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 93–114). New York: Cambridge University Press.
- Propp, V. (1968). *Morphology of the folktale*. Austin, TX: University of Texas Press.
- Radvansky, G. A., & Curiel, J. M. (1998). Narrative comprehension and aging: The fate of completed goal information. *Psychology and Aging*, 13, 69–79.

- Radvansky, G. A., Zwaan, R., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1224–1237.
- Rapp, D. N., & Gerrig, R. J. (2006). Predilections for narrative outcomes: The impact of story contexts and reader preferences. *Journal of Memory and Language*, *54*, 54–67.
- Rapp, D. N., & Gerrig, R. J. (2002). Readers' reality-driven and plot-driven analyses in narrative comprehension. *Memory & Cognition*, *30*, 779–788.
- Rapp, D. N., Gerrig, R. J., & Prentice, D. A. (2001). Readers' trait-based models of characters in narrative comprehension. *Journal of Memory and Language*, *45*, 737–750.
- Rapp, D. N., & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 988–1001.
- Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers. *Journal of Educational Psychology*, *80*, 16–20.
- Rich, S. S., & Taylor, H. A. (2000). Not all narrative shifts function equally. *Memory & Cognition*, *28*, 1257–1266.
- Richards, E., & Singer, M. (2001). Representation of complex goal structures in narrative comprehension. *Discourse Processes*, *31*, 111–135.
- Rinck, M., & Bower, G. H. (1995). Anaphora resolution and the focus of attention in situation models. *Journal of Memory and Language*, *34*, 110–131.
- Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition*, *28*, 1310–1320.
- Rinck, M., Bower, G. H., & Wolf, K. (1998). Distance effects in surface structures and situation models. *Scientific Studies of Reading*, *2*, 221–246.
- Rinck, M., Hähnel, A., & Becker, G. (2001). Using temporal information to construct, update, and retrieve situation models of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 67–80.
- Rinck, M., Hähnel, A., Bower, G. H., & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 622–637.
- Rinck, M., & Weber, U. (2004). Who when where: An experimental test of the event-indexing model. *Memory and Cognition*, *31*, 1284–1292.
- Rinck, M., Williams, P. Bower, G. H., & Becker, E. S. (1996). Spatial situation models and narrative understanding: Some generalizations and extensions. *Discourse Processes*, *21*, 23–55.

- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188–194.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Roy, D., & Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, 26, 113–146.
- Rumelhart, D. E., McClelland, J. L., & The PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Volumes 1 and 2*. Cambridge, MA: MIT Press.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437–442.
- Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, 2, 95–100.
- Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29, 37–60.
- Sanford, A. J., Clegg, M., & Majid, A. (1998). The influence of types of character on processing background information in narrative discourse. *Memory & Cognition*, 26, 1323–1329.
- Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language: Explorations of comprehension beyond the sentence*. Chichester: Wiley.
- Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26, 159–190.
- Sanford, A. J., Moar, K., & Garrod, S. C. (1988). Proper names and the control of focus. *Language and Speech*, 31, 43–56.
- Schank, R., & Abelson, R. (1977). *Scripts, goals, and understanding*. Hillsdale, NJ: LEA.
- Schank, R. C. (1982). *Dynamic memory*. Cambridge: Cambridge University Press.
- Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, 25, 279–294.
- Schneider, W., & Körkel, J. (1989). The knowledge base and text recall: Evidence from a short-term longitudinal study. *Contemporary Educational Psychology*, 14, 382–393.
- Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, 81, 306–312.
- Singer, M. (1993). Global inferences of text situations. *Discourse Processes*, 16, 161–168.

- Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology*, *30*, 1–38.
- Singer, M., & Remillard, G. (2004). Retrieving text inferences: Controlled and automatic inferences. *Memory & Cognition*, *32*, 1223–1237.
- Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*, 125–140.
- Spivey, M. J., Richardson, D. C., & Gonzalez-Marquez, M. (2005). On the perceptual-motor and image-schematic infrastructure of language. In: D. Pecher, & R. A. Zwaan (Eds), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 246–281). New York: Cambridge University Press.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153–156.
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In: R. O. Freedle (Ed.), *New directions in discourse processing: Vol. 2. Advances in discourse processes* (pp. 53–120). Norwood, NJ: Ablex.
- Suh, S., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, *32*, 279–300.
- Sundermeier, B. A., van den Broek, P., & Zwaan, R. A. (2005). Availability of objects and locations during text comprehension. *Memory & Cognition*, *33*, 462–470.
- Taylor, H. A., Naylor, S. J., & Chechile, N. A. (1999). Goal-directed influences on the representation of spatial perspective. *Memory & Cognition*, *27*, 309–319.
- Taylor, H. A., & Tversky, B. (1997). Indexing events in memory: Evidence of index dominance. *Memory*, *5*, 509–542.
- Terken, J. M. B., & Nootboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, *2*, 145–163.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.
- Therriault, D. J., Rinck, M., & Zwaan, R. A. (2006). Assessing the influence of dimensional focus during situation model construction. *Memory & Cognition*, *34*, 78–89.
- Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, *24*, 595–611.
- Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.

- Trabasso, T., van den Broek, P. W., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, *12*, 1–25.
- Trabasso, T., & Wiley, J. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes*, *39*, 129–164.
- Traxler, M. J., Bybee, M., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology*, *50(A)*, 481–497.
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*, 185–203.
- Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*, 377–394.
- van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, *27*, 1–22.
- van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In: D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds), *Comprehension processes in reading* (pp. 423–445). Hillsdale, NJ: LEA.
- van den Broek, P. (1994). Comprehension and memory for narrative texts: Inferences and coherence. In: M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). San Diego, CA: Academic.
- van den Broek, P., Lorch Jr., R. E., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*, 1081–1087.
- van den Broek, P., Lynch, J. S., Naslund, J., Ievers-Landis, C. E., & Verduin, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology*, *94*, 707–718.
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, *39*, 299–316.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press, Inc.
- Vosniadou, S. (2002). On the nature of naïve physics. In: M. Limon, & L. Mason (Eds), *Reconsidering conceptual change: Issues in theory and practice* (pp. 61–76). Dordrecht: Kluwer Academic Press.
- Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, *18*, 123–124.
- Weingartner, K. M., Guzmán, A. E., Levine, W. H., & Klin, C. M. (2003). When throwing a vase has multiple consequences: Minimal encoding of predictive inferences. *Discourse Processes*, *36*, 131–146.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology*, *78*, 1024–1037.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252.

Wolfe, M. B. W. (2005). Memory for narrative and expository text: Independent influences of semantic associations and text organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 359–364.

Zwaan, R. A. (1991). Some parameters of literary and news comprehension: Reading rate and surface-structure representation in literary and news comprehension. *Poetics*, *20*, 139–156.

Zwaan, R. A. (1993). *Aspects of literary comprehension: A cognitive approach*. Amsterdam/Philadelphia: John Benjamins.

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 920–933.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1196–1207.

Zwaan, R. A. (1999a). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88.

Zwaan, R. A. (1999b). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, *8*, 15–18.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In: B. H. Ross (Ed.), *The Psychology of learning and motivation*, (Vol. 44, pp. 35–62). New York: Academic Press.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In: D. Pecher, & R. A. Zwaan (Eds), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 224–245). New York: Cambridge University Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.

Zwaan, R. A., & Van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes*, *16*, 125–143.

Zwaan, R. A., & van Oostendorp, H. (1994). Spatial information and naturalistic story comprehension. In: H. van Oostendorp, & R. A. Zwaan (Eds), *Naturalistic text comprehension* (pp. 97–114). Norwood, NJ: Ablex.

Zwaan, R. A., & Yaxley, R. H. (2004). Lateralization of object-shape information in semantic processing. *Cognition*, *94*, B35–B43.



Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995a). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, 292–297.

Zwaan, R. A., Madden, C. J., & Whitten, S. N. (2000). The presence of an event in the narrated situation affects its availability in the comprehender. *Memory & Cognition*, 28, 1022–1028.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, 28, 611–619.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995b). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386–397.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, 2, 199–220.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science*, 13, 168–171.

## Chapter 19

### Neuroimaging Contributions to the Understanding of Discourse Processes

*Robert A. Mason and Marcel Adam Just*

Neuroimaging research is providing new types of information and insight about the cortical activity underlying discourse processing. Knowing the intensity and location of the brain activity during discourse comprehension adds significantly to the information provided by behavioral measures alone. The combination of neuroimaging data and behaviorally based discourse theories indicate that discourse processing is underpinned by a system of several distinguishable cortical networks that are activated for discourse processing, above and beyond the activation evoked by comprehension at the word and sentence level. Whereas the multiplicities of the processes in discourse comprehension are sometimes seen as a drawback to behavioral experiments, it is something of a benefit in neuroimaging research. Controlled neuroimaging experiments, with their multidimensional measures, can help determine when each of these components contributes to discourse processing. By making some assumptions about the cortical regions/network that underlie this processing, we can begin to determine when an area becomes activated and to what degree it is activated as a function of the discourse properties.

In this chapter, we describe some key neuroimaging studies of discourse processing, and observe some systematic patterns of results that apply across the described studies. To foreshadow these patterns, we list here five specialized networks we believe to be involved in discourse processing.

#### *Parallel Networks of Discourse:*

1. A coarse semantic processing network (right middle and superior temporal)
2. A coherence monitoring network (bilateral dorsolateral prefrontal)
3. A text integration network (left inferior frontal-left anterior temporal)
4. A network for interpreting a protagonist's or agent's perspective (bilateral medial frontal/posterior right temporal/parietal)

A spatial imagery network (left dominant, bilateral intraparietal sulcus)

This list should be treated with caution and with excitement. It is exciting that some of the components of discourse processing revealed by neuroimaging research, like protagonist perspective monitoring (here we use protagonist as a shorthand for any agent in the story capable of intentional action), are relatively new to the discourse processing theory (although comprehending the motivations of characters in a story is probably an ancient skill). At the same time, there is uncertainty about the reality of these networks and about their anatomical location. Moreover, these networks must function in interaction with somewhat lower level comprehension processes that operate at the lexical and sentence level (for a review see: Bookheimer, 2002; Gernsbacher & Kaschak, 2003). Despite these cautions, it seems useful to consider the recent research with some framework in mind, and we suggest that these five networks/discourse functions provide an initial attempt at such a framework. As we proceed through the chapter we will elaborate on the characteristics of these networks as they are illuminated by the various studies. At the conclusion of the chapter we will expand on a theoretical framework based on these networks.

Although the focus of this paper is decidedly not on “where” in the brain discourse processing occurs, a brief overview of which areas of the brain play a role in discourse comprehension is useful. In almost every discourse processing task, the traditional left hemisphere language network activates in a contrast with a fixation baseline condition. This traditional left hemisphere language network includes the left hemisphere inferior frontal gyrus, the superior and middle temporal gyrus, portions of the inferior temporal gyrus, and the angular gyrus. In addition to this basic language processing network, we expected additional discourse networks to activate during discourse processing. An overview of the Parallel Networks of Discourse and a rough sketch of the cortical regions in which they are localized are shown in Figure 1. This schematic representation is shown via surface projection on a rendered brain. It is not intended to depict an exhaustive account of discourse processing networks but instead highlight some key areas. It is likely that the networks are differentially engaged in the research presented here and the localization of the peak activation for a specific task could be expected to vary somewhat within an anatomical region.

Discourse theories become critical in developing this understanding of the cortical discourse processing network. In addition, neuroimaging research has led to the development of several new discourse theories such as the coarse coding theory of right hemisphere processing (Beeman, 1998), the dynamic recruitment of networks in response to text constraints (Ferstl, Rinck & von Cramon, 2005; Mason & Just, 2004; Xu, Kemeny, Park, Frattali, & Braun 2005), a Theory of Mind system responsible for awareness of different perspectives (Gallagher & Frith, 2003), and the spillover of processing to other differential specialized networks in response to capacity utilization (Just, Carpenter, Keller, Eddy, & Thulborn 1996). The cortical activation and these new theories are based on, or at least consistent with, traditional discourse theories such as Kintsch’s Construction-Integration framework (1988), Gernsbacher’s Structure Building Framework (1990), Myers and O’Brien’s Resonance model (1998), Giora’s Graded Salience Hypothesis (1997), and van den Broek’s Landscape architecture (1996) as well as others.

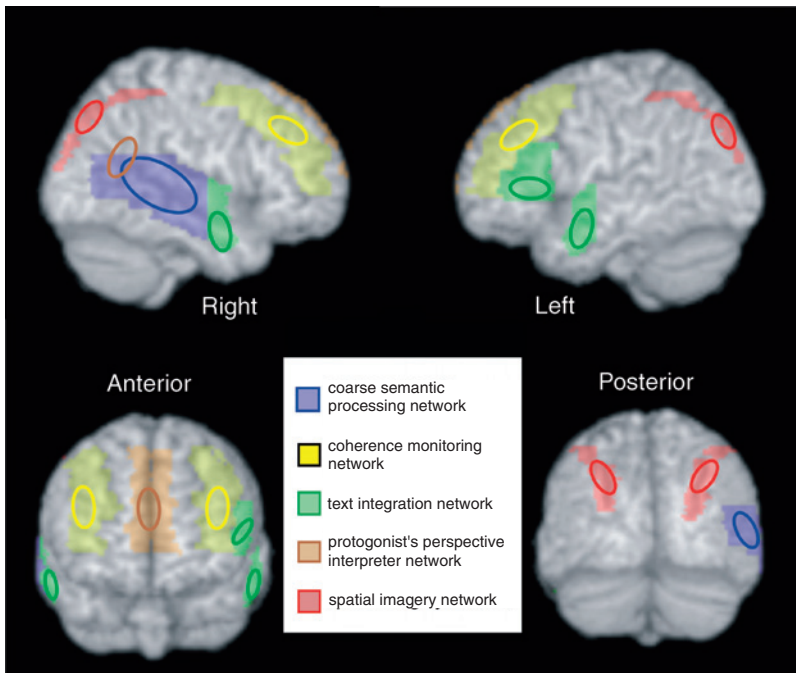


Figure 1. A Schematic representation of the Parallel Networks of Discourse. Shaded regions represent surface rendered anatomical regions as described in the text. A rough localization within anatomical regions are represented by colored ellipses. For a full color image see: [http://www.ccbi.cmu.edu/reprints/Mason\\_Figure1.tif](http://www.ccbi.cmu.edu/reprints/Mason_Figure1.tif)

## 1. THE BEGINNINGS OF BRAIN IMAGING IN DISCOURSE COMPREHENSION

Dating back to Broca's and Wernicke's findings on brain-damaged patients with specific language deficits in the late 1800s, psychologists have had some idea of the brain's functioning as a language processing mechanism. There is however, no similarly well-known case of a patient with a deficit in discourse processing abilities. One reason for this lack may be the difficulty in defining what is meant by a discourse processing deficit. Another reason for the lack of reports on patients with a discourse-processing deficit is that many patients with such deficits are either still able to function well in everyday life or they also have severe deficits at lower levels of language processing. Both circumstances would make the deficit less apparent to clinicians. In recent years, however, more sensitive neuropsychological investigations have detected impaired discourse functioning at several levels and have found such impairments to be correlated with right hemisphere damage. Unlike the situation with Broca's and Wernicke's patients, consistent focal lesions have not been found in these discourse aphasics. Various patients have lesions in right hemisphere homologues of Broca's area (e.g., the inferior frontal gyrus), Wernicke's area (e.g., posterior-superior and middle temporal gyrus) as well as the right hemisphere angular gyrus, dorsolateral prefrontal cortex and the medial frontal gyrus. Additionally,

several other populations, such as individuals with autism (Dennis et al., 2001) and Alzheimer's patients (Papagno, 2001), with non-focal lesions or no lesions at all have difficulty with discourse processing. Thus, unlike the case for lower-level language processing research, neuropsychological studies of discourse processing in brain-damaged and other special populations did not provide much information about brain function other than a generalized notion that the right hemisphere was somehow involved in discourse processing.

It was not until the late 1980s that researchers began using brain imaging as a technique to investigate cognitive process. At that time neuroimaging was primarily based on positron emission tomography (PET). One of the earliest neuroimaging investigations of a "discourse variable" was a PET study of metaphor comprehension conducted by Bottini et al. (1994). Subjects listened to either literal or metaphorical sentences, including such examples as:

plausible:                            *"The investors were squirrels collecting nuts"*

implausible:                           *"The investors were trams."*

The sentences containing metaphors elicited more activation in the right hemisphere, particularly the right inferior frontal gyrus and right posterior temporal cortex. The increased involvement of the right hemisphere may have been the result of an inference process that combined world knowledge with the contents of the sentence to resolve the ambiguity. This early finding of right hemisphere involvement in metaphor comprehension, combined with neuropsychological and visual field presentation data, contributed to the hypothesis that the right hemisphere is critically involved in metaphor comprehension and in other facets of discourse processing. This view is still widely held today although much refined. The right hemisphere coarse-coding hypothesis for example (described more fully in the lateralization section) developed by Beeman and colleagues (Beeman, 1993; Beeman, 1998; Beeman et al., 1994; Brownell, Polter, Bihrlé, & Gardner, 1986) is a broader theory which encompasses the right hemisphere discourse view.

Nichelli, et al. (1995) was among the first to visually present passages consisting of multiple sentences (Aesop's Fables) passages in a neuroimaging study. (Another early auditory study was reported by Mazoyer et al., 1993.) Participants in Nichelli's PET study were asked to monitor either semantic details, syntactic details or the moral of the fable. Nichelli et al. concluded that the extra activation observed in the right hemisphere during moral monitoring and not in the other conditions was due to the drawing of an inference about the passage. Although there may have been some alternative accounts offered for this finding, this early text comprehension study advanced the view that there was something special about the role of the right hemisphere in discourse processing. In the decade since these early discourse studies, considerable advancement has been made in both methods and theory.

There were several possible reasons for the scarcity of neuroimaging research at the discourse processing level, some of which remain as problems: (1) neuroimaging research

into language processing at any level is fairly new, (2) discourse processing is a broad field without a tradition for concern about neural mechanisms, (3) many of the key research questions that have been asked in the discourse processing area are not easily addressed with neuroimaging (due to the sluggish nature of the hemodynamic response and the low temporal sampling rate), (4) several of the frequently-used paradigms used to investigate discourse processing are not easily adaptable to the scanner environment (e.g., talk aloud protocols, naming), and (5) neuropsychological research on discourse processing deficits is limited. Recent advancements in technology and knowledge about cortical function have alleviated many of these constraining factors. The release of some of these constraints has opened many exciting new possibilities for the investigation of the neural substrate underlying discourse processing.

## **2. SPECIALIZED IMAGING PARADIGMS FOR DISCOURSE PROCESSING RESEARCH**

Perhaps the largest factor constraining discourse imaging research is the challenge of developing appropriate experimental paradigms within the constraints of brain scanning methodology. A brief consideration of imaging methodology will make this point more clear. Imaging of brain activity using PET requires uptake of a tracer substance into the blood stream and into the brain during task performance. The relative insensitivity to the tracer requires long sampling intervals, such that brain activity can be measured only over many tens of seconds. As a result, PET studies have to utilize a “blocked” or “epoch” design in which many trials or items of the same type (i.e., constituting the same experimental condition) are presented in a sequence or “block,” and brain activity is measured during the processing of the entire block of stimuli. The activity is then contrasted with other blocks (experimental conditions) in which a different set of processes occur. The activity in the experimental conditions is typically estimated with respect to a baseline task, usually a simpler task that is believed to be common to the two or more experimental conditions. The resulting image is then “subtracted” from the experimental blocks so that the non-baseline processing can be isolated. Even with some shortcomings of the subtraction paradigm (Newman, Twilg, & Carpenter, 2001) this methodology was very fruitful for the early days of imaging. At the very least it served to illuminate the extent of various networks engaged in many cognitive processes. But the nature of PET imaging led to some concessions in experimental design. Because PET is less sensitive and requires a radioactive tracer, most mainstream functional neuroimaging has turned to functional Magnetic Resonance Imaging (fMRI).

The underlying assumption of fMRI is that it is a measure of neuronal activity, which should increase in some area when a cognitive process makes use of substrate in that area. This increase in neuronal activity results in an increase in local blood flow and volume. The oxygen content is then elevated in cortical areas which are being used. This will result in an increase in the MR signal, which is affected by the ratio of deoxygenated hemoglobin to oxygenated hemoglobin (Cohen & Bookheimer, 1994).

A key advantage of the fMRI methodology is that a significantly smaller temporal window can be used in the measurement of brain activity than had been possible with PET. In some cases, the cognitive processing that we would like to measure is very short in duration and, particularly with language processing, it occurs on the order of tens of milliseconds. There is obvious difficulty when trying to measure a rapid cognitive process with a very slow measurement such as PET. In fact, PET does not really allow a consideration of the moment by moment changes in activation. It results in a single average image from a large temporal window. The fMRI can be used to acquire an image in many consecutive temporal windows. The exact temporal window (termed TR, for “time for repetition”) is related to the strength of the magnet, the amount of cortex being imaged, the criteria for signal to noise and, critically, the amount of time in which it takes the protons in the imaged substance to return to baseline after the introduction of a radio frequency pulse that causes them to tilt. Even though the temporal window in imaging varies across experiments, the range is roughly between 1 s (e.g., Just, Newman, Keller, McElency, & Carpenter, 2004; Mason, Just, Keller, & Carpenter, 2003) and 3 s (e.g., Martin & Weisberg, 2003; Robertson et al., 2000; St. George, Kurtas, Martinez, & Sereno, 1999).

Although early fMRI studies of discourse comprehension used a block design, they took advantage of the more rapid temporal window and the ability to compare conditions without subtracting out lower levels of cognitive processing. For example, Robertson et al. (2000) presented readers with blocks of sentences which either contained an indefinite article or a definite article. They found greater right hemisphere activation for the lists of sentences that used definite articles rather than indefinite articles. The definite article sentences were assumed to lead to more coherent discourse than the indefinite article sentences. They concluded that the additional right hemisphere activation reflected processes used to anaphorically relate the nouns in a text. Ferstl and von Cramon (2001) compared pairs of sentences that were coherent or incoherent as well as cohesive or incohesive. The coherence manipulation resulted in activation in the left frontal gyrus. The cohesion manipulation involved adding lexical connectives to the pairs of sentences to make them easier to understand as a single unit. Unlike Robertson et al. they did not find any additional right hemisphere activation. In both the Robertson et al. and Ferstl and von Cramon studies, the researcher utilized sentences or pairs of sentences in which a similar type of discourse level processing could be assumed to occur in all items of one type and not the items of another type.

Recently the development of slow-paced event-related fMRI has enabled the randomization of items within experiments (Buckner et al., 1996), and more importantly, has enabled the measurement of brain activity during the comprehension of individual sentences. In a slow-paced event-related fMRI, a blank interval appears (about 7–14 s, hence the term slow-paced) between the items, sufficient to allow the hemodynamic response to return to a baseline level (Mason et al., 2003; Mason & Just, 2004) so that the activations associated with individual sentences are separable. The development of the slow-paced and fast-paced event-related experimental designs for fMRI allowed for imaging researchers to both randomize presentation of items as well as isolate specific cognitive processing to single sentences as opposed to examining gross levels of processing differences.

The Mason and Just study (2004) provides an example of how the event-related approach can be used to both randomize items as well as isolate specific types of cognitive processing. The study examined causal inferencing in the comprehension of two-sentence passages, drawn from previous stimulus materials (Keenan, Baillet, & Brown, 1984; Myers, Shinjo, & Duffy, 1987). The critical sentence in each passage was the second one, and the experimental manipulation was the degree of causal relatedness to the preceding sentence. The second sentence was followed by a fixation point so that the activation could return to baseline before the next passage. Using this technique, it was possible to identify the time interval of cortical activation that corresponded to the processing of the critical sentence. The processing was expected to differ among the experimental conditions, and this expectation was confirmed, as described later in the chapter.

Even the relatively rapid 1 s temporal sampling rate used by Mason et al. (2003) seems slow relative to the time measures and effect sizes that are typical of behavioral studies of discourse comprehension. For example, the reading time difference between reading a sentence with an indefinite article and the definite article may be only a matter of tens of milliseconds. Nevertheless, even small differences in processing time can produce measurable differences in the brain activity that has been aggregated over 1 or more s, as several studies (e.g., Robertson et al., 2000) have shown. Although averaging over 1 or more s can be viewed as a weakness of fMRI, there is a sense in which it conveys an advantage, because it is often difficult to isolate the cognitive process of interest to a specific temporal window during the reading of a passage. For example, McKoon and Ratcliff (1986, 1989, 1992) have shown that predictive inferences could be drawn either on the first sentence at which it was possible to draw the inference or in the subsequent sentence. Consider how Robertson et al. (2000), Ferstl and von Cramon (2001) and Mason and Just (2004) dealt with this issue. In all three cases, they used well-constrained materials of either a single sentence or sentence pairs to maximize the probability that the effect would occur in a specific temporal window.

In modern discourse research, the trend has been toward much longer and more naturalistic passages. This creates an added methodological burden. Every second of an imaging session is precious, because a participant will lie still for only so long. Although the background and introductory sections of these longer, naturalistic passages are critical, the expectation is that cognitive processing during this context-setting period is similar across conditions. The data acquired during this temporal window would either have to be discarded or treated as an additional factor in the design. Similarly, if the critical sentence does not include the expected cortical processing (perhaps the processing is delayed to a post-target sentence) then the likelihood of finding cortical evidence of a cognitive process is greatly reduced given that often there are as few as 10–20 items per condition in neuroimaging experiments.

Perhaps the largest advantage that imaging research has over behavioral research is that it provides a fairly direct measure of the processing activity in each of the neural networks underpinning discourse comprehension. Measuring cognitive workload in most purely behavioral studies is difficult, frequently necessitating the use of secondary tasks



as means of measuring processing load. The fMRI studies allow the collection of button pressing data and response times, but many other data collection methods are problematic. Vocalization responses such as naming times and talk-aloud protocols are more difficult to acquire. Although, noise-canceling technologies have facilitated the extraction of voice responses in the noisy scanning environment, the head motion induced by jaw movements can render data unusable. The normal scanning procedure also makes it difficult to collect discourse recall data from a scanning session. As with voice responses, these problems can be overcome with some effort, but the problems have not been systematically solved so far.

In summary, brain imaging methodology imposes some constraints on discourse processing experimental design. Passages must be designed carefully to maximize the chance of finding how a complex cortical network functions in discourse comprehension. Specific process must be temporally localized to a specific point in the text. Moreover, the same process must occur across the majority of passages for the majority of subjects. For example, if the generation of a bridging inference is being examined, the inference must be generated at a similar point within the text across stimuli and across readers to be measured. The limitation on number of passages and subjects constrains the ability to average over a large set of passages.

Even with the constraints on experimental design, imaging still has the benefit of examining how the network as a whole functions. This is true even in those cases in which a process, such as inferencing, might not occur at a specific temporal point. Thus, it is possible to see the cognitive workload required for generating an inference with imaging even if it cannot be determined behaviorally whether or when an inference was drawn. The fMRI can thus be sensitive to processes that have only small effects on behavioral measures, while at the same time capturing the qualitative variation and parallel nature of the processes underlying discourse comprehension to which response times are insensitive.

### **3. NEW PERSPECTIVES ON TEXT INTEGRATION**

Text attributes at the discourse level enter into combinations with other information to allow a reader to weave individual sentences into an integrated narrative structure. The resulting conceptual structure incorporates pragmatic information and connects the text with the reader's world knowledge. This discourse process extends beyond strictly linguistic information. For example, discourse comprehension requires that the reader generate inferences and extract meaning that is not explicitly encoded in the text. Readers must make inferences in order to integrate sentences in a coherent fashion, filling in what is absent from or ambiguous in the text. Several researchers have tried to describe the properties of the internal representation of discourse. One of the most influential was the situation model as developed by van Dijk and Kintsch (1983). The situation model arises from linguistic processing of the text itself (lexical access, syntactic processing, and construction of a propositional based micro-structure) and an

interaction with non-linguistic cognitive processes. The situation model is a result of this interaction created by connecting the text with knowledge derived from the reader's long-term memory, and involves additional demands upon attention (e.g., the ability to shift points of view and parse sequences of events), working memory (the ability to retain longer term, anaphoric references), and the contribution of visual imagery, empathy, and emotional knowledge.

Some early attempts at imaging investigation of discourse processing were designed to determine the neural underpinnings of the construction of a situation model. An fMRI study conducted by St. George and colleagues (1999) presented syntactically well-structured paragraphs which were uninterpretable by virtue of never specifying the referent of the text. The paragraphs were similar to those previously used by Bransford and Johnson (1972.) These paragraphs were either preceded or not preceded by a title that disclosed the referent. For example, these paragraphs contained sentences such as "Typically, success requires that you start with your left leg, and make sure that it is securely in place. Then swing your body high into the air." Without knowledge of the referent, it is difficult if not impossible to understand the passage. But with foreknowledge provided by a title ("Riding a horse") all of the sentences become interpretable.

The fMRI results revealed that the left hemisphere, as a whole, exhibited no effect of whether the paragraph was presented along with the title, while the right hemisphere revealed significantly greater involvement during the presentation of the untitled paragraphs. More specifically, the left middle and superior temporal sulci became more active during the processing of the titled paragraphs, and conversely, the right middle and superior temporal sulci became more active during the processing of the untitled paragraphs. These results support the idea that the right hemisphere is concerned with the mapping of information into a text representation, as discussed below. In addition, these results suggest that the processing roles of the two hemispheres are, in fact, distinguishable.

Tomitch, Just, and Newman (2004) attempted to investigate the differential processing of the left and right hemisphere during text integration using fMRI. They manipulated the serial position of the topic sentence in short, three-sentence paragraphs. The topic sentence contained a unifying super-ordinate theme, while the supporting sentences in the paragraph instantiated that thematic concept. The serial position manipulation consisted of varying the position of the topic sentence in the paragraph, placing it either in the first position – *topic first* or in the third position – *topic last*, as shown below.

'Topic first' condition

This is a totally guaranteed method to completely eliminate a flea infestation on your dog or around his doghouse. (Topic first)

First, late in the evening, chain your dog to his doghouse, build a small bonfire and let it burn overnight. (Support 2nd - #1)

They are insatiably attracted to heat, become enamored of the fire, leave your dog, jump into the flames, and die. (Support 2nd - #2)

Fleas will be eliminated from your dog or his doghouse with the use of a bonfire. True or false?

‘Topic last’ condition

First, late in the evening, chain your dog to his doghouse, build a small bonfire and let it burn overnight. (Support 1st - #1)

They are insatiably attracted to heat, become enamored of the fire, leave your dog, jump into the flames, and die. (Support 1st - #2)

This is a totally guaranteed method to completely eliminate a flea infestation on your dog or around his doghouse. (Topic last)

Fleas will be eliminated from your dog or his doghouse with the use of a bonfire. True or false?

This paradigm made it possible to measure the brain activation associated with the comprehension of each of the three sentences in each paragraph separately. The results revealed differential effects in the two hemispheres. The right temporal cortex revealed greater involvement during the processing of topic sentences, regardless of their location within the paragraph. In contrast, the left temporal cortex was sensitive to the location of the topic sentence.

Tomitch et al. cited Gernsbacher’s (1990) *structure building framework* (SBF) to account for the processing underlying the cortical activation. According to SBF, discourse comprehension builds cohesive mental representations using three general processes: laying the foundation, mapping incoming information to previous information, and initiating a new substructure if the incoming information is not adequately coherent with previous information. SBF states that the first step in building a mental representation of the text is to lay a foundation to which subsequent information presented in the text can be attached. Presumably this first stage must occur across all passages regardless of the order of the topic sentence; the consistent left temporal activation on the first sentence suggests that the left temporal region is involved in laying the foundation of the text representation.

A second prediction SBF makes is based on more involvement of the “shifting” process in paragraphs whose topic sentence is in the final position. When the topic sentence is in the final position, increased shifting is, therefore, expected to result in higher activation levels. This response was also observed in the left temporal region.

The right temporal region was sensitive to the presence of a topic sentence but not to its location. This suggests that the right temporal region (1) is sensitive to whether a sentence is a potential statement of the topic and 2) performs additional processing on the

potential topic sentence. The right hemisphere may then be responsible for using the information forwarded by the left to “fill in” or map information onto the text representation built by the left and to connect it with the participant’s world knowledge. This right temporal activation “mapping” activation was also found by Robertson et al. (2000) for both indefinite and definite articles; in contrast, the less coherent, indefinite article texts also resulted in additional right frontal activation than the definite article texts.

Several other cortical regions have been found to play a role in tasks that require structure building. Partiot, Grafman, Sadato, Fitman, and Wild (1996) investigated script processing and found bilateral precuneus/posterior cingulate regions to be activated, along with bilateral medial parietal cortex, during processing of event sequences. Maguire, Frith, and Morris (1999), using Bransford and Johnson (1972) passages with or without titles, also found these same areas to be active in linking textual information with subjects’ prior knowledge. These operations must be central to construction of a situation model, connecting the narrative text with knowledge about the real world.

Narrative-specific activations have also been found in the temporo-parieto-occipital junction, angular gyrus, and superior temporal sulcus. A prevailing view has been that the angular gyrus plays a key role in grapheme to phoneme translation. But this region is in fact multifunctional, and has been implicated in a variety of cognitive processes – attention, semantic association, problem solving, and mental imagery (Cabeza & Nyberg, 2000) that are likely to be engaged in the narrative context. For example, the angular gyrus is activated when subjects visualize a scene derived from a written text (Mellet et al., 2002) – precisely the sort of mental model representation required during narrative processing.

Recently, Ferstl, Rinck, and von Cramon (2005) examined passages in which the reader might encounter inconsistent emotional or temporal information in a passage. They suggested the contrast of consistent and inconsistent information that should be coded at the situation model level would allow them to examine cortical networks specialized for situational level text representation. They concluded that activation in the frontal cortex indicated a specialization for building and maintaining a situation model representation. The specific region within the frontomedial cortex varied as a function of the type of information and whether it was involved in noticing or resolving the inconsistency. First, the ventral portion of the medial frontal cortex was involved in detecting inconsistent emotional information. In contrast, the dorsomedial frontal cortex was active in the processing of emotional consistencies suggesting that a protagonist interpreter network was engaged during the attempted resolution of the inconsistent information. The chronological inconsistency activation was located much more anteriorly in the orbital portion of inferior frontal gyrus and the frontopolar region. Interestingly, the right anterior temporal lobe was more active during the processing of inconsistent texts regardless of the text type, suggesting that as text processing became more difficult, the specialized text integration network spilled over into the right hemisphere.

To summarize, while there is still a significant amount of research to be done to clarify the contributions of the left and right hemisphere in text integration, neuroimaging studies such as those outlined here are making good progress. Although it must be true

that the two hemispheres work together to accomplish such a complex function as comprehension, it does appear as though they are involved in different aspects of discourse processing. It seems as though the identification of the main idea and the building of the text representation is separable from the mapping of that information onto the discourse structure with both hemispheres working in an interactive manner in order to construct a coherent representation of the text. Furthermore, the type of the information within the text plays a role in which areas are responsible for building and maintaining a representation of the text. Propositional level information may be processed by left frontal and left temporal areas as long as resources are available. Situational model information results in an engagement of the medial frontal cortex. Specific areas within the cortex may be dependent upon the nature of the input, for example, emotional information related to protagonist interpreter should activate more dorsal frontal regions.

#### 4. NEW PERSPECTIVES ON INFERENCE PROCESSING

Often, the links between events in a story are not explicitly expressed and the reader must connect them by generating linking inferences and integrating them with the presented information. Almost every text requires a reader to draw on a rich store of shared knowledge about the world. There is a strong relation between inference generation and text integration. Inference generation is often necessary in building accurate text representations and conversely, an accurate text representation is often necessary for inference generation.

Inferences may be drawn to fill in missing information, resolve discrepancies or to predict yet unmentioned events or facts. There are several types of inferences, including coherence, predictive, elaborative, and causal. The classification of various types of inferences has been a continuing topic in discourse research and there have been several excellent attempts to resolve this classification problem (e.g. Singer, 1994; van den Broek, 1994).

A successful inference generally occurs as a result of generating a possible inference and then integrating that inference into the internal representation of the text. The Construction-Integration (CI) model of text comprehension (Kintsch, 1988) is consistent with this general description of inferencing. According to the CI model, there is a first process in which the many possible inferences are liberally generated (inference construction), followed by a second process of integrating only those inferences that have a high degree of connection with the reader's representation of the preceding text (integration). A successful integration of an inference results in a representation of the text that involves both the specific propositions contained in the text and those propositions that were generated by the reader to connect information in the text.

Patients with lesions to the right hemisphere generally have trouble drawing inferences in order to integrate sentences and maintain coherence (Beeman, 1993; Brownell et al., 1986). These patients are less likely to mistakenly false alarm to inferences in a text recognition task, presumably because they never generated the inferences in the first

place (Grafman, Salazar, Vance, Weingartner, & Amin, 1987). Such patients also make elaborative inferences more easily than bridging inferences (Tompkins & Mateer, 1985). Furthermore, Beeman and colleagues (1994) have shown that when probes are inference-related, they are primed in the left visual field-Right Hemisphere immediately, and are primed in both left visual field-Right Hemisphere and right visual field-Left Hemisphere at a later time. This finding suggests that the right hemisphere is particularly involved in inference processing.

There have been few neuroimaging studies of inference processing. One question that has been debated is whether logic-based inferencing relies on the same processes as text-based inferencing. In an fMRI study, Caplan and Dapretto (2001) directly addressed this issue by comparing the generation of logic-related inferences versus text-based inferences. There were two types of sentence pairs as shown below.

Text based:	<i>“Do you believe in angels?”</i>
	<i>“Yes, I have my own special angel”</i>
Logic-related:	<i>“Do you like having fun?”</i>
	<i>“Yes, because it makes me happy”</i>

While the logic condition produced greater activation within the left language areas, the text-based condition revealed more activation in the right hemisphere. This study is significant in that it suggests that at the neural level, and consequently at the process level, there are significant differences between logic-based and text-based inferencing. Also, the results converge with the neuropsychological findings showing that the right hemisphere is intimately involved in text-related inference processing.

Mason and Just (2004) reported an fMRI experiment that was designed to examine the cortical areas that are involved in making causal inferences during reading. Participants read sentence pairs that varied in terms of their causal relatedness. For example, an “outcome” sentence (e.g., *The next day his body was covered with bruises.*) was preceded by one of three sentences (equated for overall length and number of propositions) that described an antecedent condition:

Highly Related:	<i>Joey’s big brother punched him again and again.</i>
Moderately Related:	<i>Joey’s crazy mother became furiously angry with him.</i>
Distantly Related:	<i>Joey went to a neighbor’s house to play.</i>

In previous behavioral studies using similar materials, Keenan, Baillet, and Brown (1984) and Myers, Shinjo, and Duffy (1987) found that reading times on the sentences increased as the degree of relatedness between the sentences decreased; that is, the

reading times increased from the highly to moderately to distantly-related conditions. Paradoxically, however, the participants' memory for the two-sentence passages (as measured using a variety of recall and recognition tests) followed an inverted U-shaped function; that is, the sentences in the moderately-related condition were remembered better than those in both the highly- and distantly-related conditions. Thus, the participants' memory for the sentences does not seem to be a simple monotonic function of either their reading times or the degree of causal relatedness between the two sentences being read.

In the fMRI study of causal relatedness, Mason and Just (2004) found three main foci of fMRI-measured cortical activation among the language areas. In the left hemisphere language areas, the activation volume did not vary across the three relatedness conditions. In contrast, the activation volume in the bilateral dorsolateral prefrontal cortices showed a marked (albeit not statistically reliable) increase as the sentences became less causally related. Finally, the most interesting pattern of activation volume was observed in the right hemisphere homologues of the language areas: the activation volume was consistent with the patterns that had been reported with recognition and recall measures.

To account for their fMRI data, Mason and Just (2004) proposed that two different cortical networks support the generation and the integration of inferences during reading. The first network, consisting of the left and right dorsolateral prefrontal cortex is more involved in generating the inferences. As the causal distance between two sentences increases, the dorsolateral prefrontal regions generate more inferences, leading to an increased volume of activation. The second network, consisting of the right inferior frontal gyrus, right superior and middle temporal gyri, and right inferior parietal lobe, is more involved in integrating the possible inferences that have been generated. Because of the relative differences in integrating the inferences in the three conditions, the volume of cortical activation that was observed in this region is described by an inverted U-shaped function, with more activation occurring with the moderately related than either the highly related or distantly related sentences.

Reichle and Mason (2005) present a working memory account for the additional right hemisphere activation that occurs as a result of generating an inference in the moderately related condition. They suggest that, as proposed by Just et al. (1996), there is an inherent limit on how much cognitive processing can be done per unit of time in left hemisphere language areas. In the context of text processing, this limit means that to the degree that working memory resources are being used to process the text and generate inferences, those resources will not be available for integrating those inferences into long-term memory. Reichle and Mason propose a computational model to demonstrate that these resources are exceeded only in the moderately related condition, such that the remaining processing load that cannot be accommodated in the left hemisphere spills over into the right hemisphere.

Research is only now beginning to map out the cortical network associated with drawing inferences in reading. Evidence to date suggests that the right hemisphere plays a key role in such a process (Beeman et al., 1994; Mason & Just, 2004). At this point, at least

three plausible theories have been developed to illuminate the right hemisphere's role. Reichle and Mason (2005) building on work by Just et al. propose that a limited capacity is exceeded across inferencing component processes. This then leads to processing being passed to the right hemisphere. This is in contrast to an account proposed by Mason et al.; they propose that the right hemisphere is utilized during the integration of an inference and that inferences are generated by utilizing the dorsolateral prefrontal cortex. Finally, Beeman's coarse coding theory leads to the prediction that inferences are accomplished as the result of activation of coarsely coded semantic information in the right hemisphere. In this account, it is not the inference per se that is processed in the right hemisphere but rather the information from which the inference is developed. A further account combines aspects of these previous explanations. Inference generation is supported by the right hemisphere coarse semantic network. Additional attempts to utilize this network will be signaled by the dorsolateral prefrontal coherence monitor, provided resources are available. As the propositionalization work of the left anterior temporal text integration network becomes more demanding (and resources are consumed), this processing will spillover into the right anterior temporal region. It is clear that further research will be needed to specify at which level of difficulty each of the networks are engaged for various readers' abilities.

## 5. NEW PERSPECTIVES ON FIGURATIVE LANGUAGE PROCESSING

The study of metaphor comprehension has long been a major area of interest in behavioral studies of figurative language (e.g., Allbritton, McKoon, & Gerrig, 1995; Gerrig & Healy, 1983; Gibbs, 1990). With the development of recent theories and methods that illuminate the role of the right hemisphere in discourse processing (see Beeman & Chiarello, 1998 for an extensive listing), the right hemisphere's role in metaphor comprehension arises as a central issue. As mentioned previously, the study of the neural basis of metaphor processing has some neuroimaging precedent. Bottini et al. (1994) showed that the processing of novel metaphors resulted in an increase in right hemisphere activation. The finding from this early PET study was one of the reasons that metaphor processing has been recently described as a right-hemisphere language function. In fact, Beeman (1998) listed metaphor processing as one of the functions for which the right hemisphere is well suited. However, that is turning out to be too simple a view of a complex process.

A recent brain imaging study conducted in our laboratory indicated that the comprehension of frozen metaphors activated the same left hemisphere language areas that were active in the processing of literal sentences, with the activation being more extensive for the frozen metaphors in the left inferior frontal gyrus (Eviatar & Just, 2006). Unlike the results of Bottini et al. (1994) which used novel metaphors, the frozen metaphors in the more recent study did not activate the right hemisphere posterior superior temporal areas more than literal passages. Eviatar and Just concluded that the processing of frozen metaphors required semantic selection of a more abstract meaning associated with the figurative phrase. In the case of frozen metaphors, these frequently used abstract meanings are lexicalized. Further, they proposed that the comprehension of such a metaphor would require



selection of appropriate aspects of the meaning and suppression of the inappropriate, or literal, meaning (Gernsbacher & Robertson, 1999). It is presumably for this reason that additional activation was observed in the left inferior frontal gyrus, an area associated with the selection and suppression of lexical content (Thompson-Schill, 2003; Keller, Carpenter, & Just 2001). Rapp, Leube, Erb, Grodd, and Kircher (2004) also examined novel metaphors in simple “An A is a B” sentence frames. They asked their participants to judge whether their metaphors had a positive or negative connotation. Even though they used novel metaphors, Rapp et al. found higher activation for metaphoric versus literal sentences in left inferior frontal gyrus and left inferior temporal gyrus but not in the right hemisphere. There were several differences between the Rapp et al. and Bottini et al. studies, but the largest may have been inconsistencies in syntax of the Bottini et al. materials. Although many of Bottini et al.’s metaphors were of the “A is a B” variety, a large number of them were presented in more complex syntax such as “The old man had a head full of dead leaves.” Rapp et al.’s lack of finding a right hemisphere effect in metaphor comprehension may have arisen due to the fact that a broader situation model did not have to be constructed to understand the metaphor in isolation. The simple metaphors only require an equation of two concepts without discourse or even syntactic processing.

Mason, Eviatar, and Just (2005), in order to reconcile the Bottini et al. novel metaphor result and the Eviatar and Just frozen metaphor results, contrasted cognitive processing during the reading of literal sentences with two different types of figurative language: novel metaphors that are created *de novo*, and frozen metaphors, which have been previously encountered and may have a stored representation. The stories contained three sentences. The first two sentences were presented simultaneously, and constituted the context for the third sentence, which was always a statement uttered by one of the characters. The character’s utterance always contained either a frozen metaphor, novel metaphor, or a literal phrase.

Frozen metaphor:

*Mary got straight A’s on her report card.*

*Her parents were proud of her.*

*They said, “You are as sharp as a razor.”*

Novel metaphor:

*It was Judy’s first time on an airplane.*

*Her mom let her have the window seat.*

*Judy said “We’re surrounded by great white mushrooms.”*

Literal:

*Johnny went on a hike with his brother.*

*Suddenly he saw a huge snake next to his foot.*

*He said, "I have always been afraid of snakes."*

As in the Eviatar and Just results, Mason et al. found that when reading a frozen metaphor passages, the same language processing areas are active that are active during normal reading (e.g., dorsolateral prefrontal cortex bilaterally, left middle and superior temporal lobe as well as left inferior frontal gyrus). In addition to the shared language processing areas, we found additional activation for frozen metaphors in right middle and superior temporal lobe and superior medial frontal gyrus and the paracingulate area. But the full story was much more complex and interesting. The novel metaphors resulted in primarily visual-spatial activation, suggesting that visual imagery processes were being used to instantiate and/or interpret the novel metaphors used in the study. In contrast, the frozen metaphors were associated with activation in a superior-medial frontal cortex. This is the same region as the proposed protagonist interpreter network and has often been associated with theory of mind processing, and indeed, the frozen metaphor passages tended to refer to a character's traits (e.g., "You are as sharp as a razor."). These results demonstrate a consistently emerging pattern in discourse processing research; during discourse processing, a complex set of cortical networks are dynamically recruited depending on qualities of the text and the reader's goals.

The few studies conducted on figurative language suggest that processing of metaphors within text utilizes the same cortical regions as do several other discourse tasks. In particular, trying to understand a metaphor has resulted in engagement of right hemisphere regions (Bottini et al., 1994; Mason et al., 2005) as well as increased processing in the left inferior frontal gyrus and left temporal regions (Bottini et al., 1994; Eviatar & Just, 2006; Mason et al., 2005; Rapp et al., 2004). Recently, Mashal, Faust, and Hendler (2005) also found a selective right hemisphere involvement in the processing of novel metaphors and a left hemisphere involvement in the processing of conventional metaphors. Specific types of metaphors also seem to activate a region in or near the medial frontal gyrus (Mason et al., 2005). As mentioned in the previous section, this frontal region was also active during the processing of inconsistent emotional information (Ferstl, Rinck, & Von Cromon 2005). Mason et al. suggested that their frozen metaphors activated this region in part because the frozen metaphors they used were rated as high on an emotional content scale. Together these results indicate that, much like the case with other discourse tasks, text variables must be carefully controlled in experiments investigating figurative processing.

Coulson and Van Petten (2002) propose that conceptual integration results from a process of *alignment* (Gentner & Wolff, 1997) or *mapping* in conceptual blending theory (Coulson, 2000). Conceptual blending theory involves the establishment of a

blended space into which concepts and relations from both the literal concepts and the target metaphoric concepts are imported. The blended space can then be combined with background knowledge so as to understand the metaphor. As suggested by Coulson and Van Petten (2002), a blended space may be necessary to understand metaphors. It is possible that this blended space may be no different than an episodic-based situation model. The additional processing required to either maintain two mental models (the literal and the figurative) in parallel or the need to combine and supplement the model with information from background knowledge may account for the observed increase in cortical activation in response to reading a metaphor. In the case of novel, discourse based metaphors, the additional right hemisphere activation may be a spillover of processing from the text integration area. In this case the interpretation of a situational model metaphor will be similar to a complex inference. Furthermore, other areas that become active in metaphor processing will be a function of the text. Metaphors require interpreting the perspective of the protagonist may be accompanied by an increase in activation of the medial frontal protagonist interpreter network. The reason for the recruitment of specific areas is still partially speculative at this point. Future research in this area will be necessary to determine how much of the right temporal activation and the medial frontal activation are due to figurative language processing per se, and how much is simply a result of text factors.

## 6. NEW PERSPECTIVES ON LATERALIZATION

The classical view, derived in 19th century clinical studies of aphasia, held that the left hemisphere is dominant for both comprehension and production of language. It is now clear that the right hemisphere plays a role in language processing as well (Gardner, Boller, Moreines, & Butter, 1973) supporting semantic operations (Koivisto, 1998) particularly global processes like inference, coherence, conceptual association, text integration (St. George et al., 1999) and prosody (Hesling, Climent, Bordessoules, & Allard, 2005; Plante, Creusere, & Sobin, 2002). Right hemisphere involvement during narrative processing has been noted in several neuroimaging studies, both for comprehension (Bottini et al., 1994; Mazoyer et al., 1993; Nichelli et al., 1995; St. George et al., 1999) and production (Braun, Guillemin, Hosey, & Varga, 2001). The right hemisphere involvement may reflect coherence and inference-related processes at the discourse level, such as when readers make connections between sentences, integrating these into a global representation, processing metaphors, and otherwise utilizing information not encoded in the text.

In addition to the neuroimaging data, there is clear evidence of a right hemisphere role in discourse comprehension from neuropsychological data. For example, patients with right-hemisphere damage have difficulty connecting and integrating semantically distant concepts (Brownell & Martino, 1998; Beeman, 1993; Bihle, Brownell, Powelson, & Gardner, 1986; Brownell, Michel, Powelson, & Gardner, 1983). However, patient studies have failed to reach a clear conclusion regarding the contribution of the right hemisphere homologues of Broca's and Wernicke's areas to discourse processing.

Beeman's coarse coding hypothesis (Beeman, 1998) applied to language is an important theoretical contribution that relates the cognitive processing of discourse to its neural basis. The coarse coding hypothesis proposes that the two hemispheres differ in the level of granularity at which they code semantic information. Beeman proposes that the left hemisphere uses fine semantic coding to quickly select a small number of relevant meanings, whereas the right hemisphere uses a coarse semantic coding scheme in which it weakly activates a broad spectrum of meanings and features (Beeman, 1993, 1998).

According to the coarse coding hypothesis, word meanings are represented bilaterally. In the left hemisphere, word meanings are represented by localized semantic fields so that their core meanings can be rapidly and reliably accessed. In the right hemisphere, word meanings are represented by more distributed (and possibly overlapping) semantic fields. The coarse right hemisphere semantic field allows for more than one sense of a word's meaning to be accessed. These coarse semantic fields facilitate processing of figurative language and are particularly useful in solving insight problems.

Strong support for the right hemisphere based coarse coding hypothesis comes from priming studies in which the hemisphere to which words are passed is controlled via visual field presentation. For example, when processing ambiguous words, priming occurs for subordinate meanings of ambiguous words after 750 ms when presented to the left visual field/right hemisphere (LVF/RH) but not in the right visual field/left hemisphere (RVF/LH) (Burgess & Simpson, 1988). Additionally, several weakly associated words prime a concept word (e.g., *cry*, *foot*, and *glass* together prime *cut*) when the words are displayed to the LVF/RH, but not when the words are displayed to the RVF/LH (Beeman et al., 1994). Similarly, distantly related concepts prime each other (e.g., *deer* primes *pony*) over longer time intervals in the LVF/RH than in the RVF/LH (Beeman et al., 1994; Chiarello, Burgess, Richards, & Pollock, 1990; Nakagawa, 1991). Beeman suggests these results indicate that right hemisphere maintains less central aspects of a word's meaning and distant associates longer than the left hemisphere.

It is less clear how a representation based hypothesis, such as coarse coding is related to inference making. Here neuroimaging results provide a database on which theories of discourse processing can be constructed, possibly using the coarse coding hypothesis as a fundamental assumption. For example, the broadly distributed, partially overlapping semantic fields in the right hemisphere are ideally suited to allow the cortical activation from several distantly related and/or weakly activated concepts to accrue and converge, bridging whatever semantic information happens to be represented in the fields. One possibility is that the right hemisphere semantic network may have developed in parallel to the left hemisphere semantic network. Because the left hemisphere plays a role in phonological processing and the majority of language input during development is auditory, the left hemisphere may have developed a finer grained semantic representation system. The coarser semantic representation of the right hemisphere semantic network then becomes a strength for the system in the case of inferencing. In many cases, generating possible inferences requires connection of distant features of words or relations between concepts. The coarse representation in the right hemisphere more easily supports this type of connection.

While many neuroimaging studies of language have tended to support the traditional notion of strong left hemisphere lateralization, this may be due to the superimposition of meta-linguistic tasks in these studies. Long and Baynes (2002) proposed that although situation model processing involves an interaction with the right hemisphere, input from the left hemisphere is also required. It is therefore not surprising that even in discourse comprehension tasks, the left hemisphere remains highly active (Mason & Just, 2004). Not only does the left hemisphere remain active, but, as the results of many of the previously mentioned neuroimaging studies have shown, there is additional left hemisphere activation accompanying the right hemisphere activation (e.g., Xu et al., 2005).

Xu et al. (2005) showed that even at the single word level, text comprehension naturally engages both hemispheres (although in this context, responses are still markedly lateralized to the left). Right hemisphere activation becomes prominent when words are presented in a sentential context, and may reflect coherence and inference at the propositional level during which readers make connections within sentences to form coherent representations. But Xu et al. found that it was during the processing of narrative that right hemisphere activity was most robust. Reading of the narrative was associated with strong bilateral activations throughout the brain, encompassing perisylvian, extrasylvian, premotor cortices, and cerebellum, indicating that both linguistic and extralinguistic processes play a role in discourse comprehension.

Xu et al. (2005) developed a step-by-step account of how a set of brain structures, particularly the right hemisphere, functions during the reading of a narrative text. They analyzed their passages using the formal structural measures developed to determine the story grammar or structural regularities in the narrative content (Mandler & Johnson, 1977; van den Broek, 1994). On the basis that

“text comprehension, as defined by Kintsch and Van Dijk (1978) must interact with formal structure: that is, since a mental model of the narrative is constructed in increments and adapted as a story unfolds, we reasoned that a reader’s cognitive effort would change—that is, both language and language-related processes should be differentially engaged during the succession of narrative segments and should be reflected in dynamically fluctuating patterns of brain activity” (Xu et al., 2005, p. 1013).

Another possible reason for the right hemisphere’s involvement in discourse comprehension tasks comes from studies of syntactic processing. Just and colleagues (Just & Carpenter, 1992; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Reichle, Carpenter, & Just, 2000) have developed a theory in which there is an inherent limitation on how much cognitive processing can be done per unit of time. Just et al. (1996) have shown that the right hemisphere becomes active in sentence comprehension as the syntactic processing demands of the sentence increase. It may be the case that additional right hemisphere activation in discourse comprehension arises in response to the left hemisphere’s capacity limitations. It is also important to note that the principle of working memory’s limited capacity has been widely used in existing theories of text processing (Frank, Koppen,

Noordman, & Vonk, 2003; Golden & Rumelhart, 1993; Goldman & Varma, 1995; Kintsch, 1988; Kintsch & van Dijk, 1978; Langston & Trabasso, 1999; Myers & O'Brien, 1998; Schmalhofer, McDaniel, & Keefe, 2002; van den Broek, Risden, Fletcher, & Thurlow, 1996; Tzeng, van den Broek, Kendeo, & Lee, 2005). Thus, in the example of inference processing, to the degree that the left hemisphere's working memory capacity is consumed by the processing of text and the generation of inferences, there may not be enough capacity left to integrate those inferences. As a result, inference processing may invoke right hemisphere processing.

Ferstl et al. (2005) specifically propose that the right anterior temporal activation they find with inconsistent situational model level information is a direct reflection of cross hemisphere spillover of processing. They suggest that propositionalization is a left anterior temporal specialization. When situational model information is inconsistent, the propositionalization of incoming text information is more difficult and it results in the engagement of the right anterior temporal lobe. It is likely that the activation of the right hemisphere in many inference tasks could be a result of both activation of coarse coded semantic information as well as spillover of situation level propositionalization. It remains to be seen if these two hypotheses can be separated.

One dominant point in the neuroimaging discourse research is that the right hemisphere plays a role in discourse comprehension. This has been seen in many of the experiments presented in this chapter. The coarse coding theory (Beeman, 1998) and the spillover of processing theory (Just et al., 1996) provide two recent accounts that attempt to provide a framework that allows for the right hemisphere homologues of Broca's and Wernicke's areas to become active in several different experimental manipulations. It is important to note that these frameworks are not intended to be in place of previous discourse theories (e.g., Kintsch, 1988; Kintsch & van Dijk, 1978) but rather to provide an explanation of underlying component processes present in these discourse theories. Moreover, they are also consistent with some attempts to tie internal representations of the discourse to specific regions, such as Grafman's script based Meaning-Knowledge-Units (Grafman, 1995), situation models (Ferstl et al., 2005; Schmalhofer, 2003), and a bilateral discourse model (Long & Baynes, 2002). This new perspective on lateralization enables us to propose that the right temporal lobe is involved in discourse processing in both a coarse semantic processing network as well as spillover of the a left anterior temporal text integration network.

## **7. A NEW COMPONENT OF DISCOURSE PROCESSING: PROTAGONIST INTERPRETER**

Narrative-specific activations have been consistently found in the medial frontal gyrus, and precuneus/posterior cingulate cortices (Ferstl & von Cramon, 2001, 2002; Ferstl et al., 2005; Xu et al., 2005). The medial frontal gyrus has been shown to play a role in theory of mind processes (Fletcher et al., 1995), defined as the capacity to intuit the beliefs, desires, and goals and predict the actions of others. An everyday understanding

of others' minds is clearly necessary for interpreting the intentions, goals, and actions of characters within a narrative. This same area has also been found to be activated in the comprehension of metaphor (Bottini et al., 1994) and identification of thematic roles within a story (Nichelli et al., 1995). All of these are considered to be discourse level variables that are likely to be engaged during narrative comprehension.

This theory of mind network has been reported to be engaged in a range of cognitive functions that include 'mentalizing' (Castelli et al., 2002), that is, the ability to attribute mental states to others (essentially theory of mind), understanding social concepts (Martin & Weisberg, 2003) and making moral judgments (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Moll, de Oliveira-Souza, Bramati, & Grafman 2002) and on this basis it has been argued to play a central role in social cognition. Yet, as mentioned previously, this system has also been shown to participate in cognitive processes that lie outside of the social domain, from inferring logical relationships between events or propositions (Ferstl & von Cramon, 2002) to evaluating and verifying facts based on personal knowledge and experience (Zysset, Huber, Ferstl, & Von Cramon, 2002).

Activation in this medial frontal area has also been found in tasks that involve emotional processing and more specifically, emotion related to memory (Canli, Desmond, Zhao, & Gabrieli 2002; M. Nakic & J. Grafman, unpublished manuscript). Activation in this area has also been found when participants had to integrate inconsistent emotional information in stories (Ferstl, et al., 2005). Thus the additional activation in medial frontal areas might be attributed to the activation associated with emotion-related processing. As previously mentioned, frozen metaphors that were high in emotionally based content also resulted in activation in the medial frontal gyrus (Mason et al., 2005).

This medial frontal area has also been shown to be active in a number of tasks that require representing the mental states of others, or Theory of Mind (Gallagher & Frith, 2003). Unlike the novel metaphors which require visualizing, frozen metaphors are much like ambiguous words and require abstraction. The ability to understand this type of abstraction requires that the reader be able to mentalize the characters' intentions and internal emotional state within the story.

The fact that the medial frontal region is often activated in discourse tasks, social cognitive processing, and theory of Mind suggests that this cortical area plays a general role that would be common in all of these tasks. Xu et al. (2005, p. 1012) suggest that the medial prefrontal cortex operates "at the interface between self and environment, yoking a variety of cognitive processes to knowledge about the world – a function that is clearly central to narrative comprehension." In this sense, processing language as discourse would be expected to engage systems that lie outside the language cortices.

The research demonstrating the activation of the medial frontal area in discourse processing shows just how powerful the neuroimaging approach is. Although, this area was found to be activated in many studies of discourse comprehension, it was not previously considered a language processing area. The medial frontal area has been more closely

linked with the processing of information that required an ability to examine a situation from a different perspective (e.g., Castelli et al., 2002). It has also been found to be active in tasks that require an understanding of emotional and moral aspects of a situation (e.g., Greene et al., 2001). The recent discourse work has shown that this area activates during narrative text and becomes differentially active in response to a manipulation of character centered emotional variables within a text. However, activation of this region in non-discourse tasks (Krause et al., 1999) as well as “inanimate” texts (Ferstl & von Cramon, 2002) indicates a domain general role. It seems evident that within narrative text, this domain general cortical region activates strongly in response to understanding protagonist oriented stimuli.

## **8. OUTLINE OF A NEUROCOGNITIVE ACCOUNT OF DISCOURSE COMPREHENSION**

Our goal here is to provide a theoretical framework based on the results reviewed in this chapter. Although this framework is predominantly based on neuroimaging results, it is intended to be consistent with discourse theories that have arisen from behavioral, neuropsychological, and neuroscience research. The purpose of this framework is to integrate these different research approaches and to extend current conceptualizations of discourse processing.

This framework is consistent with a more general approach in which it is assumed that the nature of the text and the goals of the reader affect the extent to which specific cortical regions are activated when reading. The potential exists for many different cortical networks to become active during reading; the theory should ultimately specify the conditions under which a particular network is engaged to make up the whole of the system.

Although the understanding of the complex nature of discourse processing at the cortical level is a lofty long-term goal, the outlines of a theoretical account are beginning to emerge. Our proposed separation of the discourse level of processing from various lexical and syntactic processes that underlie comprehension is only a temporary simplification, because there is surely interaction among these two broad categories of processing. All levels of processing consume resources to various degrees and affect the availability of resources required by discourse processes. It is quite likely that lexical-semantic process and syntactic processes will at times consume more resources than simple grapheme/phoneme processing as well as being constrained at times by contextual influence from the discourse-level processes. In fact, both the graded salience hypothesis (Giora, 1997) and the coarse-coding process (Beeman, 1998) can be viewed as lexical-semantic level processing which is constrained by context and likewise are utilized in developing a model of the text.

### **8.1. Basic LH Sentence Network (Not Uniquely a Part of Discourse Processing)**

As a text is being read, the individual words are being identified, the syntactic structure is being parsed and word meaning is being extracted. The basic reading processes are



primarily left hemisphere functions. These basic processes include visual/graphemic processing, phonological processing, lexical-semantic processing and syntactic processing. Obviously, discourse level processing does not wait until all the lower levels are complete; discourse processing occurs on a word-by-word, moment-to-moment level in parallel with the lower levels of language processing. As each word is read, an interpretation of the word within the context of the passage is constructed. This interpretation is informed by a salience-based lexical access process, utilizing several left hemisphere regions.

## **8.2. Coarse RH Semantic Processing Network**

The relevant coarse semantic field for each word is activated in the right hemisphere. The possibility for an inference or alternative interpretation of a concept arises based on the degree that this newly active coarse semantic field overlaps with a recently activated coarsely-coded semantic field (either based on previous text or perhaps activated in response to world knowledge or schema related to the context or topic of the passage). Thus, whenever enough information accrues to support the generation of an inference, additional activation should be seen in the right temporal region. At this point several other accompanying cortical networks might activate in parallel.

## **8.3. Dorsolateral Prefrontal Coherence Monitor Network**

The dorsolateral prefrontal cortex will increase activity bilaterally in response to either a lack of coherence in the text (signaling the need for additional right hemisphere activity) or as a result of an unusually active semantic field in the right temporal region. This dorsolateral prefrontal cortex activation can be viewed as goal directed behavior with respect to guiding the inference construction process. The guiding and reiterative nature of this process is dependent on available cognitive resources. As long as resources are available, the inferential process proceeds until a successful inference has been integrated and a reader-based standard of coherence has been achieved. If resources are consumed, the reader continues on through the text with whatever current interpretation exists; later information either supports the current interpretation or results in additional signals to an inference generation process, again checking available resources. This iterative process continues until coherence is achieved or a complete breakdown in comprehension occurs.

## **8.4. Left Frontal – Temporal Text Integration Network**

In addition to the lexical access and parsing which are traditionally viewed roles of the left hemisphere, the left inferior frontal gyrus and the left anterior superior temporal sulcus region extending into the temporal pole maintains, constructs and integrates information into the reader's understanding of the text. This "understanding" may be similar to the situation model or mental model of the text. It is still unclear where this representation is "stored" in the cortex. There have been several proposals concerning the storage of the situation model. Grafman (1995) has proposed frontal cortex storage of schema-level representations in meaning knowledge units (MKUs). Schmalhofer (2003)

has proposed a right-hemisphere storage of the situation model, and Long and Baynes (2002) has suggested a bilateral representation of situation model.

Our hypothesis is that this representation is stored diffusely, distributed over areas of the cortex specifically suited to the nature of the information (e.g., spatial information in right parietal, emotional information in the amygdala/medial frontal cortex, etc.). Thus, these left hemisphere regions reach a higher level of activity when an inference is integrated into the discourse representation than during normal reading. This higher level of activity is only seen with respect to text that does not exceed the reader's capacity to process it; in contrast, when resources are unavailable to integrate the inference (due to either text that is difficult at another level or due to a reader's reading ability), integration of a possible inference fails or processing may be passed to the right hemisphere.

### **8.5. Medial Frontal Protagonist/Agent Interpreter Network**

The medial frontal gyrus is active during most narrative processing. This region seems to be particularly well-suited to processing information related to understanding another's plans and motivations. This can also be viewed as comprehension of an alternative reality, specifically the world-view of a protagonist within a text. Any inference that would be related to a characteristic specific to a protagonist within the story should result in activity in this region in response to the process of updating the *protagonist model*. Typically, Theory of Mind tasks also activate portions of the right posterior, superior temporal gyrus and the right inferior parietal lobe. It is likely that these areas are also part of this protagonist model network, but additional experiments are necessary to determine the full extent of the network. The medial frontal region has also been seen to increase its activity in response to text that is particularly emotionally oriented or one that requires the reader to reference emotionally based memories (Ferstl & von Cramon, 2002).

### **8.6. Intraparietal Sulcus Spatial Network**

Whenever a reader encounters sentences that load heavily on a spatial referent, activation appears in the left intraparietal sulcus area (Just et al., 2004). Just et al. presented readers with sentences like, "The number eight when rotated 90 degrees looks like a pair of spectacles." These high-imagery sentences resulted in additional activation bilaterally in the intraparietal sulcus relative to low-imagery sentences, although it was much stronger in the left hemisphere. Mason et al. (2005) also found additional left intraparietal sulcus activation for their novel metaphors, which were rated to be much more visualizeable than their literal sentences and frozen metaphor sentences. It is likely that the left intraparietal sulcus activates on most narrative texts due to the spatial information likely to be encoded in a situational model (Zwaan & Radvansky, 1998; Zwaan & Singer, 2003). This spatial network has not been given as much attention as the other networks in discourse processing research but should be expected to function in a similar manner as the other parallel networks proposed here. It should activate as a function of the text constraints as well as the reader's individual ability.

This theoretical framework is by no means a complete picture of discourse processing and the functions of several networks are still highly speculative. It is offered as a possible overview of how the various cortical networks may function in discourse. It is likely that several of the descriptions here will need to be revised in the light of additional data.

## 9. RELATION TO OTHER THEORIES

Throughout this chapter we have described research in this area that is consistent with various discourse processing theories. In addition, the research contained in this chapter has been fundamental in developing new concepts such as the coarse coding theory of right hemisphere process, the dynamic recruitment of cortical networks in response to text constraints, the spillover of processing to other differential specialized networks in response to capacity utilization, and the utilization of a Theory of Mind network in discourse comprehension. What remains is to examine how the recent cortically based perspectives on discourse processing fit into the various discourse processing theories.

The CI model (Kintsch, 1988; Kintsch & van Dijk, 1978) has certainly been one of the most influential models of discourse processing. According to this model, the process of comprehending written text proceeds in two stages. The first is an initial construction stage in which the propositions of the text base are combined with the reader's knowledge base to construct a loose representation of the text. It is during this stage that whatever inferences are necessary to connect disjoint pieces of the text are generated. Then, during a subsequent integration stage, this representation is "pruned" to remove any propositions or connections that are inconsistent or contradictory. This second stage of processing is thought to result in a representation that is coherent. A relatively undefined aspect of the CI model is an initial liberal generation stage which occurs prior to the construction phase. This liberal generation stage presumably allows the generation of many inferences due to an interaction of the text and world knowledge. Readers attempt to construct and integrate only those inferences for which there is enough connection with the text. The CI model centrally addresses the types of processes executed by the coarse RH semantic processing network, the frontal (dorsolateral) coherence monitor, and the left frontal-temporal inference support network.

The 3CAPS model (Goldman & Varma, 1995) builds upon the CI framework by incorporating the construction and integration stages within a more general cognitive architecture (Just & Carpenter, 1992). Discourse processing within this model is mediated by *productions*, or condition-action rules, that operate on the contents of a limited-capacity working memory system. Working memory is thus conceptualized as consisting of a limited pool of processing resources that can be allocated dynamically in the service of generating inferences. This position is entirely consistent with the proposal that the cortical networks in the brain are utilized via an interaction between the text variables and the availability of resources.

The resonance model (Myers & O'Brien, 1998) also shares some similarities with the CI model and the 3CAPS-CI model. Like both CI models, text comprehension is a result of constructing a representation of the text on the basis connecting incoming text information with prior text information as well as information from the reader's world knowledge. Furthermore, the information that is contained in a short-term memory store at any one time is limited. The resonance model is consistent with the initial liberal inference generation mechanism of the CI model. O'Brien, Cook, and Peracchi (2004, p. 290) describe the resonance model as:

passive reactivation processes [which] cannot be shut off. The signal that emanates from active components in memory is not triggered by coherence breaks, nor is it guided by relevance; it is continuous, autonomous, and unrestricted. Any related information that resonates in response to this signal has the potential to be activated, independent of its relevance. Outdated information is no different than any other information; if the reader encounters a target sentence that is related to the outdated information, the target sentence can serve to reactivate that information, even if reactivation ultimately disrupts comprehension.

Perhaps most interestingly, these passive memory based retrieval/generation processes are similar to the manner in which the coarse-coding hypothesis describes the activation of the coarsely coded semantic fields in the Coarse RH semantic processing network. Specifically, coarse-coded information in the right hemisphere is activated continuously, autonomously, and in an unrestricted manner. The information has the potential to be activated, independent of its relevance. However in coarse coding "independent" may not be the correct concept; the information has to be part of an overlapping coarsely coded semantic field. It remains to be seen how independent these fields are.

It would be convenient if there were a set of processes that always occurs whenever anyone is processing a text and if these processes could be mapped onto specific cortical regions. This approach, however, fails to acknowledge the dynamic nature of discourse comprehension and brain function. The brain, particularly in text comprehension, is an adaptive machine. For example, we can say that, in general, drawing an inference should result in right temporal activation, but it would be incorrect to make a simple claim such as the right temporal area is where an inference is generated. In fact, we have shown that at least three plausible theories have been developed to illuminate the right hemisphere's role in inference construction: the Reichle and Mason (2005) limited capacity inferencing component processes hypothesis; the Mason and Just (2004) right hemisphere inference integration hypothesis; and the coarse coding theory (Beeman, 1998). A similar account applies to the processes underlying figurative language comprehension (Bottini et al., 1994; Rapp et al., 2004). Trying to understand a metaphor has resulted in the engagement of right hemisphere. But here too, it would be too simplistic to claim that the right hemisphere is where a metaphor is processed. Specific types of metaphors also seem to activate a region in or near the medial frontal gyrus (Mason, Eviatar, & Just, 2005), a region which also activates during the processing of inconsistent emotional information (Ferstl et al., 2005). And while the medial frontal area is active during the

processing of some types of metaphors, it may be due to the type of the metaphor rather than a metaphor per se.

This leaves us with the rather difficult task of proposing that if a text has high emotional content, invites an elaborative inference, and has a clear topic, we should see a specific type of network active. Developing a set of networks that process discourse would be as many and as varied as the texts themselves. The better approach would be to build upon already existing theories of discourse and show how evidence from brain regions support components of the various theories.

Giora's graded salience hypothesis (Giora, 1997) provides another example of how these new perspectives on discourse theories are consistent with previously proposed models. The graded salience hypothesis proposes that two different mechanisms, one modular linguistic mechanism (a bottom-up, perceptual lexical access of the salience ordered mental lexicon) and one global, contextually based access system that operates in parallel to the lexical access. Peleg, Giora, and Fein (2001) notes that it is not a specific word's predictiveness that is important for contextual facilitation but rather a concept's availability and predictability with respect to previous world-based encounters. Consider how this might function with respect to metaphor interpretation. It is possible that the access of a concept proceeds both locally in the left hemisphere and globally in the right hemisphere. Again, the similarity between this account and the processing of the coarse RH semantic processing network is clear.

## 10. CONCLUDING REMARKS

Discourse processing is a complex high-level cognitive task in that many facets of cognition are involved. The ability of neuroimaging research to examine a large scale cortical network of differentially specialized brain regions while manipulating the loading of the various facets of cognition makes the methodology ideally suited to advancing our understanding of discourse comprehension. The diverse nature of the research reviewed in this chapter shows that we are only beginning to bring to bear the strengths of neuroimaging on our understanding of comprehension.

At the outset, we proposed that neuroimaging allows a new way to study old theories as well as providing data which might suggest new theories underlying discourse comprehension. The ability to use neuroimaging to examine whole cortical networks enables speculation as to how the various component processes of diverse discourse theories can be integrated into a single whole. Much of the neuroimaging research completed so far can thus serve as existence proofs for several proposed discourse processes. While neuroimaging research may not yet be at the stage of "falsifying" a theory, it can be used to support several theories.

In conclusion, it is quite evident that neuroimaging offers an ability to investigate discourse processing in a manner that has not been done before. The greatest advantage of

this research approach is the ability to examine the network as a whole, revealing new aspects of discourse processing.

## ACKNOWLEDGMENTS

This research was supported by the National Institute of Mental Health Grant MH029617. We would like to thank Zohar Eviatar and the current members of the CCBI reading group for helpful comments on a previous draft of this chapter.

## REFERENCES

- Allbritton, D. W., McKoon, G., & Gerrig, R. (1995). Metaphor-based schemas and text representations: Making connections through conceptual metaphors. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*, 1–4.
- Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences during comprehension. *Brain and Language*, *44*, 80–120.
- Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In: M. Beeman, & C. Chiarello (Eds), *Right hemisphere language comprehension*. Mahwah, NJ: Erlbaum.
- Beeman, M., & Chiarello, C. (1998). Right hemisphere language comprehension: Perspectives from cognitive neuroscience. Mahwah, NJ: Lawrence Erlbaum.
- Beeman, M., Friedman, R. B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M. B. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience*, *6*, 26–45.
- Bihrlé, M. A., Brownell, H. H., Powelson, J., & Gardner, H. (1986) Comprehension of humorous and non-humorous materials by left and right brain-damaged patients. *Brain and Cognition*, *5*, 399–411.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188.
- Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S. J., & Frith, C. D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: A positron emission tomography activation study. *Brain*, *117*, 1241–1253.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.
- Braun, A. R., Guillemin, A., Hosey, L., & Varga, M. (2001). The neural organization of discourse: An H215O-PET study of narrative production in English and American Sign Language. *Brain*, *124*, 2028–2044.

- Brownell, H. and Martino, G. (1998). Deficits in inference and social cognition: The effects of right hemisphere brain damage on discourse. In: M. Beeman, & C. Chiarello, (Eds), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*, (pp. 309–328). Mahwah, NJ: Lawrence Erlbaum.
- Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right hemisphere patients. *Brain and Language*, 18, 20–27.
- Brownell, H. H., Potter, H. H., Bihle, A. M., & Gardner, H. (1986). Inference deficits in right brain-damaged patients. *Brain and Language*, 29, 310–321.
- Buckner, R. L., Bandettini, P. A., O’Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., & Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences, USA*, 93, 14878–14883.
- Burgess, C., & Simpson, G. B., (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language*, 33, 86–103.
- Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12, 1–47.
- Canli, T., Desmond, J. E., Zhao, Z., & Gabrieli, J. D. (2002). Sex differences in the neural basis of emotional memories. *Proceedings of the National Academy of Sciences, USA*, 99, 10789–10794.
- Caplan, R., & Dapretto, M. (2001). Making sense during conversation: An fMRI study. *NeuroReport*, 12, 3625–3632.
- Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125, 1839–1849.
- Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some don’t sometimes, some places. *Brain and Language*, 38, 75–104.
- Cohen, M. S., & Bookheimer, S. Y. (1994). Localization of brain function using magnetic resonance imaging. *Trends in Neurosciences*, 17, 268–277.
- Coulson, S. (2000). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press.
- Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, 30, 958–968.
- Dennis, M., Purvis, K., Barnes, M. A., Wilkinson, M., & Winner, E. (2001). Understanding of literal truth, ironic criticism, and deceptive praise following childhood head injury. *Brain and Language*, 78, 1–16.
- Ferstl, E. C., & Von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research*, 11, 325–340.

- Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *NeuroImage*, *17*, 1599–1612.
- Ferstl, E. C., Rinck, M., & von Cramon, D. Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *17*, 724–729.
- Eviatar, Z., & Just, M. A. (2006). Brain correlates of discourse processing: An fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia*, *44*, 2348–2359.
- Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. *Cognition*, *57*, 109–128.
- Frank, S. L., Koppen, M., Noordman, L. G. M., & Wietske, V. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, *27*, 875–910.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends in Cognitive Neuroscience*, *7*, 77–83.
- Gardner, H., Boller, F., Moreines, J., & Butters, N. (1973). Retrieving information from Korsakoff patients: Effects of categorical cues and reference to the task. *Cortex*, *9*, 165–175.
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory & Language*, *37*, 331–355.
- Gernsbacher, M. A. (1990) *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology*, *54*, 91–114.
- Gernsbacher, M. A., & Robertson, R. R. W. (1999). The role of suppression in figurative language comprehension. *Journal of Pragmatics*, *31*, 1619–1630.
- Gerrig, R., & Healy, A., (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 667–675.
- Gibbs, R. W., Jr. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 56–66.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *7*, 183–206.
- Golden, R. M., & Rumelhart, D. E. (1993). A parallel distributed processing model of story comprehension and recall. *Discourse Processes*, *16*, 203–237.
- Goldman, S. R., & Varma, S. (1995). CAPing the construction-integration model of discourse comprehension. In: C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337–358). Hillsdale, NJ: Erlbaum.



- Grafman J. (1995). Similarities and Distinctions among current models of prefrontal cortical functions. *Annals of the New York Academy of Sciences*, 769, 337–368.
- Grafman J., Salazar, A. M., Vance, S. C., Weingartner, H., & Amin, D. (1987). Immediate memory for story discourse in Vietnam veterans with penetrating brain wounds. *Journal of Clinical and Experimental Neuropsychology*, 9, 23–34.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108.
- Hesling, I., Clément, S., Bordessoules, M., & Allard, M. (2005). Cerebral mechanisms of prosodic integration: Evidence from connected speech. *NeuroImage*, 24, 937–947.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114–116.
- Just, M. A., Newman, S. D., Keller, T. A., McEleney, A., & Carpenter, P. A. (2004). Imagery in sentence comprehension: An fMRI study. *NeuroImage*, 21, 112–124.
- Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 115–126.
- Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, 11, 223–237.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Koivisto, M., (1998). Categorical priming in the cerebral hemispheres: Automatic in the left hemisphere, postlexical in the right hemisphere? *Neuropsychologia*, 36, 661–668.
- Krause, B. J., Schmidt, D., Mottaghy, F. M., Taylor, J., Halsband, U., Herzog, H., Tellmann, L., and Muller-Gartner, H. (1999). Episodic retrieval activates the precuneus irrespective of the imagery content of word pair associates: A PET-study. *Brain*, 122, 255–263.
- Langston, M., & Trabasso, T. (1999). Identifying causal connections and modeling integration of narrative discourse. In H. van Oostendorp, & S. Goldman (Eds), *The construction of mental representation during reading*. Mahwah, NJ: Erlbaum.
- Long, D. L., & Baynes, K. (2002). Discourse representation in the two cerebral hemispheres. *Journal of Cognitive Neuroscience*, 14(2), 228–242.
- Maguire, E. A., Frith, C. D., & Morris, R. G. M. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain*, 122, 1839–1850.

- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.
- Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology Special Issue: The organization of conceptual knowledge in the brain: Neuropsychological and neuroimaging perspectives*, *20*, 575–587.
- Mashal, N., Faust, M., & Hendler, T. (2005). The role of the right hemisphere in processing non-salient metaphorical meanings: Application of Principal Components Analysis to fMRI data. *Neuropsychologia*, *43*, 2084–2100.
- Mason, R. A., Eviatar, Z., & Just, M. A. (2005). *Brain activation associated with comprehending novel and frozen metaphors*. Unpublished manuscript.
- Mason, R. A., & Just, M. A. (2004). How the brain processes causal inferences in text: A theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychological Science*, *15*, 1–7.
- Mason, R. A., Just, M. A., Keller, T. A., & Carpenter, P. A. (2003). Ambiguity in the brain: What brain imaging reveals about the processing of syntactically ambiguous sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1319–1338.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467–479.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 82–91.
- McKoon, G., & Ratcliff, R. (1989). Semantic associations and elaborative inferences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 326–338.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.
- Mellet, E., Bricogne, S., Crivello, F., Mazoyer, B., Denis, M., & Tzourio-Mazoyer, N. (2002). Neural basis of mental scanning of a topographic representation built from a text. *Cerebral Cortex*, *12*, 1322–1330.
- Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, *16*, 696–703.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131–157.
- Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453–465.
- Nakagawa, A. (1991). Role of the anterior and posterior attention networks in hemispheric asymmetries during lexical decisions. *Journal of Cognitive Neuroscience*, *3*, 315–321.

- Newman, S. D., Twieg, D. B., & Carpenter, P. A. (2001). Baseline conditions and subtractive logic in neuroimaging. *Human Brain Mapping, 14*, 228–235.
- Nichelli, P., Grafman, J., Pietrini, P., Clark, K., Lee, K. Y., & Miletich, R. (1995). Where the brain appreciates the moral of a story. *NeuroReport, 6*, 2309–2313.
- O'Brien, E. J., Cook, A. E., & Peracchi, K. A. (2004). Updating Situation Models: Reply to Zwaan and Madden (2004). *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 289–291.
- Papagno, C. (2001). Comprehension of metaphors and idioms in patients with Alzheimer's disease: A longitudinal study. *Brain, 124*, 1450–1460.
- Partiot, A., Grafman, J., Sadato, N., Fitman, S., & Wild, K. (1996). Brain activation during script event processing. *NeuroReport, 7*, 761–766.
- Peleg, O., Giora, R., & Fein, O. (2001). Salience and Context Effects: Two are better than one. *Metaphor and Symbol, 16*, 173–192.
- Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: Activation interacts with task demands. *NeuroImage, 17*, 401–410.
- Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. J. (2004). Neural correlates of metaphor processing. *Cognitive Brain Research, 20*, 395–402.
- Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence-picture verification. *Cognitive Psychology, 40*, 261–295.
- Reichle, E. D. & Mason, R. A. (2006). The neural signatures of causal inferences: A preliminary computational account of brain-imaging and behavioral data. In: F. Schmalhofer, & C. A. Perfetti (Eds), *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ: Erlbaum.
- Robertson, D. A., Gernsbacher, M. A., Guidotti, S. J., Robertson, R. R. W., Irwin, W., Mock, B. J., & Campana, M. E. (2000). Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. *Psychological Science, 11*, 255–260.
- Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33*, 105–132.
- Schmalhofer, F. (2003). Construction and Integration Processes in Inference Generation, Storage and Retrieval. *International Hanse-Conference on Higher Level Language Processes in the Brain: Inference and Comprehension Processes*. Hanse-Advanced Study Institute, Delmenhorst, Germany.
- Singer, M. (1994). Discourse Inference Processes. In: M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 479–515). San Diego, CA: Academic Press.
- St. George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain, 122*, 1317–1325.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring “how” from “where”. *Neuropsychologia Special Issue: Functional neuroimaging of memory*, 41, 280–292.

Tomitch, L. M. B., Just, M. A., & Newman, S. D. (2004). Main Idea Identification: A Functional Imaging Study of a Complex Language Comprehension Process. In: L. M. B. Tomitch, & C. Rodrigues (Eds), *Ensaio sobre a linguagem e o cérebro humano: Contribuições multidisciplinares* Porto Algre: Artmed (pp. 167–175).

Tompkins, C. A., & Mateer, C. A. (1985). Right hemisphere appreciation of prosodic and linguistic indications of implicit attitude. *Brain and Language*, 24, 185–203.

Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the Landscape Model: Modeling inferential processes and memory representations of text comprehension. *Behavioral Research Methods, Instruments & Computers*, 37, 277–286.

van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In: M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 539–588). San Diego, CA: Academic Press.

van den Broek, P., Ridsen, K., Fletcher, C. R., & Thurlow, R. (1996). A ‘landscape’ view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In: B. K. Britton, & A. C. Graesser (Eds), *Models of understanding text* (pp. 165–187). Hillsdale, NJ: Erlbaum.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage*, 25, 1002–1015.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185.

Zwaan, R. A., & Singer, M. (2003). Text comprehension. In: A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds), *Handbook of discourse processes* (pp. 83–122). Mahwah, NJ: Lawrence Erlbaum.

Zysset, S., Huber, O., Ferstl, E., & von Cramon, D.Y., (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *NeuroImage*, 15, 983–991.

This page intentionally left blank

## Chapter 20

### Comprehension Ability in Mature Readers

*Debra L. Long, Clinton L. Johns, and Phillip E. Morris*

#### 1. INTRODUCTION

Readers construct detailed mental representations of texts with relative ease; this tends to obscure the fact that reading is a complex, intellectual skill, requiring the coordination of multiple component processes. At the word level, processes are necessary to encode the printed word, access its sound-based representation, and retrieve its meaning from memory. At the sentence level, processes are devoted to the formation of structures that specify the syntactic and conceptual relations within and across phrases; these processes are involved in understanding “who did what to whom” in a sentence. This results in a representation that some theories call propositional–abstract units of meaning that encode the explicit ideas in a sentence. At the discourse level, the explicit ideas in a text are integrated across sentences and with contextually relevant semantic and pragmatic knowledge. This results in a mental representation that researchers call a discourse or situation model. It reflects features of the real or imaginary world that the text describes. In order to construct a discourse model, readers must engage in active inferential processing to interpret and restructure text information in light of their prior understanding of the relevant knowledge domain.

The ability to learn and execute the component processes of reading varies widely. The challenges faced by beginning readers are detailed elsewhere in this volume (see Torgesen & Wagner, this volume); therefore, we focus here on the comprehension performance of adult readers. These readers differ on a broad range of component reading abilities. At the word and sentence levels, poor comprehenders, relative to good ones, have slower and less efficient word-identification skills and greater difficulty processing low-frequency syntactic structures (Bell & Perfetti, 1994; Curtis, 1980; Frederiksen, 1981; Jackson & McClelland, 1979; Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; Palmer, MacLeod, Hunt, & Davidson, 1985; Pearlmutter & MacDonald, 1995). The greatest variation, however, is seen in readers’ ability to execute high-level interpretive processes involved in constructing a coherent discourse model. Poor reading comprehension is frequently associated with a systematic

failure to make appropriate inferences during reading. Poor comprehenders have difficulty making inferences to integrate ideas in a text, to answer questions, and to identify main ideas and themes (Garnham, Oakhill, & Johnson-Laird, 1982; Long & Golding, 1993; Long, Oppy, & Seely, 1994, 1997).

Reading comprehension involves language specific processes as well as domain-general cognitive abilities—sensation, perception, attention, memory, and reasoning. Variation in any of these abilities potentially underlies individual differences in comprehension performance. Thus, one important question concerns the extent to which variation in some ability is central to individual differences in comprehension performance and the extent to which variation in the ability is derivative. For example, researchers have found that poor comprehenders have more difficulty parsing complex syntactic structures than do good comprehenders. Does variation in parsing ability explain individual differences in comprehension, or is this variation secondary to some other ability, such as fast and accurate word recognition?

In the following sections, we review five reader characteristics that are associated with comprehension ability in mature readers. We used two criteria in selecting these particular characteristics from all of those that are potentially involved in comprehension. First, we selected characteristics that have the strongest correlations with comprehension performance. Second, we chose characteristics that play a central role in different theories of comprehension skill. The characteristics that we selected are: word-level ability, working memory (WM) capacity, suppression ability, print exposure, and background knowledge.

We review each characteristic in a separate section; however, it is important to note that they are not independent. They are correlated with each other as strongly as they are correlated with measures of comprehension performance. The theories of comprehension skill that we review explain these correlations by means of different causal mechanisms. We include figures in each section to illustrate how these theories partition variance among the five reader characteristics in different ways. These figures do not represent all conceivable relations among the characteristics; they represent only those that are emphasized according to different explanations of individual differences in comprehension ability.

## **2. INDIVIDUAL DIFFERENCES IN WORD-LEVEL ABILITIES**

The most straightforward explanation of poor comprehenders' failure to construct coherent discourse models is that they result from deficits in basic linguistic abilities—in particular, word-identification skill (Perfetti & Lesgold, 1977; Perfetti, 1985, 1989, 1994; Vogel, 1975). Individual differences in the component processes of word-recognition are well documented (Bell & Perfetti, 1994; Curtis, 1980; Frederiksen, 1981; Jackson & McClelland, 1979; Palmer et al., 1985).

In beginning readers, word-identification skill is strongly related to measures of phonological awareness—explicit knowledge about the phonological structure of the language (Perfetti, 1991; Perfetti, Beck, Bell, & Hughes, 1987). Mature readers, however, show little variability on phonological awareness tasks. Nonetheless, they vary widely in the speed and accuracy with which they can map a letter string onto a sound-based representation (Bell & Perfetti, 1994; Olson, Kliegl, Davidson, & Foltz, 1985; Perfetti, 1989). The three tasks that are most commonly used to assess adult readers' word-level abilities are naming (i.e., rapid pronunciation of words and pseudowords), phonological decision (i.e., a decision about which of two pseudowords—e.g., brane and blane—would be a real word if pronounced aloud), and orthographic decision (i.e., a decision about which of two letter strings—e.g., brane and brain—is correctly spelled).

The view that comprehension problems at the discourse level have a lexical basis is the primary assumption of Perfetti's (1985) *verbal efficiency theory*. Perfetti argues that comprehension depends on the rapid retrieval of high-quality lexical codes during word recognition. A lexical code is high quality to the extent that it is specific—it has a fully specified orthographic representation—and redundant—it has a representation that can be retrieved from both spoken language and from orthographic-to-phonological mapping.

*Verbal efficiency theory* suggests two ways that deficits in word-identification skill can influence comprehension performance. First, comprehension processes that depend on high quality, lexical representations, such as syntactic analysis, will be negatively affected if readers retrieve low quality lexical codes. Second, slow retrieval of lexical codes can compromise higher-level interpretive processes by consuming WM resources that would otherwise be devoted to these processes.

Figure 1 depicts the five reader characteristics that are discussed in this chapter. *Verbal efficiency theory* emphasizes three of these: word-level ability, WM capacity, and print exposure. Figure 1 shows a direct effect of word-level ability on word-level processes and on print exposure. Readers with fast and efficient word-identification skills learn more about words than do readers with poorer skills and this is related to their enthusiasm for reading. Individuals who read often will improve their word-level ability, further increasing their desire to read. Figure 1 also depicts indirect effects of word-level ability on sentence-level and discourse-level processes via a direct effect on WM capacity. Slow and inefficient word-level processes will consume WM resources that would otherwise be devoted to high-level interpretive processes.

### 2.1. Are Comprehension Problems Secondary to Poor Quality Representations?

According to *verbal efficiency theory*, if readers retrieve low-quality lexical representations, then “down-stream” processes that depend on high-quality codes will be affected. Do adult readers fail to execute high-level interpretive processes because they fail to construct adequate representations at lower-levels of processing? Answering this question is difficult because researchers have yet to identify the minimum level of word-identification



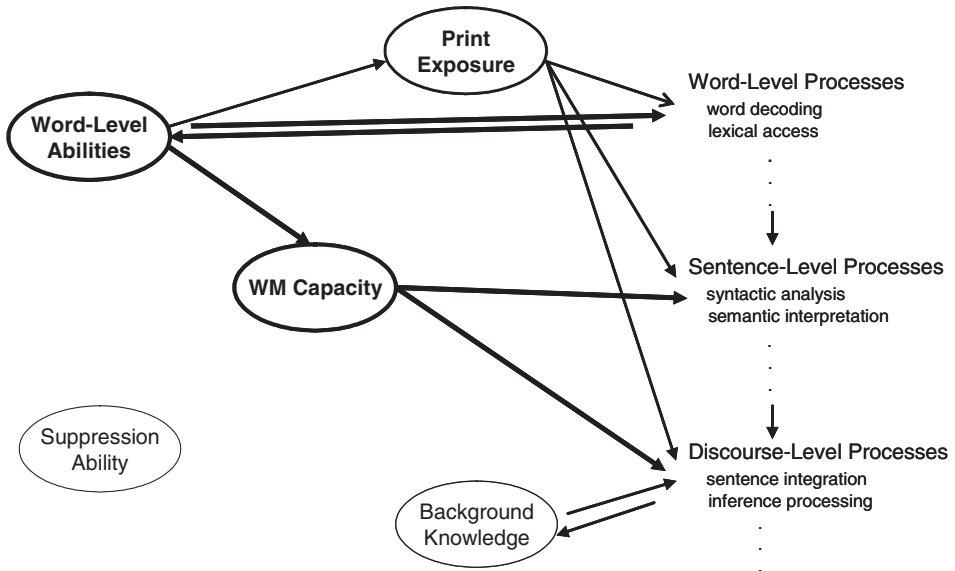


Figure 1. Hypothesized relations among the five reader characteristics according to verbal efficiency theory. The figure is gray-scaled to indicate strength of relations. Dark print corresponds to stronger relations than does light print.

skill that is necessary for constructing an accurate sentence-level representation. The assessment of word-level ability is always a relative one; if poor comprehenders perform worse than good ones on some measure of word-level processing (e.g., word-naming speed), then researchers often assume that these readers lack the ability to construct accurate sentence-level representations. This assumption, however, may not be warranted. Poor comprehenders may have less accurate word-identification processes than good comprehenders, but these processes may be adequate to the task of constructing an accurate sentence-level representation.

Long and her colleagues have conducted a series of studies in which they examined the accuracy of poor comprehenders' sentence-level representations as well as their ability to execute processes at the discourse level (Long et al., 1994, 1997). In one study (Long et al., 1994), they contrasted good and poor comprehenders' ability to execute a process necessary to represent the meaning of a sentence (i.e., to select the context-appropriate sense of an ambiguous word) with their ability to make an inference related to the sentence topic.

Long et al. (1994) argued that sense selection depends on reasonably accurate representations of sentence contexts. If poor comprehenders can quickly and accurately select context-appropriate senses of ambiguous words, their failure to make topic-related inferences should be attributed factors other than poor quality sentence representations. Long et al. (1994) had participants read short-passages containing a sentence that ended

with a homograph prime (e.g., The townspeople were amazed to find that all the buildings had collapsed except the *mint*. Obviously it had been built to withstand natural disasters) and respond to lexical-decision targets. The targets were (a) context-appropriate associates of the homographs (e.g., money), (b) context-inappropriate associates of the homographs (e.g., candy), (c) words related to the topics of the sentences (e.g., earthquake), (d) words unrelated to the topics of the sentences (e.g., breath), and (e) nonwords. The time course of sense selection and inference processing was examined by presenting the targets at different stimulus onset asynchronies (SOAs).

Long et al. (1994) found that both good and poor comprehenders responded faster to appropriate than to inappropriate associates of the homographs within 300ms of processing. In contrast, they found that only good comprehenders responded faster to appropriate than to inappropriate topic words. They also found that poor comprehenders' failure to make topic-related inferences was not due to lack of knowledge about the sentence topics; when the two groups were asked to provide single-word descriptions of the topics of the sentences, they showed no differences.

In a second set of studies, Long et al. (1997) provided additional information about poor comprehenders' ability to construct accurate sentence-level and discourse-level representations. They used an item-priming-in-recognition procedure to examine the structure readers' memory representations. The logic of the paradigm is that activation of a concept in memory facilitates recognition of other concepts to which it is linked (Ratcliff & McKoon, 1978).

Long et al. (1997) used this procedure to investigate readers' structural (propositional) representations of sentences, their representation of context-appropriate senses of ambiguous words, and their representation of topic-related information. Participants received a series of study-test trials. Each trial consisted of a set of passages similar to those used by Long et al. (1994) (e.g., The townspeople were amazed to find that all the buildings had collapsed except the *mint*. Obviously, the architect had foreseen the danger because the structure withstood the natural disaster). The passages were followed by a list of single-word recognition items. Embedded in each test list were three types of priming pairs: (a) propositional-priming pairs consisted of a target that was preceded by a prime from the same proposition or by a prime from a different proposition in the same sentence (e.g., *disaster-structure* versus *danger-structure*), (b) associate-priming pairs consisted of a target that was either the appropriate or the inappropriate associate of a homograph in the sentence and was preceded by a prime from the sentence containing the homograph (e.g., *buildings-money* versus *buildings-candy*), and (c) topic-priming pairs consisted of a target that was either the topic of a passage or was an unrelated word and was preceded by a prime from the final sentence of the passage (e.g., *architect-earthquake* versus *breath-earthquake*). (Note: targets in the associate-priming and topic-priming pairs did not appear in the passages; thus, the correct answer to these items was "no.")

Long et al. (1997) found a strong propositional-priming effect. Both good and poor comprehenders recognized targets faster when they were preceded by primes from the

same than different propositions. Both groups also showed reliable priming for targets in associate-priming pairs. They had difficulty rejecting appropriate associates of ambiguous words because these items resonated with information in their memory representations. Only good comprehenders, however, showed topic-priming effects—difficulty rejecting topic words that were related to the passages.

Using a similar method, Long et al. (1997) examined the extent to which good and poor comprehenders integrated ideas from different parts of a text. They contrasted the groups' ability to make connections among idea units that were relatively distant in the surface structure of a story and between stories that shared the same theme. Participants received study-test trials in which they read pairs of stories that were either thematically related or were unrelated. They then made recognition judgments to a list of sentences. Target items in the test list were sentences from one of the stories that described the outcome of a character's attempt to achieve a goal. The targets were paired with different types of primes: (1) in story-priming pairs, the outcome was preceded by the character's goal from the same story or a character's goal from the other story in the pair; (2) in thematic-priming pairs, the outcome was preceded by the outcome of a story that shared the same theme or the outcome of a story that had an unrelated theme.

Good and poor comprehenders showed differences in their ability to integrate information from different parts of the story and their ability to elaborate their representations with topic-related information. Good, but not poor, comprehenders made connections between a character's goal and a later description of the goal outcome. In addition, good comprehenders, but not poor ones, represented connections in memory between stories that shared the same theme.

One explanation for why poor comprehenders fail to integrate information from different parts of a text is that they fail to activate prior text ideas when they read new incoming information. Long and Chong (2001) tested this idea using an inconsistency paradigm. First, they demonstrated that poor comprehenders could detect an inconsistency when relevant information was adjacent in the text, but not when it was separated by intervening sentences. Good and poor comprehenders received a set of passages in which a target action (e.g., Ken enrolled in boxing classes) was either consistent or inconsistent with a description of the character presented earlier in the passage (e.g., Ken loved/hated contact sports). In one condition, the target and character description were separated by a single statement (the local condition). In another condition, the target action and character description were separated by several sentences (the global condition). Both good and poor comprehenders showed an inconsistency effect in the local condition; they were slow to read the target when it was inconsistent with the character description presented earlier in the text. Only good comprehenders, however, showed the inconsistency effect in the global condition, when the target action and character description were separated by several intervening sentences.

In their second experiment, Long and Chong (2001) used a probe-verification paradigm to determine whether poor comprehenders failed to detect the inconsistency in the global

condition because they failed to reactivate the character description when they read the inconsistent action. Probe sentences asking about the character description were presented (a) immediately after the character description, (b) after several intervening filler sentences, but before the target, and (c) after the target. They found that both groups exhibited the same pattern of results. They responded faster to probes presented after the character description and after the target than to probes presented after the filler sentences. This suggests that the character description was replaced in WM by information from the filler section and then reactivated when readers comprehended the target. Thus, poor comprehenders had a representation of the character description that was reactivated when they read the target action; however, activation of this information had no effect on their comprehension of the target.

In summary, the studies described above suggest that poor comprehenders have word-level and sentence-level processes that are accurate enough to encode structural relations among concepts in a sentence (i.e., propositional relations), to use content in selecting the appropriate sense of an ambiguous word, and to support the reactivation of prior text information. Thus, poor comprehenders appear to construct quality sentence representations; why, then, do they fail to execute high-level interpretive processes to construct a coherent discourse model? We address this question in the next section.

## 2.2. Are Comprehension Problems Due to Slow Word-Level Processing?

A central claim of *verbal efficiency theory* is that slow word-level processes can consume resources that would otherwise be devoted to higher-level interpretive ones. Only a few studies have been conducted to examine the unique influence of rapid word-identification on mature readers' comprehension ability separate from other skills that affect comprehension. In one of these, Bell and Perfetti (1994) investigated the relative contributions of general language ability and word-decoding skill. Participants read a set of texts about different topics: science, history, and fiction. They then received a multiple-choice comprehension test for each passage. Bell and Perfetti conducted a series of regression analyses; the critical predictors were listening comprehension, vocabulary, reading speed, and pseudoword decoding. General language ability–listening comprehension, vocabulary, and reading speed—were significant predictors of comprehension for all text types. Pseudoword decoding emerged as a significant predictor for the difficult science texts. Their results suggest that word decoding made a contribution to comprehension that was independent of general language ability, but only when readers comprehended the more difficult science texts.

Long, Prat, Blozis, Widaman, and Traxler (2006) provided somewhat stronger evidence that word-decoding ability plays a unique role in adult comprehension ability using a technique called multilevel modeling. They assessed participants' performance on several information processing and language tasks. These tasks included (a) phonological decision, (b) orthographic decision, (c) speeded naming, (d) WM span, (e) print exposure, and (f) The Nelson–Denny Vocabulary and Comprehension Test. Participants also read several full-length, narrative texts and their sentence reading times (RTs) were recorded.

The initial step in constructing the multilevel model was to analyze the level-1 data (i.e., sentence RTs for each participant). Three text characteristics were used in the analysis: (a) number of function words—a crude index of grammatical complexity, (b) number of new argument nouns—a measure that is sensitive to the introduction of new entities into the discourse model, and (c) number of repeated argument nouns—an index of anaphoric reference. The second step was to identify the latent variable structure for the level-2 variables, measured at the participant level. Performance on the individual difference tests was factor analyzed. The analysis yielded five factors: (a) decoding speed—RTs on the naming, phonological, and orthographic decision tasks; (b) decoding accuracy—accuracy on the naming, phonological and orthographic decision tasks; (c) print exposure—performance on Author and Magazine Recognition Tests; (d) verbal ability—vocabulary and comprehension subsections of the Nelson-Denny reading test; and (e) WM capacity—performance on operation span and reading span tests.

The final step in the analysis was to regress the level-1 coefficients on the five individual-differences factors (level-2 latent variables). The model showed that individual differences in the coefficient relating RTs to the number of function words in a sentence was influenced by verbal ability alone, suggesting that readers who were high in overall verbal ability were less affected by grammatical complexity than readers who were low in verbal ability. The coefficient associated with number of new argument nouns was predicted by decoding speed alone, suggesting that the ability to encode new entities into the discourse model primarily depended on the ability to decode words rapidly. Finally, the coefficient relating to the number of repeated argument nouns was predicted by decoding speed and print exposure, suggesting that the ability to process anaphoric references depended on both rapid word decoding and reading practice.

Two findings in this study are notable. First, word-decoding speed was a unique predictor of sentence processing, whereas word-decoding accuracy was unrelated to performance. This is consistent with the evidence that we reviewed in the previous section suggesting that poor comprehenders have word-identification skills that are adequate for constructing quality sentence representations. Second, WM capacity was not a reliable predictor of any level-1 coefficient. One explanation for this is that sentence RT may be insensitive to the effects of WM capacity. Long et al. (2006) tested this explanation by conducting an analysis in which they examined the influence of the WM factor alone. They found that the coefficient associated with the number of function words in a sentence was predicted by WM capacity, but only when the other individual-difference factors were eliminated. Thus, WM capacity failed to predict the level-1 coefficients because it shares variance with the other factors. This finding is relevant to current theoretical debates about the nature of WM capacity and how it relates to reading comprehension, issues that we discuss in the next section.

In summary, *verbal efficiency theory* claims that slow and inaccurate word-level ability is associated with reading comprehension in adult readers as it is in children. The research that we have reviewed here suggests that adult readers have word-level processes that are accurate enough for the construction of reasonably good sentence-level representations.

Slow word-level processing, however, appears to be predictive of reading comprehension, independent of other factors such as general verbal ability and WM capacity.

### 3. INDIVIDUAL DIFFERENCES IN WORKING MEMORY

Reading comprehension, like most complex cognitive tasks, involves multiple processing steps. Successful performance requires the ready availability of task goals, task-relevant information, and the intermediate results of cognitive operations. WM is the theoretical construct used to refer to the system that is responsible for maintaining such information (Baddeley & Hitch, 1974; Cowan, 1988, 1995; Klapp, Marshburn, & Lester, 1983).

The hallmark characteristic of WM is its limited capacity. Although the existence of capacity limitations is uncontroversial, the factors responsible vary from theory to theory. These factors include constraints on the amount of activation available to the WM system (Just & Carpenter, 1992; Engle, Cantor, & Carullo, 1992; Lovett, Reder, & Lebiere, 1999), similarity-based interference (Schneider, 1999; Young & Lewis, 1999), processing speed (Kieras, Meyer, Mueller, & Seymour, 1999; Salthouse, 1996), lack of skill or knowledge for efficient encoding and retrieval (Ericsson & Kintsch, 1995; Ericsson & Delaney, 1999), and the ability to inhibit irrelevant information (Stoltzfus, Hasker, & Zacks, 1996; Rosen & Engle, 1998).

Two measures are commonly used to assess WM capacity: the reading-span task and the operation-span task. In the reading-span task, participants are asked to read aloud a set of unrelated sentences, presented one at a time, and to recall the final word of each sentence once the entire set has been presented. The operation span task is similar except that participants perform simple arithmetic problems rather than read aloud. In both tasks, span is the largest set of words that the participant can recall.

The role of WM in reading comprehension has been important in theoretical debates about the nature of WM and its limitations. In the next two sections, we describe theories that make different claims about how WM and comprehension are related. One class of theories attributes variation in comprehension to limitations in capacity; the other class attributes variation in both comprehension and WM to individual differences in skill and knowledge. (We describe a third type of WM model in the section on suppression ability; it attributes variation in performance on complex tasks, such as reading comprehension, to individual differences in controlled attention.)

#### 3.1. Limitations Due to Capacity Constraints

In this section, we review two theories that attribute variation in reading comprehension to individual differences in WM capacity: the *capacity theory of comprehension* and the *separate-sentence-interpretation-resource (SSIR) theory*. These theories have important differences, but both claim that high-level interpretive processes involving the

integration of ideas across sentences and the use of sentence meaning to make inferences are constrained by capacity limitations.

### 3.1.1. *The capacity theory of comprehension*

Just and Carpenter (1992) present one view of WM limitations in their *capacity theory of comprehension*. According to the theory, the storage and processing functions necessary for language are fueled by activation, a commodity that maintains knowledge elements in memory and supports computation. Activation is shared among storage and processing functions such that activation-consuming processes limit the amount of activation available to support storage and vice versa. The *capacity theory* attributes individual differences in reading comprehension to variation in capacity, the total amount of activation available to the system.

Support for *capacity theory* has been found in studies of the relation between language comprehension and performance on WM tasks—in particular, the reading span task devised by Daneman and Carpenter (1980). Reading span correlates with a number of verbal measures, including verbal SAT ( $r=0.5$  to  $r=0.6$ ) and the ability to answer questions about explicit information in a text ( $r=0.7$  to  $r=0.9$ ; Daneman & Carpenter, 1980; Masson & Miller, 1983). Reading span interacts with text complexity to influence RT. Reading time differences between low-span and high-span individuals are small for easy texts, but large for difficult ones (Just & Carpenter, 1992).

Several studies have reported a correlation between WM performance on the reading-span task and the speed and accuracy of syntactic processing (Just & Carpenter, 1992; King & Just, 1991; MacDonald et al., 1992; Pearlmutter & MacDonald, 1995). One of the first studies to document this relation examined how high-span and low-span readers process garden-path sentences. Just and Carpenter (1992) presented readers with sentences containing reduced relative (RR) clauses, such as The defendant examined by the lawyer shocked the jury. When the initial noun phrase was a plausible agent of the main verb (MV) interpretation (e.g., The defendant examined...), both high-span and low-span readers exhibited long gaze durations on the disambiguating information in the sentence (i.e., the word by). The performance of the two groups differed, however, when the initial noun phrase was an implausible agent of the MV interpretation (e.g., The evidence examined ...). The plausibility manipulation decreased processing time for high-span readers, but did not affect the performance of low-span readers. Just and Carpenter argued that only high-span readers had the WM capacity necessary to use information about plausibility during the comprehension process.

Other studies have reported similar results. King and Just (1991) found that low-span readers were slower and less accurate than high-span readers in processing difficult, object-relative sentences, such as The senator that the reporter attacked admitted the error. MacDonald et al. (1992) and Pearlmutter and MacDonald (1995) also found a relation between reading span and syntactic ambiguity resolution, although high-span, rather than low-span, readers showed slower processing of syntactically ambiguous relative to unambiguous sentences.

Similar results have been found in studies examining the relation between WM and discourse-level processes. For example, high-span, compared to low-span, readers are more accurate in finding the antecedent of a pronoun when the pronoun and its antecedent are separated by intervening sentences (Daneman & Carpenter, 1980). High-span readers are also more likely than low-span readers to show faster recognition of sentences that are thematically related (Cantor & Engle, 1993).

Figure 2 depicts hypothesized relations among WM capacity and other reader characteristics according to *capacity theory*. *Capacity theory* emphasizes the role of WM in comprehension. Capacity limitations should have direct effects on all comprehension processes, but these effects should be particularly strong for resource-consuming, sentence-level and discourse-level processes.

3.1.2. *Separate-sentence-interpretation-resource (SSIR) theory*

Waters and Caplan (1996) argue for a model of WM called the *SSIR theory*. Their model of WM is highly modularized. Part of the WM system is specialized for analyzing syntactic structure and using it to determine sentence meaning. This part of the system is dedicated to sentence interpretation and individuals do not differ with respect to its capacity. Another part of the system is devoted to activities that involve conscious controlled processing, activities that Waters and Caplan call “post-interpretive.” These activities include making inferences to integrate ideas across sentences, using world knowledge in the interpretation of a text, remembering sentence content, and planning

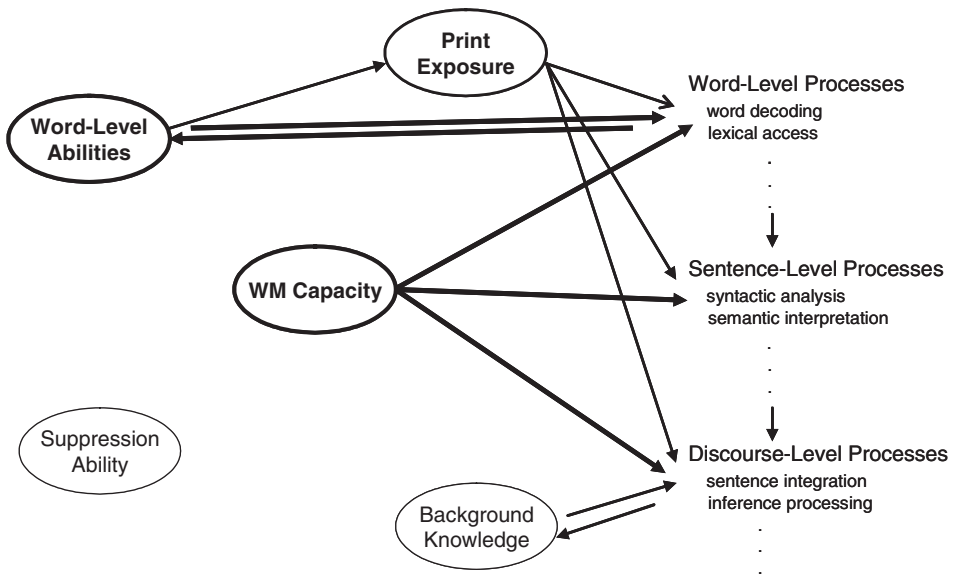


Figure 2. Hypothesized relations among the five reader characteristics according to the capacity theory of comprehension.



actions based on the meaning of sentences and texts. This part of the WM system shows variation across individuals and is the one that is tapped when individuals perform complex span tasks, such as reading span and operation span.

The *SSIR theory* has its foundation in neuropsychological data concerning the ability of patients to understand sentences containing complex syntactic structures (Caplan & Waters, 1995; Martin, 1995; Waters, Caplan, & Hildebrandt, 1991). Many patients have extremely limited verbal memory spans, but are able to use a variety of complex syntactic structures in deriving the meaning of sentences (see for a review Waters & Caplan, 1996). Thus, Waters and Caplan (1996) argue that WM must involve at least two systems. Brain damage can impair the system that underlies performance on span tasks, leaving intact the system that underlies sentence interpretation.

In addition to neuropsychological studies, Waters and Caplan (1996) have conducted studies of the syntactic processing abilities of normal adults and found no evidence for a correlation between reading span and syntactic parsing. For example, they examined the processing of sentences containing a temporary syntactic ambiguity, similar to those used by MacDonald et al. (1992). Although Waters and Caplan found that ambiguous sentences were more difficult to comprehend than were unambiguous ones, they found no processing-time differences as a function of reading span (see also, Clifton et al., 2003; Traxler, Williams, Blozis, & Morris, 2005). Similarly, Caplan and Waters (1999) found no differences between high and low-span readers in the processing of sentences with object-relative clauses like those used by King and Just (1991). These findings cast some doubt on the role of WM capacity in explaining individual differences at the sentence level. Nonetheless, the *SSIR theory* is similar to the *capacity theory* in attributing individual differences at the discourse level to a limited capacity WM system.

Figure 3 depicts hypothesized relations among WM capacity and other reader characteristics according to *SSIR theory*. In this figure, WM is partitioned into two pools of resources. One is devoted exclusively to word and sentence-level processing and shows no individual variation (denoted by a square in the figure). Individuals do differ, however, in the capacity of the WM system that is devoted to post-interpretive processing. Thus, the *SSIR theory*, like the *capacity theory*, predicts a strong relation between WM capacity and discourse-level processing.

### 3.2. Limitations Due to Poor Word-Level Ability and Insufficient Experience

A second view of WM limitations emphasizes the role of skill and practice in comprehension, rather than capacity, per se. This view is represented in two models of WM: a *connectionist-based account* proposed by MacDonald and Christiansen (2002) and the *long-term working memory (LTWM) model* proposed by Ericsson and Kintsch (1995). These two accounts differ in many respects, but both emphasize the importance of skill and experience in the relation between capacity and comprehension performance.

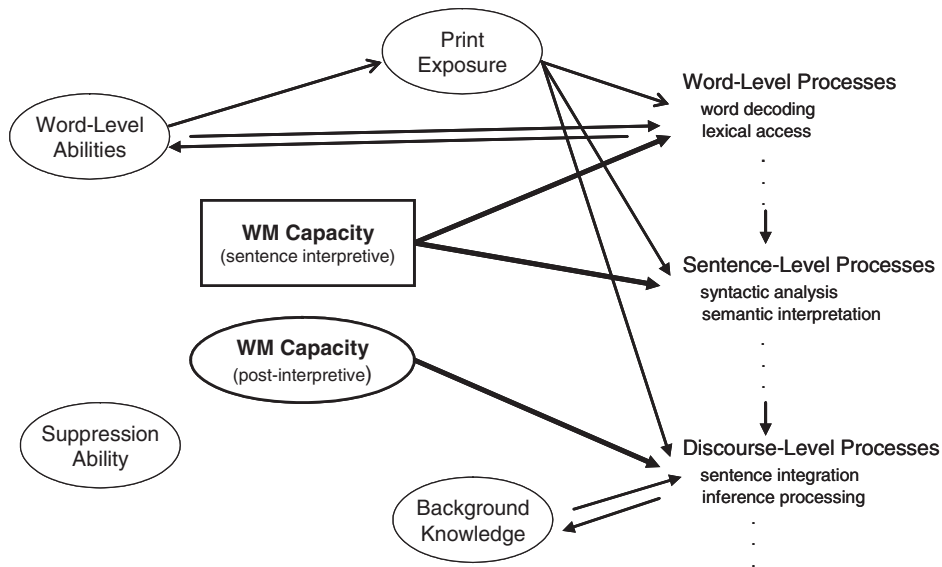


Figure 3. Hypothesized relations among the five reader characteristics according to the SSIR theory.

3.2.1. *The connectionist-based account*

The *connectionist-based account* of variation in WM is based on connectionist approaches to language processing. In a connectionist network, the capacity of the system arises from its architecture (i.e., the number of processing units, how activation passes through weights, etc.) and the network’s experience (i.e., how often it has processed similar input in the past). In this view, capacity is not a separate pool of resources; it is a property of the processing network itself.

In the *connectionist-based account*, individual differences in performance on WM tasks arise from variation in two factors. First, individuals can vary with respect to basic sensory/perceptual abilities—primarily the ability to represent phonological information accurately. Second, individuals can vary in reading experience. The *connectionist-based account* emphasizes this second factor. Variation in practice can lead to individual differences that appear qualitative, such as differences in the nature of Frequency  $\times$  Regularity interactions. Consider, for example, the Frequency  $\times$  Regularity interaction that is found in word recognition. High-frequency words are recognized faster than low-frequency words and the effect is larger for words with an irregular than a regular orthography. Moreover, frequency and regularity interact with skill (Seidenberg, 1985). Good comprehenders exhibit regularity differences only in the low-frequency range. In contrast, poor comprehenders exhibit regularity differences for all words except those in the high-frequency range. Seidenberg (1985) attributes the Skill  $\times$  Frequency  $\times$  Regularity interaction to variation in

reading experience. Good comprehenders, who read often, encounter irregular words more frequently than do poor comprehenders, who read less. Thus, good comprehenders have a broad frequency range of irregular words for which they can quickly compute the appropriate pronunciation. Poor comprehenders, in contrast, have sufficient experience only in computing the appropriate pronunciation of high-frequency irregular words. Thus, they show irregularity effects at all other word frequencies.

A similar explanation can be applied to the relation between WM capacity and sentence processing. Relative clauses are a low-frequency syntactic structure. Low-span readers are likely to encounter these structures infrequently. In addition, they should have particular difficulty understanding object-relative clauses because these clauses have an irregular word order (i.e., Noun-Noun-Verb as opposed to the canonical Noun-Verb-Noun).

The influence of frequency on ambiguity resolution was recently demonstrated in a study by Long and DeLey (2000). They examined how pronoun resolution is affected by the implicit causality inherent in certain verbs (e.g., the subject of the verb annoy performs some action or has some characteristic that “causes” a response from the grammatical object, whereas the object of the verb praise performs some action that “causes” a response from the grammatical subject). Knowledge about the implicit causality of a verb can be used to resolve an ambiguous pronoun (e.g., John praised Paul because he won the race).

Long and DeLey (2000) found that readers’ use of implicit causality depended on important characteristics of both the reader and the stimuli. First, good comprehenders showed an effect of implicit causality when they encountered the pronoun, whereas poor comprehenders showed the effect at the end of the sentence. Second, good comprehenders showed an effect that was limited to verbs in which the implied cause of the event was the grammatical object of the sentence. In order to explore the locus of the effect, Long and DeLey examined the use of these verbs in a large corpus of natural text. They found that the “object verbs” were better predictors of the referent of a subsequent anaphor than were the “subject verbs.” A *connectionist-based account* of these results would suggest that good comprehenders, who have considerable reading experience, encoded more information about the contextual use of these verbs than did poor comprehenders. Thus, good comprehenders expected the implied cause of an “object verb” to be the referent of a subsequent pronoun.

Figure 4 depicts hypothesized relations among reader characteristics and comprehension according to a *connectionist-based account* of WM. The figure is similar in some respects to the one illustrating *verbal efficiency theory* (see Figure 1). Word-level ability and print exposure are emphasized in both these figures. They differ, however, with respect to the role of WM capacity. According to the *connectionist-based account*, variation in performance on span tasks is not due to limitations in capacity, per se; but is due to the same factors that influence all language tasks: word-level ability and print exposure.

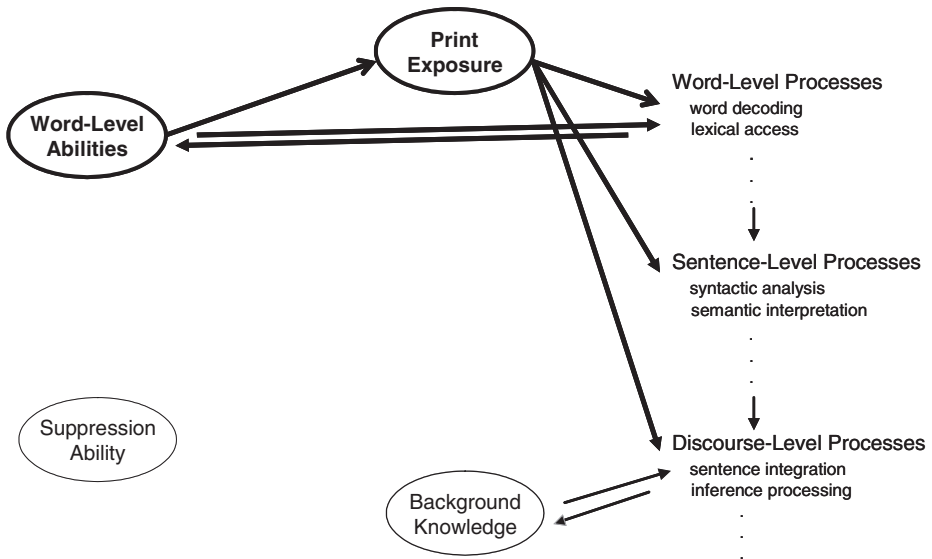


Figure 4. Hypothesized relations among the five reader characteristics according to the connectionist-based model of working memory.

3.2.2. *The long-term working memory (LTWM) Model*

Ericsson and Kintsch (1995) have argued that the standard definition of WM as a small capacity, temporary storage mechanism is too restrictive to account for skilled performance on complex tasks, such as text comprehension. They propose a model of WM that consists of the standard limited-capacity mechanism that they call short-term working memory (ST-WM), and a mechanism based on skilled storage and retrieval in long-term memory that they call long-term working memory (LTWM). In this model, the amount of information that can be actively maintained in LTWM is not limited by a fixed capacity. As individuals become skilled at a task, they develop mechanisms for encoding and retrieving information from long-term memory that meet the demands of the task.

This view is supported by evidence that individuals who exhibit large WM capacities do so only for skilled activities. Individuals who are skilled in mental calculation have large WM capacities as measured by digit span (Ericsson, 1985; Hatano, Amaiwa, & Shimizu, 1987; Jensen, 1990), but do not show large capacities for other types of materials. Experienced waiters and waitresses show large WM capacities for dinner orders, but their memory for other information is in the average range (Ericsson & Polson, 1988a, 1988b). Chess experts show superior WM for meaningful configurations of chess pieces, but not for random configurations (Chase & Simon, 1973).

Evidence for the role of LTWM in text comprehension has focused on the relative influence of domain expertise and general verbal ability on comprehension (Recht &

Leslie, 1988; Schneider, Körkel, & Weinert, 1989; Walker, 1987). Several studies have examined groups of readers who differ with respect to reading ability (e.g., high versus low performance on standardized reading tests) or general aptitude (e.g., high versus low performance on IQ tests) and with respect to domain knowledge (e.g., high versus low knowledge about baseball). In all of these studies, domain knowledge was the dominant factor in predicting comprehension performance. Ericsson and Kintsch (1995) argue that high-knowledge readers perform better than low-knowledge readers because their domain expertise gives them better strategies for encoding structures in long-term memory that can be accessed quickly and easily based on retrieval cues in ST-WM.

The same argument can be applied to why good comprehenders recall more information from texts than do poor comprehenders. Good comprehenders have strategies that are effective for encoding large and integrated structures in memory. These structures are activated when new, incoming information in a text provides cues to their retrieval. In contrast, poor comprehenders encode ideas from a text in isolation or in poorly integrated clusters. Thus, the retrieval cues in short-term memory activate small, relatively impoverished, structures from long-term memory.

Figure 5 depicts the role of LTWM and background knowledge in comprehension. The *LTWM model* is similar to the *connectionist-based account* in that it emphasizes the role of print exposure in text comprehension (see Figure 4). Reading experience helps comprehenders develop strategies for building text structures in long-term memory that can be

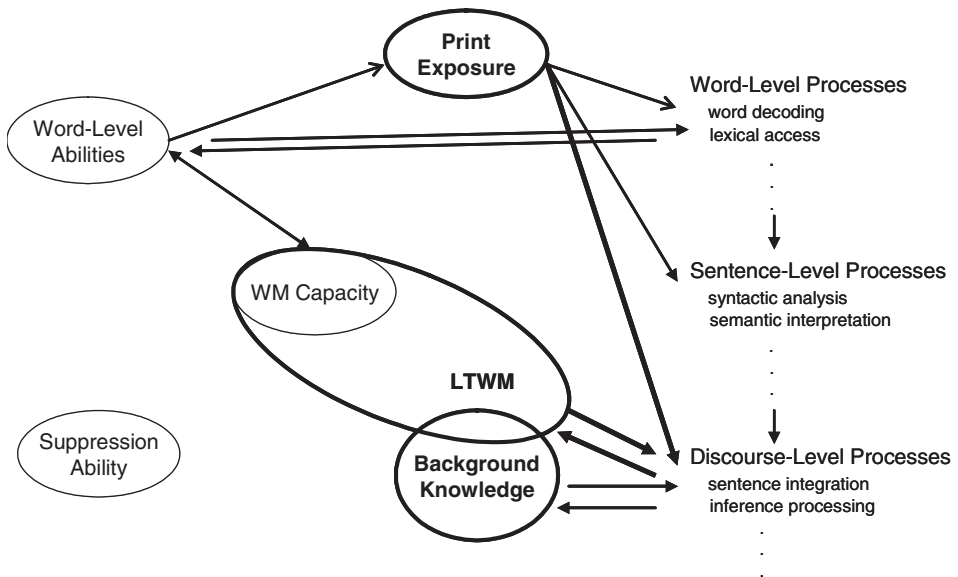


Figure 5. Hypothesized relations among the five reader characteristics according to the LTWM model.

accessed easily based on retrieval cues in short-term memory. The *LTWM model* is different from the *connectionist-based account* in that it includes a traditional view of a fixed-capacity WM system. Although this system is involved in the performance of novel tasks, it plays little role in the performance of skilled activities, such as reading comprehension.

#### 4. INDIVIDUAL DIFFERENCES IN SUPPRESSION ABILITY

Suppression ability (also called cognitive inhibition) is an individual's skill at ignoring or inhibiting distracting information and overcoming interference from a prepotent response. Suppression appears to play an important role in constructing a coherent representation of a text. Readers often activate contextually irrelevant information during comprehension. Activated, but irrelevant, information has the potential to interfere with comprehension processes. Suppression reduces the interference from such information by dampening its activation.

Most measures of suppression assess an individual's ability to resist or overcome interference. Such measures include (a) Stroop Interference—participants receive words or other stimuli (e.g., xxxx) in colored print and they name the color as quickly as possible, (b) the Go/No-go Task—participants are asked to respond when one visual target appears (e.g., a circle), but not to respond when a different visual target appears (e.g., a square), and (c) the Eriksen Flanker Task—participants respond to a target letter that is presented with distractor letters (flankers) that are either the same as the target or different from the target (Eriksen & Hoffman, 1973; Yeh & Eriksen, 1984).

Suppression plays a prominent role in Gernsbacher's (1990) *structure building framework*. According to her framework, a reader's goal is to build a coherent mental representation or "structure." Readers begin the process by establishing a foundation based on some initial information. They develop their mental structure by adding new incoming information when it relates or coheres to this representation. When readers receive information that is unrelated to previous information, they shift to initiate a new substructure. Thus, a mental representation often consists of several branching structures.

According to the theory, mental structures are built out of previously stored memory traces; these traces are activated by incoming information. Activation is modulated by two different mechanisms: enhancement and suppression. Enhancement increases the activation of memory traces when their content is relevant to the mental structure being developed. Suppression dampens activation of the traces when their content is unrelated to the structure.

Gernsbacher and her colleagues have argued that failure to suppress activated, but irrelevant, information during comprehension underlies individual differences in comprehension skill (Gernsbacher, 1993; Gernsbacher & Faust, 1991, 1995; Gernsbacher & Robertson, 1995; Gernsbacher, Varner, & Faust, 1990). This failure can be seen when readers comprehend ambiguous words. Gernsbacher et al. (1990) had participants read short sentences that ended with a homograph (e.g., The man dug with a *spade*) and then

judge whether a test probe fit the meaning of the sentence. The probes were presented either 100 or 850 ms after the offset of the sentence-final word. When the probe was related to the meaning of the sentence (e.g., garden), both good and poor comprehenders experienced facilitation. Likewise, when the probe was an inappropriate associate of the homograph (e.g., ace), both good and poor comprehenders had difficulty rejecting the probe as unrelated in the 100 ms condition. This finding suggests that the inappropriate meaning was activated at 100 ms, making it difficult to reject the test probe even though it was inappropriate in this context. In contrast, only the poor comprehenders had difficulty rejecting the probe at the 850ms delay. Gernsbacher et al. argued that good comprehenders suppressed the inappropriate sense of the homograph, whereas poor comprehenders failed to do so; thus, poor comprehenders experienced interference at the long delay.

Poor comprehenders' suppression problems have important implications for their ability to create coherent discourse representations. When readers encounter irrelevant information, they shift from mapping information onto the current structure to initiate a new substructure. Poor comprehenders' failure to suppress irrelevant information causes them to shift too often; they initiate new substructures when they should continue mapping information onto their current structure. Thus, they construct discourse representations that are less integrated than those constructed by good comprehenders.

Several experiments have been conducted to investigate the extent to which suppression is an automatic inhibitory mechanism or a controlled, strategic one (Gernsbacher & Faust, 1995; Long, Seely, & Oppy, 1999). Gernsbacher and Faust (1995) examined good and poor comprehenders' ability to suppress the irrelevant meanings of ambiguous words. They manipulated the proportion of trials on which suppression was needed and found that readers were more likely to inhibit irrelevant information when the proportion of conflict trials was high (i.e., when the target was a context-inappropriate associate of the ambiguous word) than when the proportion of such trials was low. In other words, readers inhibited irrelevant information when suppression had high utility.

Long et al. (1999) also found evidence that suppression is a strategic process. They had good and poor comprehenders read sentences that ended with a word that was unambiguous (e.g., The presence of the stranger upset the *baby*) and then respond to test probes that were backward associates of the sentence-final word (e.g., *stork*). If suppression is an automatic mechanism, then responses to the test word should be unaffected by the preceding sentence. The word *baby* does not activate the word *stork*, so suppression should not be triggered. If, however, suppression is a strategic process, it may be invoked as a consequence of the response conflict that readers experience when they compare *stork* to the preceding context. Long et al. found that both good and poor comprehenders experienced interference to the test probes when the interval between the test sentence and probe was short (100ms); however, only poor comprehenders experienced interference when the interval was long (850ms).

If suppression during comprehension is under readers' strategic control, why do poor comprehenders fail to execute it? One possibility is suggested by recent research

investigating the relation between cognitive inhibition and individual differences in WM memory. Engle and his colleagues have described a view that attributes variation in WM capacity to limitations in the ability to inhibit task-irrelevant information and to maintain activation in the face of distracting or interfering events (Conway & Engle, 1994; Engle, Conway, Tuholski, & Shisler, 1995; Engle, Kane, & Tuholski, 1999; Engle, Tuholski, Laughlin, & Conway, 1999, Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2000; Rosen & Engle, 1997). According to their view, WM is a system comprised of activated memory traces, procedures and skills necessary to achieve and maintain activation, and controlled attention (Engle, Kane, & Tuholski, 1999). Controlled attention is involved in maintaining activation of task-relevant information, particularly in the presence of distraction or interference, and in inhibiting task-irrelevant information. The capacity of controlled attention is limited and is the primary source of individual differences in the performance of complex tasks, including reading span and reading comprehension.

Figure 6 depicts a view of comprehension skill in which skill at the sentence-level and discourse-level is heavily influenced by domain-general, controlled attention—in particular, the ability to suppress activated, but irrelevant, information. This view of WM capacity and its relation to comprehension is similar to *capacity theory* in that both attribute variation in sentence and discourse-level processes to limitations in a domain-general ability (see Figure 2). They differ, however, as to the nature of this ability. In *capacity theory*,

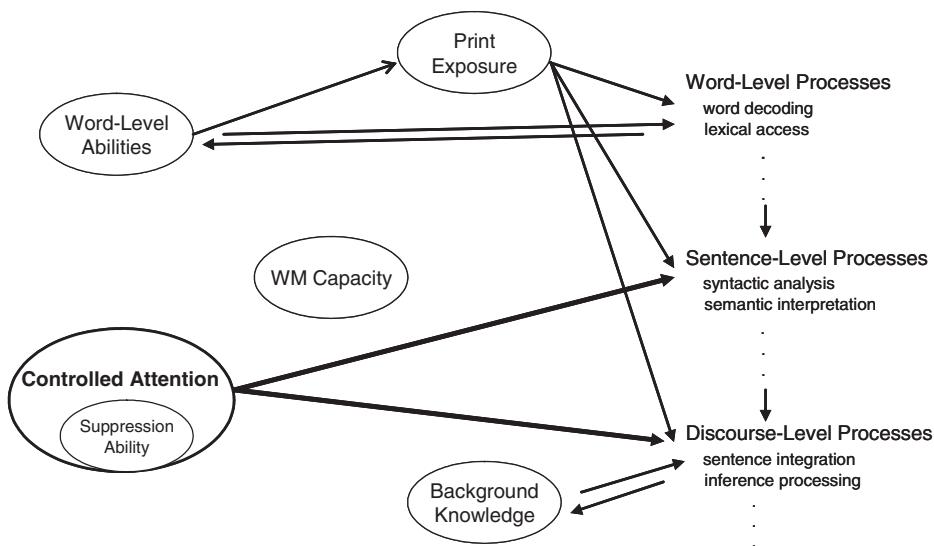


Figure 6. Hypothesized relations among the five reader characteristics according to a view of working memory that emphasizes the role of controlled attention in suppression ability and comprehension.



the limitation is in the amount of activation available to the system. In the *controlled-attention view*, variation in performance on both comprehension tasks and complex span tasks are due to individual differences in the ability to control attention, including the ability to suppress irrelevant information.

## 5. INDIVIDUAL DIFFERENCES IN PRINT EXPOSURE

Readers differ greatly in their enthusiasm for reading, in how much they value it, and in the time they spend engaged in the activity. Researchers have been interested in a variety of questions about how reading practices relate to comprehension ability and the acquisition of general world knowledge. One question has involved whether print exposure affects comprehension ability primarily by automating word-identification processes or whether print exposure also has important effects on vocabulary development, syntactic knowledge, and the acquisition of cultural and domain knowledge.

Print exposure can be assessed by means of questionnaires and interviews (Guthrie, 1981; Guthrie & Greaney, 1991; Walberg & Tsai, 1984); however, these techniques are susceptible to social desirability factors, primarily the tendency to report more time spent reading than actually occurs. Activity diaries, in which individuals record their daily activities, can provide more reliable results (Carp & Carp, 1981; Greaney, 1980; Rice, 1986); however, this technique can be expensive and time-consuming.

Stanovich and his colleagues have developed a checklist method of assessing print exposure that is immune to social desirability effects and is quickly and easily administered (Stanovich & Cunningham, 1992; Stanovich & West, 1989). The *Author Recognition Test* and the *Magazine Recognition Test* are both recognition checklists that assess participants' ability to discriminate author names and magazine titles from foils. Signal detection logic is applied such that correct responses are adjusted for guessing by examining the number of foils that are selected. Performance on these checklists correlates with measures of reading comprehension in both children and adults.

Stanovich and his colleagues have used regression techniques to examine the influence of print exposure separate from other reader characteristics, such as word-decoding ability, comprehension skill (e.g., performance on standardized measures of reading comprehension), and general cognitive ability (e.g., performance on IQ or reasoning tasks such as Raven's Matrices) (Cunningham & Stanovich, 1990, 1991; Stanovich, 1986; Stanovich & Cunningham, 1992; Stanovich & West, 1989; West & Stanovich, 1991; West, Stanovich, & Mitchell, 1993). They have found that print exposure is uniquely related to spelling ability, vocabulary, and general world knowledge after controlling for other reader characteristics.

Print exposure is likely to influence comprehension skill in at least three ways. First, individuals who read often are more likely to learn about rare words than are individuals who read seldom. This is because rare words appear more often in print than they do in

speech. Hayes and Ahrens (1988) found that rare words (those that ranked lower than 10,000 in a frequency ordered list of words; Carroll, Davies, & Richman, 1971) occurred 50% more often in children’s books than in adult conversation or prime-time television shows. Vocabulary growth, therefore, is likely to be accelerated in individuals who read often. Second, individuals are more likely to encounter complex syntactic structures in print than in speech. This is particularly true in genres such as newspapers where optional function words are often deleted due to space considerations. Finally, individuals who read often are likely to acquire more world knowledge than individuals who read seldom. Text comprehension is the primary means of knowledge acquisition in many domains.

Figure 7 depicts hypotheses about how print exposure may be related to other reader characteristics. Note that the relations among these characteristics are similar to those hypothesized in the *LTWM* model illustrated in Figure 5. Print exposure and background knowledge are emphasized in both figures. One difference is that Figure 7 emphasizes the relation between print exposure and word-level ability. Individuals who read often will learn more about words than individuals who read seldom and good word-level abilities will increase these individuals’ enthusiasm for reading. The reciprocal relation between word-level ability and print exposure is also emphasized in *verbal efficiency theory* (see Figure 1). Finally, Figure 7 depicts a strong effect of print exposure on background knowledge via its effect on discourse-level processing. Individuals who read often will learn more from texts than individuals who read seldom; increasing their general world knowledge and domain expertise.

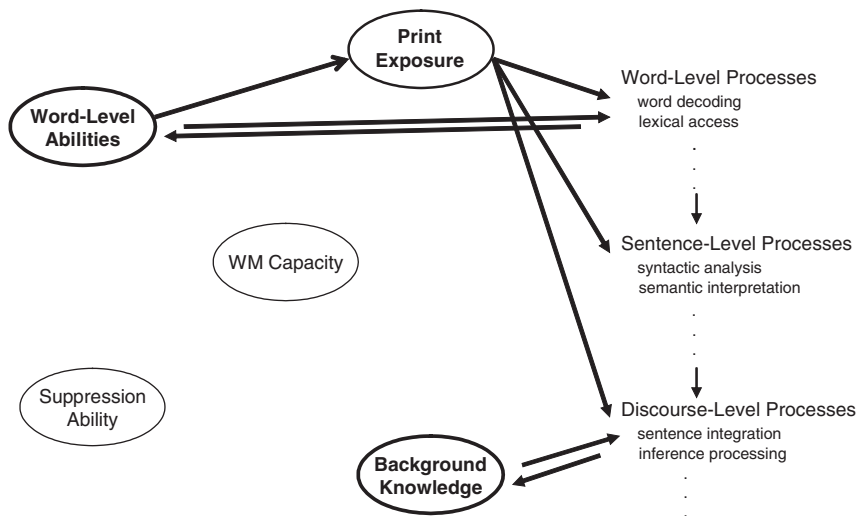


Figure 7. Hypothesized relations among the five reader characteristics according to a view of comprehension that emphasizes the role of print exposure.

## 6. INDIVIDUAL DIFFERENCES IN BACKGROUND KNOWLEDGE

The focus of our review to this point has been on reader characteristics that discriminate good comprehenders from poor ones. Variability in comprehension performance, however, can be seen even among very skilled readers. One of the earliest findings in the field of text comprehension was that readers who have relevant knowledge about the topic of a text understand it and remember it better than do readers who lack such knowledge (Bartlett, 1932; Bransford & Johnson, 1972). Experts—readers who have extensive domain knowledge—access a richly interconnected network of learned facts when reading a text relevant to their domain of expertise (Chi, Feltovich, & Glasser, 1981, Chiesi, Spilich, & Voss, 1979; Means & Voss, 1985). Moreover, experts employ more effective reading strategies than do novices (Afflerbach, 1986; Lundeberg, 1987) and are faster and more efficient at retrieving information from their knowledge domain (Ericsson & Smith, 1991).

The comprehension advantage associated with background knowledge has been well documented in two different paradigms. In one, participants read texts that contain numerous vague referring expressions. Recall for the text improves when readers are given relevant contextual knowledge, such as a title that denotes the topic of the passage (e.g., “Washing Clothes”) (Alba, Alexander, Hasher, & Caniglia, 1981; Bransford & Johnson, 1972; Summers, Horton, & Diehl, 1985). In a second paradigm, participants read coherent texts that contain information about a specific domain. Those who are knowledgeable about the domain recall more information from the text than do those who are less knowledgeable (Schneider, Körkel, & Weinert, 1989; Spilich, Vesonder, Chiesi, & Voss, 1979; Sulin & Dooling, 1974). For example, Spilich et al. (1979) found that baseball “experts” recalled more information from baseball texts than did “novices.” Moreover, experts recalled more propositions relating to actions and events that were closely associated with the goal structure of a baseball game than did novices. Schneider et al. (1990) reported similar results using soccer as the knowledge domain.

Recently, Long and her colleagues have argued that high-knowledge readers construct qualitatively different text representations than do low-knowledge readers. High-knowledge readers construct discourse models in which text ideas are integrated with each other and with a large network of relevant prior knowledge (Long & Prat, 2002; Long, Wilson, Hurley, & Prat, in press). These models support recall, problem-solving, generalization, and knowledge-based inferences. Low-knowledge readers, in contrast, construct text representations that are coherent at the sentence-level, but they lack the knowledge necessary to construct coherent discourse models. Thus, low-knowledge readers can recognize ideas from a text, but cannot use their representations to perform tasks that require conscious, reflective access to a discourse model.

Long and Prat (2002) used a recognition-memory paradigm to examine qualitative differences in high-knowledge and low-knowledge readers’ text representations. Many memory researchers believe that recognition involves at least two component processes: recollection and familiarity. The nature of these two processes differs somewhat across dual-process models (for a review see Yonelinas, 2002). For example, Yonelinas and his

colleagues argue that recollection involves retrieval of specific information about a studied item, such as information about the context in which the item appeared (Dobbins, Kroll, & Liu, 1998; Yonelinas, 1997, 1999, 2001, 2002). Familiarity, in contrast, involves an assessment of the similarity (perceptual and conceptual) between a test item and a memory trace. Rajaram (1996) has argued that recollection reflects elaborative and distinctive processing that occurs at study, whereas familiarity reflects the fluency of processing that occurs at test. Wixted and Stretch (2004) have argued that both recollection and familiarity are continuous variables reflecting memory strength and that the two are combined into a single memory signal. Although these models differ in critical ways, they share the core assumption that recollection and familiarity are distinct processes and can be empirically dissociated.

According to Long and Prat (2002), the processes involved in constructing sentence-level representations give rise to familiarity at test, whereas the processes involved in constructing a discourse model give rise to recollection. Familiarity arises from the perceptual and semantic processing that occurs when participants encode a to-be-learned item. A substantial amount of this type of processing occurs when readers comprehend sentences in texts, even when they do not possess domain-relevant knowledge. Thus, familiarity should support recognition of text ideas even in the absence of the elaborative processing involved in constructing a discourse model.

In contrast, discourse-level processing involves forming associative relations between text ideas and prior knowledge. If a text idea activates extensive knowledge during comprehension, a network of connections will be formed that integrates the idea with the reader's prior knowledge. When the text idea is presented at test, it will resonate with its item representation in memory, reactivating the network of contextual information that was constructed during comprehension. Retrieval of contextual information about the study context will give rise to an experience of recollection. In addition, some text ideas may evoke conscious inferences when readers have extensive knowledge about a topic. For example, readers who are knowledgeable about a particular genre of stories, such as horror stories, may make an explicit prediction in response to a character's action (e.g., if the character says, "I'm going outside, I'll be right back," the reader may consciously predict that the character will be eaten by the monster). If the action is presented at test, the reader may retrieve the inference that was associated with it at study, leading to an experience of recollection.

These hypotheses were tested using the remember/know paradigm developed by Tulving (1985). Participants made judgments concerning the nature of their memory for recognized items, responding "remember" to items that were accompanied by recollection of details about the item's prior occurrence, and responding "know" to items that were recognized from the study episode, but were not accompanied by recollection. Readers were tested to assess their knowledge about the science-fiction saga Star Trek. They then read a short story about Star Trek and received a recognition test consisting of sentences from the story that they read as well as distractor sentences from a Star Trek story that they did not read. In addition, they read a chapter from an introductory psychology textbook and received a subsequent recognition test.

Long and Prat (2002) found no effect of prior knowledge on overall recognition for either the *Star Trek* story or for the Psychology chapter. They did find an effect, however, when they examined remember and know responses. High-knowledge readers were more likely to report a vivid, conscious experience of recollection in response to text ideas than were low-knowledge readers, but only for the *Star Trek* items. Long and Prat found similar effects when they examined recollection and familiarity by means of the process-dissociation procedure, a procedure that assesses the extent to which individuals can remember the specific context in which an item appeared.

Similar results were found in a study by Long et al. (in press). They examined the influence of both domain knowledge and text coherence on readers' memories for text ideas. Previous research has shown that the influence of text coherence on comprehension depends critically on the reader's prior knowledge (McNamara, 2001; McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). Somewhat surprisingly, manipulations that decrease coherence can improve text comprehension and recall. This occurs when readers use their background knowledge to fill conceptual gaps in the text. Only high-knowledge readers, however, benefit from low-coherence texts because only they have the knowledge that is necessary to establish coherence among text ideas.

Long et al. (in press) examined the interaction of prior knowledge and text coherence on recollection and familiarity in recognition. If recollection is a consequence of processing at the discourse-level, then high-knowledge readers should have higher recollection estimates than low-knowledge readers. Moreover, high-knowledge readers should have higher recollection estimates in response to low-coherence than to high-coherence texts. Low-coherence texts are more likely than high-coherence texts to involve retrieval of relevant domain knowledge during comprehension, leading to a more elaborate discourse model at encoding and to the experience of recollection at test. Long et al. found support for these hypotheses in experiments involving both the remember/know task and the process-dissociation procedure.

The role of background knowledge in comprehension is emphasized in the *LTWM model* (see Figure 5) and in research on individual differences in print exposure (see Figure 7). In Figure 5, background knowledge is essential for building retrieval structures in *LTWM* that expand a reader's ability to hold large amounts of information in an accessible form. In Figure 7, background knowledge is facilitated by print exposure. Individuals who read often are exposed to more information about the world than are individuals who read seldom and they are more likely to create coherent discourse models that expand their knowledge base.

## 7. CONCLUSIONS

What is it that skilled adult readers do when they comprehend a text that less-skilled readers fail to do? The answer to this question is fairly clear. Good comprehenders construct quality representations of individual sentences and then reinterpret, reorganize, and integrate their sentence representations in light of prior knowledge that is relevant to the

text—that is, they construct coherent discourse models. Poor comprehenders, in contrast, understand individual sentences reasonably well, but fail to integrate their sentence representations into a coherent whole. They represent text ideas in isolation or in clusters, loosely organized by theme.

Why do poor comprehenders fail to execute the high-level interpretive processes that result in a coherent discourse model? The answer to this question is unclear even after decades of research. We do know that good and poor comprehenders differ in ways that are strongly correlated with comprehension performance. We still do not understand, however, the exact nature of these correlations. What accounts for the correlation between complex span tasks and comprehension performance? Is the relation between span and comprehension secondary to deficits in word-level ability? Do good and poor comprehenders differ primarily with respect to their ability to construct effective retrieval structures and store them in LTWM?

One of the obstacles in answering the questions posed above is methodological. Individual differences in reading are typically studied using quasi-experimental designs (e.g., Gernsbacher & Faust, 1991, 1995; Just & Carpenter, 1992; Long & DeLey, 2000; Long et al., 1994, 1997; McNamara & Kintsch, 1996; Pearlmutter & MacDonald, 1995). Readers are grouped according to some reader characteristic (e.g., performance on a span task) and then participate in an experiment in which some text variable is manipulated. Such studies have provided important information about how reading is affected by various combinations of reader and text characteristics, but they also have important limitations. One limitation is that quasi-experimental designs necessarily involve small numbers of variables. One of our goals in this chapter has been to illustrate the difficulty in understanding the role of one reader characteristic in comprehension when that characteristic is correlated with other reader characteristics.

In our view, we are fast approaching the limit of what can be learned about individual differences in comprehension from small quasi-experiments or simple regression studies. Substantial progress will depend on the use of more sophisticated modeling approaches (e.g., factor analysis, multilevel modeling), approaches that can be used to test the different relations among reader characteristics that we have described in this chapter.

Research on individual differences in the ability to construct coherent discourse models is important for understanding the nature of reading ability and disability. We hope that our review has also shown that such research is important for understanding how language-specific processes interact with more domain-general abilities and for understanding the nature of controlled attention, working memory, and the development of expertise.

## REFERENCES

Afflerbach, P. (1986). The influence of prior knowledge on expert readers' importance assignment processes. *National Reading Conference Yearbook*, 35, 30–40.

- Alba, J. W., Alexander, S. G., Hasher, L., & Caniglia, K. (1981). The role of context in the encoding of information. *Journal of Experimental Psychology: Human Learning & Memory*, 7, 283–292.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In: G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.
- Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology*, 86, 244–255.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, 11, 717–726.
- Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, 12, 637–649.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77–126.
- Carp, F. M., & Carp, A. (1981). The validity, reliability and generalizability of diary data. *Experimental Aging Research*, 7, 281–296.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston, MA: Houghton Mifflin.
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In: W. G. Chase (Ed.), *Visual information processing* (pp. 215–218). New York: Academic Press.
- Chi, M., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152.
- Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, 18, 257–273.
- Clifton, C. Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49, 317–334.
- Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource dependent inhibition model. *Journal of Experimental Psychology: General*, 123, 354–373.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104, 163–191.

- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology, 82*, 733–740.
- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*, 264–274.
- Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology, 72*, 656–669.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450–466.
- Dobbins, I. G., Kroll, N. E. A., & Liu, Q. (1998). Confidence-accuracy inversions in scene recognition: A remember-know analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1306–1315.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 972–992.
- Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A resource account of inhibition. *Psychological Science, 6*, 122–125.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In: A. Miyake, & P. Shah (Eds), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York, NY: Cambridge University Press.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.
- Ericsson, K. A. (1985). Memory skill [Special issue]. *Canadian Journal of Psychology, 39*, 188–231.
- Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In: A. Miyake, & P. Shah (Eds), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257–297). New York, NY: Cambridge University Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.



- Ericsson, K. A., & Polson, P. G. (1988a). An experimental analysis of the mechanisms of a memory skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 305–316.
- Ericsson, K. A., & Polson, P. G. (1988b). A cognitive analysis of exceptional memory for restaurant orders. In: M. T. H. Chi, R. Glaser, & M. J. Farr (Eds). *The nature of expertise* (pp. 23–70). Hillsdale, NJ: Erlbaum.
- Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In: K. A. Ericsson, & J. Smith (Eds). *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). New York, NY: Cambridge.
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*, 155–160.
- Frederiksen, J. R. (1981). Sources of process interactions in reading. In: A. M. Lesgold, & C. A. Perfetti (Eds). *Interactive processes in reading* (pp. 361–386). Hillsdale, NJ: Erlbaum.
- Garnham, A., Oakhill, J. V., & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. *Cognition*, *11*, 29–46.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science*, *4*, 294–298.
- Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 245–262.
- Gernsbacher, M. A., & Faust, M. E. (1995). Skilled suppression. In: F. N. Dempster & C. J. Brainerd (Eds), *Interference and inhibition in cognition* (pp. 295–327). San Diego, CA: Academic Press.
- Gernsbacher, M. A., & Robertson, R. R. W. (1995). Reading skill and suppression revisited. *Psychological Science*, *6*, 165–169.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 430–445.
- Greaney, V. (1980). Factors related to amount and time of leisure time reading. *Reading Research Quarterly*, *15*, 337–357.
- Guthrie, J. T. (1981). Reading in New Zealand: Achievement and volume. *Reading Research Quarterly*, *17*, 6–27.
- Guthrie, J. T., & Greaney, V. (1991). Literacy acts. In: R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds). *Handbook of reading research* (Vol. 2, pp. 68–96). New York, NY: Longman.

- Hatano, G., Amaiwa, S., & Shimizu, K. (1987). Formation of a mental abacus for computation and its use as a memory device for digits: A developmental study. *Developmental Psychology*, *23*, 832–838.
- Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of “motherese?” *Journal of Child Language*, *15*, 395–410.
- Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, *108*, 151–181.
- Jensen, A. R. (1990). Speed of information processing in a calculating prodigy. *Intelligence*, *14*, 259–274.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.
- Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 336–358.
- Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In: A. Miyake, & P. Shah (Eds). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 183–223). New York, NY: Cambridge.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.
- Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the “working memory” of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, *112*, 240–264.
- Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders’ abilities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 1424–1429.
- Long, D. L., & De Ley, L. (2000). Understanding anaphors in story dialogue. *Memory & Cognition*, *28*, 731–738.
- Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes*, *16*, 55–73.
- Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1456–1470.

- Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual differences in readers' sentence and text-level representations. *Journal of Memory and Language*, *36*, 129–145.
- Long, D. L., Oppy, B. J., & Seely, M. R. (1999). The strategic nature of less skilled readers' suppression problems. *Discourse Processes*, *27*, 281–302.
- Long, D. L., & Prat, C. S. (2002). Memory for Star Trek: The role of prior knowledge in recognition revisited. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 1073–1082.
- Long, D. L., Prat, C. S., Blozis, S. A., Widaman, K. F., & Traxler, M. J. (2006). A multilevel model of sentence reading time. In preparation.
- Long, D. L., Wilson, J., Hurley, R., & Prat, C. S. (in press). Assessing Text Representations with Recognition: The Interaction of Domain Knowledge and Text Coherence. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.
- Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In: A. Miyake, & P. Shah (Eds). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). New York, NY: Cambridge.
- Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis, *Reading Research Quarterly*, *22*, 407–432.
- Martin, R. C. (1995). Working memory doesn't work: A critique of Miyake et al.'s capacity theory of aphasic comprehension deficits. *Cognitive Neuropsychology*, *12*, 623–636.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.
- Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, *75*, 314–318.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51–62.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.
- Means, M. L., & Voss, J. F. (1985). Star Wars: A developmental study of expert and novice knowledge structures. *Journal of Memory & Language*, *24*, 746–757.

- Olson, R. K., Kliegl, R., Davidson, F., & Foltz, G. (1985). Individual and developmental differences in reading disability. In: G. E. MacKinnon, & T. Waller (Eds). *Reading research: Advances in theory and practice* (pp. 1–64). San Diego, CA: Academic Press.
- Palmer, J., MacLeod, C. M., Hunt, E., & Davidson, J. (1985). Information processing correlates of reading. *Journal of Memory & Language*, 24, 59–88.
- Pearlmuter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory & Language*, 34, 521–542.
- Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford.
- Perfetti, C. A. (1989). There are generalized abilities and one of them is reading. In: L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 307–335). Hillsdale, NJ: Erlbaum.
- Perfetti, C. A. (1991). Representation and awareness in the acquisition of reading competence. In: L. Rieben, & C. A. Perfetti, (Eds). *Learning to read: Basic research and its implications* (pp. 33–44). Hillsdale, NJ: Erlbaum.
- Perfetti, C. A. (1994). Psycholinguistics and reading ability. In: M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego, CA: Academic Press.
- Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly*, 33, 283–319.
- Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In: P. A. Carpenter, & M. A. Just (Eds). *Cognitive processes in comprehension* (pp. 141–183). Hillsdale, NJ: Erlbaum.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 365–377.
- Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning & Verbal Behavior*, 17, 403–417.
- Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, 80, 16–20.
- Rice, G. E. (1986). The everyday activities of adults: Implications for prose recall. *Educational Gerontology*, 12, 173–186.
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126, 211–227.
- Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory & Language*, 39, 418–436.

- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.
- Schneider, W. (1999). Working memory in a multilevel hybrid connectionist control architecture (CAP2). In: A. Miyake, & P. Shah (Eds). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 340–374). New York, NY: Cambridge.
- Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*, 306–312.
- Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1–30.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, *18*, 275–290.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–406.
- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, *20*, 51–68.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*, 402–433.
- Stoltzfus, E. R., Hasher, L., & Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. In: J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds), *Working memory and human cognition* (pp. 66–88). New York, NY: Oxford.
- Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, *103*, 255–262.
- Summers, W. V., Horton, D. L., & Diehl, V. A. (1985). Contextual knowledge during encoding influences sentence recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 771–779.
- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*, 204–224.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.
- Vogel, S. A. (1975). *Syntactic abilities in normal and dyslexic children*. Baltimore: University Park Press.
- Walberg, H. J., & Tsai, S. (1984). Reading achievement and diminishing returns to time *Journal of Educational Psychology*, *76*, 442–451.

- Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction, 4*, 25–42.
- Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*, 761–772.
- Waters, G., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. *Cognitive Neuropsychology, 8*, 81–126.
- West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2*, 325–330.
- West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly, 28*, 35–50.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.
- Yeh, Y., & Eriksen, C. W. (1984). Name codes and features in the discrimination of letter forms. *Perception & Psychophysics, 36*, 225–233.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*, 747–763.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 1415–1434.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*, 361–379.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language, 46*, 441–517.
- Young, R. M., & Lewis, R. L. (1999). The Soar cognitive architecture and human working memory. In: A. Miyake, & P. Priti (Eds). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 224–256). New York, NY: Cambridge.

This page intentionally left blank

## Chapter 21

### Figurative Language

*Raymond W. Gibbs, Jr. and Herbert L. Colston*

What defines “figurative language” as opposed to any other kind of speech? Consider the opening paragraphs of the following article from the *Good Times*, a Santa Cruz, California news and entertainment weekly (Nov 4–10, 2004, p.8). The article is titled “David vs. Goliath: Round One,” and describes the University of California, Santa Cruz’s controversial plan to double in physical size and increase enrollment by over 6000 students. Read through the following text and pick out those words and phrases that appear to express figurative meaning.

“Hidden in the shadows of a massive election year, tucked under the sheets of a war gone awry and a highway scuffle, another battle has been brewing.”

“When UC Santa Cruz released the first draft on its 15-year Long Range Development Plan (LRDP) last week, it signaled an ever-fattening girth up on the hill. While some businesses clapped their hands with glee, many locals went scrambling for belt-cinchers.”

“The LRDP calls for 21,000 students by the year 2020 – an increase of 6,000 over today’s enrollment . . . . The new enrollment estimate may have startled some residents, but as a whole it merely represents a new stage in a decades-long battle that has been fought between the city and the City on the Hill. While some students are boon to local businesses and city coffers, many residents complain students are overrunning the town—clogging the streets, jacking up rents and turning neighborhoods and the downtown into their own party playground . . . .”

“The bottom line is that the university can do what it wants to,” explains Emily Reilly, Santa Cruz City Council member and head of a committee developed to open up dialogue between “the campus and the city.”

These paragraphs are not atypical of the writing that appears in many newspapers, and most readers easily understand the text’s meaning, both at an individual sentence level, and its overall message. But picking out the words and phrases that express specific



figurative meanings is more difficult. For instance, the first line opens with a phrase “Hidden in the shadows of a massive election year” that seems to be figurative in the sense that election years cannot really cast shadows. Yet does it matter whether this phrase is termed “metaphoric” or ‘metonymic’ or perhaps even “ironic”? Similarly, the next phrase “tucked under the sheets of a war gone awry” also seems figurative given that wars are not really beds with sheets. But should these expressions be classified as idiomatic, metaphoric, or just more generally as figurative? The final phrase in the opening line, “another battle has been brewing,” seems less figurative, because many things besides liquids are frequently talked about as if they are “brewing” (e.g., arguments, ideas, emotions), enough so that “brewing” has a conventional meaning of “slowing developing” that can be applied to both physical and nonphysical entities. Some readers, nonetheless, claim that “another battle has been brewing” does have a figurative meaning, more specifically, a metaphorical reading, given that only liquid entities can, strictly speaking, be brewed.

Lastly, the article contained many other phrases that do not seem particularly figurative, but which nonetheless reflect a kind of nonliteral meaning, such as “UC Santa Cruz released the first draft” and “some businesses clapped their hands.” Both phrases are figurative in that universities and businesses do not literally release anything or clap their hands – only the people running the university and businesses can do so. Consider also the utterance opening the last paragraph; “The bottom line is that the university can do what it wants to.” Does “the bottom line” have literal or figurative meaning? Once more, if it is judged to be nonliteral, what kind of figurative meaning does it express?

Of course, during ordinary language use people are rarely aware of whether words and phrases have literal, figurative, or some other type of meaning – they simply try to understand the discourse given the present context and their own personal goals. This fact about ordinary language use raises the question of whether there is anything special about figurative language, such that it necessarily requires different cognitive processes to produce and understand compared to nonfigurative speech. A great deal of psycholinguistic research suggests that many forms of figurative language may be interpreted as readily as most nonfigurative discourse, although there are instances where some forms of figurative meaning may require a good deal of effort to understand and may produce special cognitive effects, or meanings (Gibbs, 1994).

Our concern in this chapter is to take a new look at the continuing debates in psycholinguistics over what is special about figurative language use. We suggest that figurative language does not constitute a unified class of linguistic materials that are understood by special figurative processes. Nonetheless, the indeterminate nature of many aspects of figurative meaning, a fact that is not properly acknowledged in many psycholinguistic studies, raises important issues about the possible trade-off between minimizing cognitive effort and maximizing cognitive effects during figurative language processing. We suggest ways that this trade-off can be empirically studied and form the basis for future psycholinguistic research on figurative language.

## 1. DISTINGUISHING FIGURATIVE FROM LITERAL LANGUAGE

One of the continuing difficulties with the psycholinguistics literature on figurative language understanding is that few scholars ever attempt to define the terms “literal” and “figurative.” A traditional assumption in many academic disciplines is that literal meaning is primary and the product of default language comprehension. Thus, in psycholinguistic terms, the human language processor is designed for the analysis of literal meanings. Nonliteral, indirect, and figurative meanings are secondary products, and dependent on some prior analysis of what words and expressions literally mean. This general theory implies that nonliteral meanings should always take more time to interpret than are literal meanings.

Psycholinguistic research over the past 40 years has struggled to create adequate accounts of sentence parsing and discourse processing. Although there has been significant progress in our understanding of different aspects of on-line sentence processing in regard to specific topics (e.g., the interaction of syntax and semantics in sentence parsing, reference assignment, ambiguity resolution, establishing coherence relations in text), there is no single agreed upon position as to what people ordinarily do as they encounter language word by word in speech and reading. Thus, there is really not a single position on literal meaning processing. This state of affairs highlights the absurdity of theories of figurative speech processing that are often based on unverified assumptions as to how so-called literal language is usually understood.

In fact, it is not clear what the operational definition of “literal” meaning is in most psycholinguistic experiments. These studies individually compare metaphoric vs. literal meaning, ironic vs. literal meaning, idiomatic vs. literal meaning, metonymic vs. literal meaning, and so on. But across the vast number of empirical studies that have compared “literal” and “figurative” meaning, the variety of forms for literal utterances is as great as are the differences between metaphors, metonymies, ironies, and so on. Yet scholars continue to assume that the literal meaning they examine empirically somehow is the same variable that other researchers investigate in their respective experiments.

A related tendency in research on figurative language has been to note the difficulty in making a principled distinction between literal and figurative language, or meanings, and to suggest, alternatively, that literal and figurative represent different ends of a continuum of meaning. This idea is seen as especially useful in recognizing that some instances of figurative language, such as novel, poetic metaphor seem more nonliteral than are highly conventionalized phrases which almost seem to express literal meanings (e.g., “kick the bucket” has “to die” as one of its literal meanings). Individual word meanings may also vary along this literal vs. figurative continuum.

But making these distinctions, even along some graded continuum makes little sense, especially if one is trying to squeeze all aspects of literal and figurative meanings onto a similar scale. Without some consistent idea of what constitutes the notions of “literal” and “figurative” meanings, there is no way of defining the extremes of this proposed continuum. For example, the most novel, poetic instances of metaphor and irony differ from each other in numerous ways (e.g., irony requires meta-representational inferences to be

understood in a way that metaphor does not – see Colston & Gibbs, 2002). Even novel metaphors may differ dramatically with some being spectacular instantiations of well-known conceptual metaphors (e.g., “Our marriage was a roller coaster ride through hell” related to RELATIONSHIPS ARE JOURNEYS) and others reflecting completely new “one-shot” mappings (e.g., “The soul is a rope that ties heaven and earth”). On the literal side of the continuum, different instances of so-called literal meanings may vary along numerous dimensions, depending in part on what aspects of literality are being emphasized (e.g., subject-matter literality, conventional literality, context-free literality, truth-conditional literality) (Gibbs, 1994). For these reasons, the well-intended move toward thinking about literal and figurative meanings as existing along some continuous dimension makes little sense. There is simply no single dimension along which all instances of literal and nonliteral meanings nicely align.

One general implication of the above is that there may not be a unified theory of figurative language use and understanding, precisely because the reasons for using different tropes, and the mental processes involved in understanding metaphor, metonymy, irony, and so on are quite different and cannot be subsumed under a single umbrella that is distinct for figurative language alone.

## 2. TRADITIONAL THEORIES AND EMPIRICAL RESULTS

Following the traditional belief about differences between literal and figurative language, psycholinguists have focused a great deal on examining the possibility that figurative language is understood after some sort of preliminary analysis of an expression’s literal meaning (Gibbs, 1994, 2002). The most famous, and now traditional, view of how listeners understand nonliteral meaning comes from H. Paul Grice’s theory of conversational implicature (Grice, 1989), often dubbed the “standard pragmatic” view. Grice argued that the inferences needed to understand nonliteral meaning are derived from certain general principles or maxims of conversation that participants in talk – exchange are mutually expected to observe (Grice, 1989). Among these are the expectations that speakers are to be informative, truthful, relevant, and clear in what they say. When an utterance appears to violate any of these maxims, as do many of the figurative expressions in the opening newspaper article, listeners, or readers, are expected to subsequently derive an appropriate “conversational implicature” about what the speaker intended to communicate in context given the assumption that he or she is trying to be cooperative.

The results of many psycholinguistic experiments have shown the standard pragmatic view to be incorrect as a psychological theory (see Gibbs, 1994; Glucksberg, 2001). Numerous reading-time and phrase classification studies demonstrate that listeners/readers can often understand the figurative interpretations of metaphors, irony/sarcasm, idioms, proverbs, and indirect speech acts without having to first analyze and reject their literal meanings when these expressions are seen in realistic social contexts. For instance, people can read figurative utterances (i.e., “You’re a fine friend” meaning “You’re a bad friend”) as quickly, sometimes even more quickly, as literal uses of the same expressions in different

contexts, or equivalent nonfigurative expressions. These experimental findings demonstrate that the traditional view of figurative language as *always* requiring additional cognitive effort to be understood has little psychological validity.

But the idea that people can use context to infer figurative meaning without a literal analysis of an expression has been criticized on various grounds. First, there has been misunderstanding of the claim that figurative language can be understood “directly.” This suggestion does not imply that people do not process the meanings, literal or otherwise, of the individual words in each expression. The work showing that people can process many instances of figurative language as quickly as they do nonfigurative speech only implies that a complete analysis of an expression need not be completed before any interpretation of its intended figurative meaning can begin (Gibbs, 2002).

Second, some studies have found evidence that people take longer to process figurative language than corresponding literal speech, exactly as would be predicted by the traditional view (Blank, 1988; Giora, 2002; Schwoebel, Dews, Winner, & Srinivas, 2000). Yet in at least some cases, the contexts used in these studies were relatively weak in supporting figurative meanings. For instance, remarks like “You’re just in time” took longer to read in ironic context (i.e., when someone was quite late) than in literal ones (Giora, Fein, & Schwartz, 1998), especially when the irony was unexpected. But in other studies, the context in which an ironic remark appeared set up an ironic situation so that the speaker’s utterance was easily understood as having ironic meaning and took no longer, and occasionally less time, to process than literal statements (Gibbs, 1986a, 1986b). Similar effects have been reported in regard to metaphor understanding where some contexts set up metaphorical conceptualizations of topics that make following metaphoric utterances easy to interpret (Gentner, Imai, & Boroditsky, 2002; Pfaff, Gibbs, & Johnson, 1997). People may still need to draw complex inferences when understanding some figurative statements, but part of these inferences can occur before one actually encounters a figurative utterance. Other studies show that familiar conventional instances of figurative language (e.g., “John kicked the bucket,” “A fine friend you are”) take less time to interpret than do novel figurative expressions (Giora et al., 1998; Temple & Honeck, 1997). Listeners may take longer to understand a novel expression because of the difficulty in integrating the figurative meaning with the context and not because listeners are first analyzing and then rejecting the expression’s literal meaning (Schraw, 1995; Shinjo & Myers, 1987). For these reasons, we simply should not infer that the literal meaning for an entire phrase or expression must have been analyzed simply because people take longer to read novel instances of figurative language than to process either familiar figurative expressions or equivalent literal statements (Brisard, Frisson, & Sandra, 2000). Bowdle and Gentner (2005) also caution that equating conventionality with directness of processing may be an oversimplification. The processing required to interpret novel figurative language depends on many factors, including grammatical form, context, and whether different instances are related to pre-existing figurative schemes of thought (Bowdle & Gentner, 2005; Gibbs, 1994), enough so that even novel expressions may require as little time to understand as do conventional figurative utterances.

Another body of research has suggested that on-line studies may be better indicators of literal meaning activation than are more global measures of utterance comprehension, such as reading time and phrase classification techniques (Brisard et al., 2000). These on-line studies usually examine the activation of literal and figurative meanings at different points during and at the end of figurative utterance comprehension. For instance, one research project examined comprehension of familiar and less familiar metaphorical expressions (Blasko & Connine, 1993). Participants in these experiments heard different sentences and made lexical decisions at various times to visually presented word strings. For instance, as participants heard the sentence “The belief that hard work is a ladder is common to this generation,” they were visually presented a letter string immediately after hearing the word “ladder.” The letter string visually presented was related to some aspect of the sentence’s literal meaning (e.g., “rungs”), a letter string related to the sentence’s metaphoric meaning (e.g., “advance”), or a control word unrelated to the sentence (e.g., “pastry”). The results revealed that participants were equally fast in responding to the literal and metaphorical targets, which were both faster than the latencies to the controls. This was true both when participants made their lexical decisions immediately after hearing the critical word (e.g., “ladder”), and when the same decisions were made 300 ms after hearing the critical word. However, when participants made these same types of lexical decisions to literal and metaphorical targets having heard less familiar expressions, such as “The thought that a good professor is an oasis was clung to by the entire class,” only literal targets were primed immediately after hearing the critical word (e.g., “oasis”), while responses to the metaphorical targets were facilitated only 750 ms after the critical word.

But these studies have one important methodological flaw in their equating different aspects of meaning (word vs. phrasal) with response times to literal (word) and metaphoric (phrasal) targets. For example, in Blasko and Connine (1993) the literal target “rung” is a simple semantic associate of the word “ladder,” while the metaphoric target “advance” only relates to the general meaning of the entire expression. This makes it difficult to conclude anything about the time-course under which literal meanings of an entire sentence are activated compared to figurative meanings of these expressions. Even if one conceives of literal meaning as only relating to individual word meaning, this study does not compare activation of literal word meanings with figurative word meanings. Moreover, the words used as literal and metaphoric targets do not seem to reflect very distinctive literal and figurative meanings. The literal target “rung,” for instance, is related to the idea of advancing (i.e., the figurative target) given that climbing ladders, even literally speaking, is an instance of advancing along some physical path. We believe these problems plague a good deal of the studies using lexical priming techniques to examine figurative language processing.

A different issue with many studies is the assumption that the activation of a particular meaning (i.e., literal or idiomatic) reflects the output of entirely different linguistic processes. The possibility remains that activation of different kinds of meaning (i.e., literal or idiomatic) reflects different types of meaning accessed by a single linguistic process. The fact that psycholinguists label one kind of meaning as “literal” and another “figurative” does not necessarily indicate that different processes operate (i.e., a literal

processing mode and a idiomatic or figurative processing mode) to access these meanings (either in a serial or parallel manner). There are many types of figurative meaning, including metaphoric, idiomatic, metonymic, ironic, satirical, proverbial, hyperbolic, oxymoronic, and so on. Scholars often assume within the context of a single set of studies that there are two processes at work during figurative language understanding, such as literal vs. idiomatic, literal vs. metaphoric, or literal vs. ironic. Yet if there are numerous types of meaning, must there be dozens of types of linguistic processes all at work, or potentially at work, when language is understood? Psycholinguists have not addressed this question primarily because they focus too narrowly on only one kind of figurative meaning against a simple view of literal meaning.

A related type of study that examined the time-course for understanding literal and figurative interpretations of simple sentences used a signal, speed-accuracy trade-off procedure (McElree & Nordlie, 1999). Participants were presented strings of words, one at a time, at a rate that approximated fast reading (250 ms/word). The final word in each string produced a literal (e.g., “some tunnels are sewers”), a figurative (e.g., “some mouths are sewers”), or a nonsensical interpretation (e.g., “some cattle are sewers”). Participants judged whether each word string was meaningful when a tone appeared at varying times after the critical, last word. No differences were found in the comprehension speed for literal and figurative strings. McElree and Nordlie argued that the lack of time-course differences is inconsistent with the claim that figurative interpretations are computed after a literal meaning had been analyzed. In general, the time-course data presumably support the idea that literal and figurative interpretations are computed in parallel.

But we question whether the null results (e.g., no difference in processing literal and figurative sentences) obtained in these experiments necessarily provide evidence in favor of a parallel processing model. The activation of a particular meaning (i.e., literal or idiomatic) is assumed to reflect the output of entirely different linguistic processes. Once again, the possibility remains, however, that activation of different kinds of meaning (i.e., literal or idiomatic) may arise from a single linguistic process.

### 3. NEW MODELS AND FINDINGS

The continuing debates over the traditional view of figurative language understanding have led to the development of several alternative theories, specifically focused on the role of context in figurative language processing. These new models generally aim to describe the influence of context on figurative language processing at a more fine-grained level than the earlier proposals. Thus, the newer models suggest when and how context prompts figurative meanings during word-by-word linguistic processing. At the same time, these newer models attempt to offer general accounts that may apply to all aspects of figurative language, compared to most theories that aim to describe individual tropes (e.g., metaphor, irony, proverbs). Although these models recognize that some trope-specific types of processing may be necessary, they suggest that some obligatory linguistic processes operate with all types of figurative language.

Perhaps the most prominent of these new models is the “graded salience hypothesis” (Giora, 2002). This account specifically claims that context functions to constrain figurative meanings only after salient word or phrase meanings have already been accessed. Salient word or phrase meanings are not necessarily “literal” meanings. Instead, salient meanings reflect the most common, conventional use of a word or phrase. Unlike the standard pragmatic view, however, context may facilitate activation of figurative meanings before people analyze the semantic, or literal, meanings of the entire linguistic expression. For instance, processing familiar metaphors (e.g., “step on someone’s toes”) should activate both of their literal (e.g., foot) and metaphoric (e.g., offend) meanings, even when these metaphors are seen in appropriate discourse contexts. Processing unfamiliar metaphors (e.g., “Their bone density is not like ours”) may, on the other hand, only initially activate their literal meanings, as these are most salient.

Different empirical studies, ranging from reading-time to word-fragment completion experiments, support this general idea for how people interpret different kinds of figurative language, in addition to how jokes may be understood. For example, consider the findings of a set of studies looking at irony comprehension (Giora & Fein, 1999). These studies examined people’s understanding of familiar (e.g., “Very funny”) and less familiar (e.g., “Thanks for your help”) ironies in comparison to literal uses of the same expressions in appropriate contexts. Participants read stories ending with either literal or ironic remarks. After reading the final sentence, participants were presented with a letter string and had to quickly respond whether that string was a meaningful word. For instance, after reading the statement “Thanks for your help,” participants were presented with either a ironic test word (e.g., “angry”) or a literal test word (e.g., “useful”). These test words were presented either 150 or 1000 ms after participants read the final statements.

The results showed that when people read less familiar ironies they responded faster to the literal test words than to the ironic test words in the 150 ms condition, but there were no differences in the lexical decision times to the literal and ironic test words after 1000 ms. In contrast, the literal and ironic test words were responded to equally fast after both 150 and 1000 ms when people read familiar ironies. This pattern of data suggests that when people read familiar ironies both literal and ironic meanings are quickly accessed, but only literal meanings are initially activated when people read less familiar ironic statements. Although Giora and Fein (1999) favor a salience-first processing model, as opposed to the standard pragmatic account, their results support the idea that salient meanings, of perhaps both words and sentences, are always accessed first. In this way, the graded salience view is similar to modular views of linguistic processing in which context operates to narrow appropriate meaning after some initial context-independent word and phrase meanings have been activated.

One difficulty with the graded salience view is that is unclear what defines a word’s, or expression’s, salient meaning. Giora (2002) suggests, “The salient sense of a word, or an expression, is the one directly computable from the mental lexicon irrespective of inferences drawn on the basis of contextual information” (P. 18). Salience is a graded notion, and includes senses that are more frequent, conventional, or prototypical/stereotypical.

The best empirical method for assessing the salient meaning of any word is to use standardized norms such as word frequency and word familiarity, although these alone do not necessarily indicate which of several alternative senses of a word are most salient. Ordinary speakers can, however, be asked to judge the frequency or familiarity of alternative word senses to obtain a measure of salience. Giora (2002) also suggests other behavioral tasks may be employed to assess salient meanings such as asking people to write down the meanings of words, or phrases, that “came to mind first” (p. 22), or to provide speeded responses to probes related and unrelated to words placed in neutral contexts. In general, however, it is not clear that these different methods all lead to the same salient meaning for individual words and phrases, even for a single person.

A different problem with the graded salience view is that it posits automatic activation of both salient word and phrase meanings. The motivation for this facet of the proposal comes from the fact that the conventional meanings of certain phrases, such as “kick the bucket” (meaning “to die”), are automatically activated even when the context specifies a different interpretation (e.g., a dairy farmer striking his foot against a pail). Yet according to the graded salience hypothesis, the salient meanings of individual words should also be automatically activated regardless of context. Thus, the salient meaning of the word “kick” should be quickly accessed. But this salient word meaning differs from the putative salient meaning of the entire phrase (e.g., “to die”). It is unclear how this conflict is resolved or whether context comes into play to determine contextually appropriate word meanings before conventionalized phrasal meanings are accessed.

A related recent theory of figurative language processing claims that the language processor initially accesses an interpretation that is compatible with both a word’s literal and figurative meanings (Frisson & Pickering, 2001). Consider the verb “disarmed” in “Mrs. Graham is quite certain that they disarmed about every critic who was opposed to spending more money on art.” The “underspecification model” assumes, for example, that the initial meaning recovered when reading the verb “disarmed” in any context is underspecified as to whether it refers to removing literal or figurative arms. Over time, however, the language processor uses context to hone in on the word’s appropriate meaning, where the honing in process is faster when the preceding context is strong and slower when the preceding context is neutral.

Support for the underspecification model comes from several eye-movement studies. In one study, Frisson and Pickering (2001) examined people’s processing of ambiguous verbs, such as “disarmed” in the above sentence. The eye-movement data showed that the processing difficulty with the subordinate sense of “disarmed,” relative to when the word was used in a literal, dominant sense (e.g., “After the capture of the village, we disarmed about every rebel and sent them to prison”), did not emerge until *after* the critical verb was read. Thus, context reduces processing difficulty, but the difference did not emerge until much after the verb was seen. Frisson and Pickering suggest that people did not initially access either a specific sense or several senses for an ambiguous verb. Instead, readers initially recovered a general, underspecified meaning for the verb and then created a further concrete instantiation of its meaning later on. According to the underspecification model,



then, context does not operate to judge between different word meanings, but functions to change an underspecified, or highly general meaning, into a specific interpretation.

A different set of studies in support of underspecification investigated processing of sentences containing place-for-institution metonymies such as “That blasphemous woman had to answer to the convent” by measuring participants’ eye-movements as they read (Frisson & Pickering, 1999). Results showed that people were as fast to understand these familiar metonymies as to read literal sentences, and that processing unfamiliar metaphors took more time than did reading compatible literal sentences. Thus, figurative language processing need not be delayed for familiar metonymies. A second study showed similar findings for familiar place-for-event metonymies such as “A lot of Americans protested during Vietnam,” and unfamiliar ones such as “A lot of Americans protested during Finland.” Frisson and Pickering argued that the overall findings do not support either a literal-first or figurative-first model, but fit best with a model where a single underspecified representation that is compatible with both literal and figurative (e.g., metonymic) senses. Eventually, context comes in to hone the very general interpretation into a contextually appropriate meaning.

The underspecification model does not assume that different linguistic processes must exist for different meaning products (i.e., literal vs. figurative uses of words) to arise during on-line linguistic understanding. In this way, the putative distinction between literal and figurative senses of a word is irrelevant, at least in terms of ordinary processing. However, similar to the graded salience model, the underspecification model embraces a modular view of linguistic processing, at least in the sense that lexical access is encapsulated from contextual effects. But similar to the graded salience view, the underspecification model suffers from the problem of not being able to specify what constitutes the initial, underspecified meaning that is accessed when a word is first encountered. Many linguists reject the underspecification view precisely because they have failed to discover senses that are rich enough to capture the wide range of meanings (up to 100 for some polysemous words) many words possess (Gibbs, 1994). More generally, both the graded salience and underspecification views face the challenge of demonstrating consistent bottom-up activation of context-free word meanings even in the presence of strong supporting context.

Finally, a different model of figurative language understanding embraces the notion of constraint satisfaction, an idea that has gained much support in psycholinguistics and cognitive science (Katz & Ferratti, 2001; Katz, 2005). When people comprehend a text, or a figurative utterance, they must construct an interpretation that fits the available information (including context) better than alternative interpretations. The best interpretation is one that offers the most coherent account of what people are communicating, which includes meanings that best fits with certain other information and excludes meanings that do not fit this other information. Under this view, understanding a figurative utterance requires people to consider different linguistic and nonlinguistic information that best fits together to make sense of what a speaker or writer is saying. Constraint satisfaction models are computationally efficient, and perhaps psychologically plausible, ways of showing how different information is considered and integrated in everyday cognition.

Katz and Ferretti (2001) argue that a “constraint satisfaction model” provides the best explanation for experimental data on proverb understanding. They employed a self-paced moving window paradigm to show that context affects people’s immediate reading of familiar (e.g., “Lightning never strikes the same place twice”) and unfamiliar proverbs (e.g., “Straight trees have crooked roots”) that have both well-formed literal and figurative meanings. Familiar proverbs were understood more easily than unfamiliar expressions, and the speed-up in processing for familiar proverbs occurred as soon as the second word of the expression was read. But the first words of unfamiliar proverbs were read more quickly in contexts supporting their figurative, rather than literal, meanings. Yet the analysis of an unfamiliar proverb’s figurative meaning was not always complete when the last word was read.

These findings support a constraint satisfaction model by positing how different sources of information (i.e., syntactic, lexical, conceptual) compete for activation over time in parallel. Constraints interact to provide probabilistic evidence in support of various alternatives with the competition ending when one alternative fits best. For example, when reading an unfamiliar proverb, people immediately focus on a literal interpretation because there is less competition from other sources of information supporting a figurative meaning. Similarly, familiar proverbs are easier to process than unfamiliar expressions because there is more information available from the context and the words in familiar proverbs to support a figurative interpretation.

Another test of the constraint satisfaction view examined people’s immediate understanding of expressions like “Children are precious gems” as having metaphoric (children are valuable) or ironic (children are burdens) meaning (Pexman, Ferretti, & Katz, 2000). Several sources of information could induce either the metaphoric or ironic meaning, including the occupation of the speaker, whether the statement was counterfactual to information in the previous discourse, and the familiarity of the expression. Results from an on-line reading task (i.e., moving window) demonstrated that the “A is B” statements were initially read as metaphors, but that the speaker’s occupation and counterfactuality of the statement given the previous context play an early role in processing, thus slowing processing at the space following the statement or by the time the first word of the next statement is read. Furthermore, knowing that a speaker is often associated with irony slows down reading of the first word in the following statement if the context leads one to expect a metaphoric reading, yet acts immediately to speed up processing right after the target statement if the context induces an ironic meaning. The complex interaction between the three sources of information is consistent with the idea that understanding whether an expression is meant metaphorically or ironically depends, similar to other aspects of language, on multiple sources of information being examined and interpreted continuously during on-line reading (McRae, Spivey-Knowlton, & Tannenhaus, 1998).

Related findings using a moving window paradigm showed that context modulated relative processing of literal and ironic statements (Ivanko & Pexman, 2003). When context induced neither a literal or ironic bias, reading times for literal and ironic utterances were roughly equivalent, with faster reading times more locally for the fifth word of the target statements. When the context led to a bias for literal criticism, ironic remarks were

read more slowly than literal ones. Once again, there are complex interactions between the type of context and the speed with which figurative utterances are understood, such that literal readings of utterances, or salient ones, are not obligatory in all cases. This pattern is most consistent with probabilistic, constrain-satisfaction models of figurative use, and is inconsistent with modular approaches to linguistic processing.

#### 4. INDETERMINACY OF FIGURATIVE MEANING AND PROCESSING

The important emphasis on on-line-processing figurative language in experimental psycholinguistics often ignores exactly what people have understood when they seem to have successfully comprehended a particular figurative expression. For the most part, psycholinguists and others tacitly assume that any figurative statement can be paraphrased by a linguistic expression that states in literal terms what people must have attempted to communicate when speaking figuratively (e.g., “blow your stack” means “to get very angry”). The reduction of figurative meaning to simple, short linguistic paraphrases in psycholinguistics is reasonable in the context of designing experimental studies that, for instance, contrast figurative language processing with nonfigurative, or sometimes literal, understanding.

However, the belief that figurative expressions can be readily paraphrased misconstrues the complexities of what many figurative expressions actually communicate in real-life contexts. Studies show, for example, that when people read “John blew his stack,” they readily infer information about the cause, intentionality, and manner by which John got angry, inferences they did not draw when they read literal paraphrases of similar length such as “John got very angry” (Gibbs, 1992). Furthermore, reading idioms in contexts that violate any of these inferences slows down processing for these phrases, but not so for literal paraphrases (Gibbs, Bogdonovich, Sykes, & Barr, 1997). These empirical findings strongly suggest that even highly conventional metaphors, which are often incorrectly assumed to be dead metaphors or long lexical items, convey rich conceptual and pragmatic information, more so than do so-called literal paraphrases.

This conclusion about the richness of figurative meaning comes as no surprise to many interdisciplinary metaphor scholars who have long argued that metaphor, and many other tropes, are “pregnant with meaning.” In fact, a large body of research has discovered that different forms of figurative language communicate a wide variety of propositional, social, and affective meanings, or pragmatic effects. Verbal irony, for instance (e.g., saying “This is fantastic” when losing one’s keys), has been shown to both enhance and diminish the condemnation expressed by an individual relative to speaking more directly (Colston, 1997). By saying something positive about a negative situation, the situation is made to look worse relative to saying something directly negative (e.g., “This is just awful”), which enhances the speaker’s condemnation. Verbal irony, along with hyperbole (e.g., “He was so hungry he ate the table with his meal”), and understatement (e.g., “This might require a bit of work” about a huge task), also express predictably variable degrees of humor, expectation/reality deviance demonstration and speaker protection (Colston & O’Brien, 2000a, 2000b; Colston, 2002). Hyperbole expresses surprise either through

increasing hearers' attention toward expectation/reality discrepancies via the distinctiveness of the inflation, or by an audacity demonstration process whereby a speaker breaks with conversational congruity to make a point (Colston, in press; Colston & Keller, 1998).

One form of ironic discourse, called rebuttal or ironic analogy, performs the dual pragmatic functions of argument and social attack (Whaley & Holloway, 1996; Colston & Gibbs, 1998; Colston, 1999, 2000a). So if a speaker says, for instance, "calling Chili's just another restaurant is like saying the Great Wall of China is just a fence," she causes the hearer to map the ironic structure of the base, "saying the Great Wall of China is just a fence", onto the target, "calling Chili's just another restaurant." This acts to argue against the proposition in the target, as well as to belittle the proponent(s) of that proposition.

Many forms of figurative language also bolster persuasiveness and the social standing of speakers (Holtgraves, 2001; Sopory & Dillard, 2002). A "truth externalization" process is particularly well performed by proverbs, metaphors and some idioms, for instance. Thus, a speaker who claims, "It is best to let sleeping dogs lie" relies on a cultural norm as expressed in the proverb to convey the best course of action in a potentially difficult situation (Gibbs, 2001; Curco, 2005). By using language that leverages a significant degree of meaning outside of a proposition directly proclaimed by a speaker, the speaker places the "truth" of the intended message outside of him/herself. This lending of objectivity can make the meaning seem stronger. A speaker's demonstrated skill in shepherding the intended message of a figurative utterance can also increase others' admiration, which can in turn additionally enhance the message. These and other similar processes can contribute to a more general "mastery demonstration" function where a figurative speaker can gain in their social standing by using figurative language (Gibbs & Izett, 2005). Indeed, many people have a positive subjective experience when they comprehend figurative language (for a review, see Colston, in press). Such a positive feeling can reflect well on a speaker and lead to many sociocognitive and persuasive effects (e.g., liking the speaker more, paying greater attention to what the speaker says subsequently, more strongly adopting the speaker's viewpoint, etc.) that can cascade and contribute to other of the effects discussed here.

Finally, research shows that some forms of figurative language evoke different kinds of emotional reactions. Thus, hearing ironic statements leads listeners to feel more intense emotions than when literal speech is heard (Leggitt & Gibbs, 2000). Sarcasm, rhetorical questions, and overstatements all evoke relatively negative emotions, compared to understatement and satire. People also tend to speak metaphorically more so when feeling intense emotions, something that listeners readily pick up on in many conversations and attribute affective meanings to speakers' messages (Gibbs, Leggitt, & Turner, 2003). One large study demonstrated, more generally, that different types of figurative language can fulfill as many as 20 different discourse goals, including many of the social and affective effects described here (Roberts & Kreuz, 1994).

Not surprisingly, inferring pragmatic effects may come at a processing cost. But what determines the stopping point for the various indeterminate aspects of figurative

meanings? One well-known proposal from linguistic pragmatics suggests that there is a trade-off between the amount of cognitive effort put into linguistic understanding and the cognitive effects, or meanings, that are inferred (Sperber & Wilson, 1995), a theory that applies to all aspects of linguistic communication, not just figurative language processing. Relevance theory generally claims, again, that interpretation of figurative language occurs in the same way as with any other nonfigurative utterance. A listener stops processing when he thinks that every further implication he could get is not worth the effort it takes to obtain these additional cognitive effects. Sperber and Wilson (1995) claim that newly presented information is relevant in a context only when it achieves cognitive effects in that context, and the greater the cognitive effects, the greater the relevance. They specifically define a notion of “optimal relevance” that outlines what listeners look for in terms of cognitive effort and effect: an utterance, on a given interpretation, is optimally relevant if and only if (a) it achieves enough effects to be worth the listener’s attention, and (b) it puts the listeners to no gratuitous effort in achieving these effects.

Consider the metaphorical utterance “My surgeon is a butcher.” Listeners generally have immediate access to stereotypical knowledge about surgeons and would normally infer that the speaker here means, “My surgeon is crude and sloppy in his practice.” Speaking loosely like this requires that speakers have in mind some further idea or cognitive effect beyond the single thought “My surgeon is crude and sloppy in his practice.” For instance, the speaker may wish to convey an image of surgeons that is beyond most people’s experience and will expect the listener to make some effort toward exploring a wide range of cognitive effects (e.g., having to do with the nature of surgeons, their imprecision, their insensitivity toward dealing with human beings, and perhaps their appearance and demeanor). These implications are relatively weak, but they best resemble the speaker’s thought about his surgeon. Understanding this range of weak implications may require additional cognitive effort on the part of the listener, but this is offset, according to the principle of relevance, by extra effects not achievable by saying directly “My surgeon is crude and sloppy in his practice.” In general, metaphorical utterances, like all figurative and indirect language, are simply one means of optimizing relevance in verbal communication.

Very few psycholinguistic studies have explicitly explored the trade-off between cognitive effort and effects in figurative language processing. One study suggests that there must be extra processing associated with understanding a well-chosen metaphor (Noveck, Bianco, & Castry, 2001). Yet it is not clear how to operationalize the idea of individual metaphorical, or figurative, meanings within the “more cognitive processing = more cognitive effects” hypothesis. Consider the stock metaphor “Some jobs are jails.” There are a variety of meanings that people may understand when reading this expression, including that some jobs are poorly paid, confining, stifling, unpleasant, demoralizing and so on. But how does one actually distinguish between these impressions to clearly establish which meanings are independent in order to test the idea that more cognitive processing equals more cognitive effects? This problem is complicated by the possibility, as noted above, that listeners may draw a range of pragmatic effects, or

weak implicatures, from figurative utterances. There are also cases where people can put a good deal of cognitive effort into understanding a speaker's utterance without gaining appropriate cognitive effects. For example, where people assume that the producer of a metaphor was a famous poet, they put in a good deal more effort to try to understand anomalous phrases, such as "A scalpel is like a horseshoe," than when told that these phrases were randomly generated by a computer program (Gibbs, Kushner, & Mills, 1990). Finally, people may also infer complex figurative meanings with little cognitive effort, or at least less time than is needed to comprehend corresponding nonfigurative expressions, as shown by Gibbs (1992).

In general, it is impossible to predict the processing effort needed to comprehend figurative utterances given the number, or types of cognitive effects that may arise from interpreting these statements. It may be the case, as Noveck et al. (2001) argue, that some figurative expressions, such as certain novel metaphors, may take longer to process than synonymous nonfigurative expressions, if these are encountered in neutral contexts, precisely because of the additional cognitive effects they communicate. But proving this point will require an independent measure of the cognitive effects that utterances convey. We see this as one of the great challenges for figurative language scholars, as well as for all psycholinguists.

## **5. EXAMINING THE COGNITIVE EFFORT AND EFFECTS TRADE-OFF**

Despite some of the difficult questions regarding the nature of cognitive effects, we believe that the time is ripe for psycholinguistic research on the trade-off between effort and effects during figurative language processing. Our claim here is that present debates over whether figurative language is understood directly or indirectly should evolve into a more systematic examination of the complex interactions between many cognitive and linguistic factors associated with any psycholinguistic act. One way to begin this type of exploration is to adopt an old tetrahedral model of cognitive processes (Jenkins, 1979), which suggests that several factors must shape processing, including (1) the participants (e.g., their abilities, interests, beliefs, motivations, goals), (2) the understanding task (e.g., understanding to solve a problem, make a decision, remember something, be emotionally affected by something said), (3) the criterial task (e.g., different measures of cognitive processes and product), and (4) the materials (e.g., type of language, modality of presentation). Fortunately, there is a fair amount of research relevant to each of these factors in regard to figurative language use, even if at present these findings have not been placed within a larger theoretical framework.

### **5.1. Participants**

There are a variety of participant variables that can influence processing fluency for figurative language. For example, if a speaker is known to be the type of person who regularly uses verbal irony, based on occupation or gender for instance, and if the situation has been set up to likely create ironically intended utterances, then ironic utterances

will be comprehended relatively fluently, as several studies have demonstrated (Katz & Pexman, 1997; Pexman & Olineck, 2002; Katz, Piasecka, & Toplack, 2001). Gender is another important variable, given research showing that men tend to use figurative language in describing other people's emotions, while women use figurative speech more in talking about their own feelings (e.g., "I would feel like my heart would just jump out of my chest . . .") (Link & Kreuz, 2005). The relationship between speakers (e.g., close friends vs. strangers), their social status, occupation, geographic origin, religious or political background, ethnicity, and personalities, show effects on comprehension of a diverse range of figurative forms, including metaphor, irony, metonymy, proverbs, idioms, indirect requests, analogies, litotes, and metaphorical gestures (Colston & Katz, 2005). To note just a couple of examples, consider the normal ironic banter that often accompanies groups of friends (Gibbs, 2000; Pexman & Zvaigzne, 2004) or the quintessential British form of understatement (e.g., "He clearly has issues" in reference to a suicidal character).

Another emerging characteristic of participants that has been shown to influence figurative language processing is their past and current embodied experiences (Gibbs, 2005). For example, research shows that people's previous bodily experiences of hunger partly predicts their use and understanding of metaphorical expressions about difference forms of desire, as seen in statements like "I hunger for fame" or "I craved her affection" (Gibbs, Lima, & Francuzo, 2004). In another series of studies on metaphorical talk about time, students waiting in line at a café were given the statement "Next Wednesday's meeting has been moved forward two days" and then asked "What day is the meeting that has been rescheduled?" (Boroditsky & Ramscar, 2002). Students who were farther along in the line (i.e., who had thus very recently experienced more forward spatial motion) were more likely to say that the meeting had been moved to Friday. Similarly, people riding a train were presented the same ambiguous statement and question about the rescheduled meeting. Passengers who were at the end of their journeys reported that the meeting was moved to Friday significantly more than did people in the middle of their journeys. Although both groups of passengers were experiencing the same physical experience of sitting in a moving train, they thought differently about their journey and consequently responded differently to the rescheduled meeting question. These results suggest how ongoing sensorimotor experience has an influence on people's comprehension of metaphorical statements about time.

## 5.2. Orienting Task

In many communicative settings, people are not given, nor are they following any explicit directions or rules that might orient them to process or comprehend what is said or written in a particular way. Exceptions to this would be in those occasional situations where orienting rules or directions have been given or are being followed, (e.g., a law clerk is told to scan through court transcripts looking for when a defendant said...). People would, though, likely adhere to implicit rules that might affect figurative language comprehension and processing (c.f., eavesdropping vs. listening to poetry).

For example, asking people to solve problems with metaphoric language, simply understand metaphors, recall metaphors, or produce metaphor may all produce varying empirical findings in regard to the relative primacy of metaphor to nonmetaphoric language. Thus, research shows that people's decisions about common dilemmas are strongly shaped by the presence or absence of metaphor (Robins & Mayer, 2000). When a metaphor is critical to frame, or understand, a problem, people readily use that information in making decisions about a common dilemma. But when metaphor is not necessary to understanding a dilemma, the presence of such language adds ambiguity to people's decision-making process. On the other hand, studies also show that the persuasiveness of metaphor depends more on the way such language resonates with a person's own preferences than it does with whether the metaphor is needed to frame a topic (Ottati, Rhoads, & Graesser, 1999).

In a different context, although people clearly understand familiar metaphors faster than they do unfamiliar ones, they recall these two types equally well (Blasko & Brihl, 1997). Finally, asking people to verbally describe the conceptual connection between two word often yields metaphoric descriptions, which take longer to produce than do nonmetaphoric descriptions (Flor & Hadar, 2005), a result that is contrary to the typical pattern in reading time studies showing that metaphors do not generally take more time to process than nonmetaphorical statements. Once again, the orienting task given to experimental participants can have an important effect on whether figurative language is seen as primary compared to nonfigurative speech.

At the same time, orienting tasks can powerfully adjust the fulcrum location upon which language comprehension tips toward the more figurative or less figurative. For instance, when operating under the criterion of achieving genuine or deep understanding, listeners/readers can use figurative comprehension to fertilize rich interpretation. Many heady, moving experiences of language comprehension are evidence of this (e.g., hearing powerful speeches, emotional song lyrics or poetry, highly apt metaphors or other figures). Conversely, when listening/reading for less cooperative and indeed, combative purposes (e.g., as in arguments, debates), people will often scramble for the golden fleece of a "literal meaning," to serve those purposes (e.g., for preparing rebuttals, to find weaknesses or attack points in others' comments). One study for instance found evidence for such a link between criterial task and degree of "literal mindedness." When people were placed in high-stress situations, as are often the case in arguments, disagreements, debates, etc., their ability to comprehend figurative language subtly broke down (Colston, 2005b). Barr and Keysar (2005) also argue that people tend to interpret figurative expressions egocentrically, and thus do not take common ground information immediately into account, when they are under time pressure.

### 5.3. Criterial Task

The best method for assessing comprehension or interpretation of figurative language has always been a significant source of concern in psycholinguistics. Tasks that have used off-line measures (e.g., rating studies, judgments of metaphor aptness, memory tasks) as



indicators of figurative language comprehension have often been criticized for their inability to distinguish processes that might take place during reading or hearing in the comprehension process vs. those that might occur later in the processing stream. Reading time measures were long considered superior because they could use overall reading time as a more precise, and presumably outside of subjective control, indicator of on-line processing – relying upon the assumption that, all else being equal, longer reading times indicated greater processing. But reading time studies also differ in their specific task requirements. Some experiments ask participants to simply read individual sentences in a story, and push a button as soon as the expression on the computer screen has been understood. Yet other studies ask people to sometimes read an expression, such as a figurative remark, and make a speeded judgment as to whether it fits within the preceding story context. As it turns out, judgments of appropriateness or relatedness often result in longer comprehension times for figurative expressions compared to literal ones (Schwobel et al., 2001; Temple & Honeck, 1997). But figurative and literal expressions can be read equally fast when only simple comprehension time is measured (Gibbs, 2002). Thus, the precise task used leads to different results with very different theoretical implications.

More recently, more sophisticated mechanisms have enabled word-to-word reading time measures with moving scanning windows that readers control, and eye-tracking measures that remove unnatural reader responses altogether from the reading/measurement. This progression in research methodologies has been viewed as an improvement in our ability to tap into figurative language processing, and undoubtedly it has afforded greater precision. But often overlooked in this perspective is the potentially remaining disconnect between even the very precise eye-movement measures and what one genuinely and subjectively experiences as comprehension, as if that is ever even a delineated, all-or-none accomplishment, universal across all kinds of language, interlocutor types, goal-situations, etc. (but see Rayner and Pollatsek, this volume for a different perspective). Eye movements, although rich in their potential, may not necessarily be deterministically related to states of comprehension in completely reliable ways (e.g., a reader may pause and stare at a random word while processing some text that is largely irrelevant to that word). Moreover, the problem that text comprehension, spoken language comprehension from say anonymous audio recordings, and genuine conversational comprehension with known interlocutors are very different things is mostly overlooked given the primary emphasis on appropriate dependent measures. One possible solution would be to further increase the sophistication of “comprehension” measures, such as combining emotional response indicators, eye-movement trackers, video facial expression recordings and other measures in a linked time course measurement (Colston, 2005a).

Within interdisciplinary discussions of figurative language, scholars from fields outside of psycholinguistics often see psychological experiments as being rather distant from their own concerns with the deep, meaningful, complex interpretation of different forms of figurative discourse (e.g., metaphor, irony, metonymy). Psycholinguists’ primary interest in immediate, fast, mostly unconscious mental processing of figurative speech ignores slower, more reflective aspects of linguistic interpretation. Gibbs (1994)

suggested that figurative language understanding does not constitute a single event, or moment in time, but can exist along a continuum of temporal processing ranging from fast comprehension, slower interpretation, nonobligatory recognition (i.e., that statement I just heard was ironic) to reflective appreciation. In the past, scholars have mistakenly made theoretical claims about fast processing from slower, and consciously held, interpretations and appreciations, while psycholinguists, again, have mostly neglected the rigorous study of cognitive effects, or the products of figurative language understanding.

Yet once more, paying systematic attention to cognitive effects, both those that arise immediately during fast comprehension, and those that emerge more slowly during reading (and re-reading!) is critical to creating more comprehensive theories of figurative language use in different real world, communicative contexts.

#### 5.4. Materials

There are several aspects of the materials that have been shown to have a strong impact on figurative language use and processing. First, as described above, the conventionality of a figurative expression plays a major role in the way it is processed. But conventionality is not a single dimension, given that several factors contribute to the impression that some utterance is conventional or novel, including its grammatical form, frequency in the language, appropriateness to the specific context, and appropriateness for the speaker. For example, consider verbal irony. Conventional ironies will often take the form of rhetorical or tag questions (e.g., “wasn’t that brilliant?”, “that was brilliant wasn’t it?”). As mentioned earlier, conventional ironies also often contain noun or verb modifiers, usually intensifiers (e.g., simply, utterly, just, absolutely, etc.). With regard to semantics, a highly conventional ironic pattern is to use utterances with positive meanings, usually to comment about negative situations. Other semantic conventions in verbal irony are to express agreement (e.g., “yup,” “uh huh,” “sure,” etc.), to understate (e.g., “it seems to be snowing” said during a blizzard), to exaggerate (e.g., “her two foot tall husband,” said about a short man) or to express the semantic content that was predicted or that might be expected in the situation (e.g., “soccer is an ‘easy’ game”), often when the situation has not gone as expected (e.g., the game is difficult). For example, the ironic phrase, “wise guy,” is highly conventional in some American-English speaking communities because it is almost never used directly to state that a person is intelligent. Other utterances might have both ironic and literal conventional meanings. Examples here, again in some American-English speaking communities, are “oh, sure,” “I’m sure,” etc. Lastly, still other kinds of utterances are not at all conventionally ironic but might be used ironically in a given particular conversation.

The prototypicality of the material is also an important factor. The general pattern of results is that many kinds of processing (e.g., recognizing, recalling, reading, comprehending, etc.) are more readily accomplished to the degree that the target material they work upon is more prototypical. So, for instance, prototypical items of furniture (e.g., chair) are more easily and quickly recalled than less prototypical ones (e.g., lamp).

More prototypical forms of syntactic structure (e.g., active) are more easily read than less prototypical ones (e.g., passive), etc. For verbal irony then, one can readily predict that more prototypical forms (e.g., positive semantic content, intensifying modifier, tag question) will be more fluently processed than less prototypical ones (e.g., negative semantic content, etc.). Indeed, the particular finding in this example has been born out by research (Gibbs, 1986a, 1986b, Kruez & Glucksberg, 1989; Utsumi, 2000).

To demonstrate this prototypicality influence, recall the earlier discussion on semantic conventions for verbal irony. Often ironic utterances will contain positive semantic content, but they will be used to comment about a negative situation (e.g., saying or writing, "Excellent, just what I needed," to complain about an unexpected extra workload). This frequently observed pattern of verbal irony illustrates a semantic contrast, which is a major characteristic, and indeed necessary condition, for verbal irony (Colston, 2000b). Ironic utterances that present strong contrasts between expectations and reality are funnier, more criticizing, less self protective, more expressive of surprise, etc., than verbal ironies that provide relatively weak contrasts (Colston & Keller, 1998; Colston & O'Brien 2000a, 2000b; Gerrig & Goldvarg, 2000). These pragmatic functions are not direct measures of comprehension fluency but they do indicate the power of the contrast variable on the expressiveness of the irony and strongly suggest a processing fluency difference between strong and weak contrasts. One might thus predict that, as usual with all else being held constant, the greater the contrast between expectations and reality, the easier verbal irony processing might be.

One major influence on processing fluency that is idiosyncratic to verbal irony is the clarity with which a mention, echo, reminder, or allusion to prior predictions or expectations that have been violated by occurring events can be achieved (Sperber, 1984; Kreuz & Glucksberg, 1989; Kumon-Nakamura, Glucksberg, & Brown, 1995). Indeed such a contrast between the semantic content of the utterance and its referent situation is in fact a hallmark of the ironic figurative form (Colston, 2000b).

Metaphoric language also builds off from previous discourse structures that can quickly lead readers to metaphoric, as opposed to literal interpretations (Keysar, 1994), and in some instances are specific to this form of speech. Thus, contexts that describe topics in metaphorical ways make it easier to infer subsequent metaphoric utterances when the underlying conceptual metaphors are similar (Albritton, McKoon, & Gerrig, 1995), and more difficult to process when a new metaphorical utterance is based on a different conceptual metaphor (Langston, 2002). Although the vast amount of work on figurative language comprehension examines interpretation of a single utterance after a nonfigurative context, it is evident that different figurative contexts, and previously spoken figurative utterances, have a strong effect of on-line figurative language comprehension. This is one topic that demands further attention.

Finally, speakers use a wide variety of metalinguistic devices to indicate figurative intent. The phrase "strictly speaking" often accompanies metaphorical expressions (Goddard, 2004), and studies show that the presence of markers like "proverbially speaking" facilitates

people's comprehension of proverbs (Katz & Ferratti, 2003). One direct way to affect the comprehension fluency of verbal irony is by providing or omitting markers for the irony (Bryant & Fox Tree, in press). These markers can be controlled by a speaker or supplied in written context. For instance, nonverbal indicators of irony (e.g., gestures, facial expressions, etc.) can be performed or described (e.g., "she rolled her eyes and said"). Intonational patterns (e.g., nasal pronunciation, elongated phonemes, exaggerated pitch magnitudes, etc.), can be used or mentioned (e.g., "in a mocking tone he said"), etc. (Kreuz & Roberts, 1995). Speakers can also make simultaneous contradictory expressions to achieve expressional irony. For example a speaker could say, "Oh, yeah, he is brilliant," while making a gesture of tracing a circle around their ear, as if expressing that someone is crazy, and achieve an ironic perlocution. These factors would also contribute to the fluency of utterance processing, again to the extent that the speaker successfully uses them, that they are readily incorporated by the hearer, etc. In general, one might expect that the presence of such markers in the verbal and written situations would aid processing relative to their absence, again assuming equivalence of other influences.

### 5.5. Summary

Determining the constraints that shape the trade-off between cognitive effort and effects during figurative language understanding requires, in our view, the empirical study of how these various factors have their individual effects, and very likely interact in complex, even nonlinear, ways. The goals and motivations of individual speakers and listeners surely shape the cognitive effort expended when producing and understanding figurative language, and the specific orienting perspective, the methods used for assessing understanding, and the types of materials employed clearly determine the cognitive effects drawn in linguistic communication. Our hope is that specific recognition of these factors will enhance attempts to create a more complete picture of the trade-off between cognitive effort and effects in figurative language use.

## 6. CONCLUSION

The complexities of figurative language processing are such that there may not be a single theory or model that explains how all aspects of figurative language are understood. Part of the reason for this conclusion is that figurative language does not constitute a homogenous kind of language that is necessarily used and understood in completely distinct ways from nonfigurative, or what some call "literal" speech. Of course, one message of this chapter is that it makes little sense to suggest theories of figurative language understanding, as different from "normal" discourse comprehension, unless there is a well-developed, and consistently applied, theory of literal language and meaning. Given the long history to provide a theory of literal meaning (Gibbs, 1994), and the failure to come up with a unified account of this kind of language, we frankly are doubtful whether any such proposal will come forward that is widely embraced by psychologists, linguists, and philosophers.

None of this implies that different aspects of figurative language have no special features, both in terms of the cognitive processes involved (e.g., cross-domain mappings for metaphor, determining the source of echos for irony, inferring part to whole relationships with metonymy) and the meaning products that arise from interpretive processes. We have argued that the study of both cognitive processes and effects, or products, is critical to future theoretical work on figurative language, and that exploring the real-time trade-off between effort and effects is one specific direction for new experimental studies. In this way, figurative language research should provide another arena within psycholinguistics more generally where the traditions of language as product and language as action perspectives may be bridged.

## REFERENCES

- Albritton, D., McKoon, G., & Gerrig, R. (1995). Metaphor-based schemas and text comprehension: Making connections through conceptual metaphor. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 612–625.
- Barr, D. J., & Keysar, B. (2005). Making sense of how we make sense: The paradox of egocentrism in language use. In: H. Colston, & A. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 21–42). Mahwah, NJ: Erlbaum.
- Blank, G. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity*, *3*, 21–36.
- Blasko, D., & Brihl, D. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol*, *12*, 261–285.
- Blasko, D., & Connine, C. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 295–308.
- Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*, 185–189.
- Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*, 192–216.
- Brisard, F., Frisson, S., & Sandra, D. (2001). Processing unfamiliar metaphors in a self-paced reading task. *Metaphor and Symbol*, *16*, 87–108.
- Bryant, G., & Fox Tree, J. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*, 257–277.
- Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticism. *Discourse Processes*, *23*, 25–45.
- Colston, H. L. (1999). The pragmatic functions of rebuttal analogy. *Metaphor and Symbol*, *14*, 259–280.
- Colston, H. L. (2000a). Comprehending speaker intent in rebuttal analogy use: The role of irony mapping, absurdity comparison and argumentative convention. *Language and Speech*, *43*, 337–354.

- Colston, H. L. (2000b). On necessary conditions for verbal irony comprehension. *Pragmatics and Cognition*, 8, 277–324.
- Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics*, 34, 111–142.
- Colston, H. L. (2005a). On sociocultural and nonliteral: A synopsis and a prophesy. In: H. Colston, & A. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 1–20). Mahwah, NJ: Erlbaum.
- Colston, H. L. (2005b). Social and cultural influences on figurative and indirect language. In: H. Colston, & A. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 99–130). Mahwah, NJ: Erlbaum.
- Colston, H. L. (in press). Irony, analogy and truth. In: B. Nerlich, & A. Burkhardt (Eds), *Tropical truth*. Amsterdam: John Benjamins.
- Colston, H. L., & Gibbs, R. W. (1998). Analogy and irony: Rebuttal to “Rebuttal Analogy.” *Metaphor and Symbol*, 13, 69–75.
- Colston, H. L., & Gibbs, R. W. (2002). Are irony and metaphor understood differently? *Metaphor and Symbol*, 17, 57–80.
- Colston, H., & Katz, A. (Eds). (2005). *Figurative language comprehension: Social and cultural influences*. Mahwah, NJ: Erlbaum.
- Colston, H. L., & Keller, S. B. (1998). You’ll never believe this: Irony and hyperbole in expressing surprise. *Journal of Psycholinguistic Research*, 27, 499–513.
- Colston, H. L., & O’Brien, J. (2000a). Contrast and pragmatics in figurative language: Anything understatement can do, irony can do better. *Journal of Pragmatics*, 32, 1557–1583.
- Colston, H. L., & O’Brien, J. (2000b). Contrast of kind vs. contrast of magnitude: The pragmatic accomplishments of irony and hyperbole. *Discourse Processes*, 30, 179–199.
- Curco, C. (2005). On mosquitoes and camels: Some notes on the interpretation of metaphorically transparent sayings. In: H. Colston, & A. Katz, (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 283–334). Mahwah, NJ: Erlbaum.
- Flor, M., & Hadar, U. (2005). The production of metaphoric expressions in spontaneous speech: A controlled-setting experiment. *Metaphor and Symbol*, 20, 1–34.
- Frisson, S., & Pickering, M. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1366–1383.
- Frisson, S., & Pickering, M. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol*, 16, 149–172.
- Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space – time metaphors. *Language and Cognitive Processes*, 17, 537–565.

- Gerrig, R., & Goldvarg, Y. (2000). Additive effects in the perception of sarcasm: Situational disparity and echoic mention. *Metaphor and Symbol, 15*, 197–208.
- Gibbs, R. W. (1986a). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General, 115*, 3–15.
- Gibbs, R. (1986b). Comprehension and memory for nonliteral utterances: The problem of sarcastic indirect requests. *Acta Psychologica, 62*, 41–57.
- Gibbs, R. (1992). What do idioms really mean? *Journal of Memory and Language, 31*, 485–506.
- Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language and understanding*. Cambridge: Cambridge University Press.
- Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol, 15*(1&2), 5–27.
- Gibbs, R. (2001). Proverbial themes we live by. *Poetics, 29*, 167–188.
- Gibbs, R. (2002). A new look at literal meaning in understanding what speakers say and implicate. *Journal of Pragmatics, 34*, 457–486.
- Gibbs, R. W. (2005). *Embodiment and cognitive science*. New York: Cambridge University Press.
- Gibbs, R., Bogdonovich, J., Sykes, J., & Barr, D. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language, 37*, 141–154.
- Gibbs, R. W., & Izett, C. (2005). Irony as persuasive communication. In: H. Colston, & A. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 131–152). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gibbs, R., Kushner, J., & Mills, R. (1991). Authorial intentions and metaphor comprehension. *Journal of Psycholinguistic Research, 20*, 11–30.
- Gibbs, R., Leggitt, J., & Turner, E. (2003). What's special about figurative language in Emotional communication? In: S. Fussell (Ed.), *Verbal communication of emotions* (pp. 131–164). Mahwah, NJ: Erlbaum.
- Gibbs, R. W., Lima, P., & Francuzo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics, 36*, 1189–1210.
- Giora, R. (2002). *On our mind: Salience, context, and figurative language*. New York: Oxford University Press.
- Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol, 14*, 241–257.
- Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Graded salience and indirect negation. *Metaphor and Symbol, 13*, 83–101.
- Grice, H. (1989). *Studies in the way with words*. Cambridge, MA: Harvard University Press.

- Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. New York: Oxford University Press.
- Goddard, C. (2004). The ethnopragmatics and semantics of 'active metaphors'. *Journal of Pragmatics*, 37, 1211–1230.
- Holtgraves, T. (2001). *Language as social action: Social psychology and language use*. Mahwah, NJ: Erlbaum.
- Ivanko, S., & Pexman, P. (2003). Context incongruity and irony processing. *Discourse Processes*, 35, 241–279.
- Jenkins, J. (1979). Four points to remember: A tetrahedral model of memory experiments. In: L. Cermak, & F. Craik (Eds), *Levels of processing in human memory* (pp. 429–446). Hillsdale, NJ: Erlbaum.
- Katz, A. (2005). Discourse and sociocultural factors in understanding nonliteral language. In: H. Colston, & A. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 1–20). Mahwah, NJ: Lawrence Erlbaum Associates.
- Katz, A., & Ferratti, T. (2001). Moment-by-moment comprehension of proverbs in discourse. *Metaphor and Symbol*, 16, 193–221.
- Katz, A., & Ferratti, T. (2003). Reading proverbs in context: The role of explicit markers. *Discourse Processes*, 36, 19–46.
- Katz, A., & Pexman, P. (1997). Interpreting figurative statements: Speaker occupation can change metaphor to irony. *Metaphor and Symbol*, 12, 19–41.
- Katz, A., Piasecka, I. L., & Toplack, M. (2001). Comprehending the sarcastic comments of males and females. Paper presented at the 42nd annual meeting of the Psychonomic Society, Orlando, FL.
- Keysar, B. (1994). Discourse context effects: Metaphorical and literal interpretations. *Discourse Processes*, 18, 247–269.
- Kreuz, R., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, 118, 374–386.
- Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity*, 10, 21–31.
- Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, 124, 3–121.
- Langston, W. (2002). Violating orientational metaphors slows reading. *Discourse Processes*, 34, 281–310.
- Leggitt, J., & Gibbs, R. W. (2000). Emotional reactions to verbal irony. *Discourse Processes*, 29, 1–24.



- Link, K., & Kreuz, R. (2005). Do men and women differ in their use of nonliteral language when they talk about emotions? In: H. Colston, & A. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 153–181). Mahwah, NJ: Erlbaum.
- McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin and Review*, 6, 486–494.
- McRae, K., Spivey-Knowlton, M., & Tannenhaus, M. (1998). Modeling the influence of thematic fit (and other constraints) in online sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Noveck, I., Bianco, M., & Castry, A. 2001. The costs and benefits of metaphor. *Metaphor and Symbol*, 16, 109–121.
- Ottati, V., Rhoads, S., & Graesser, A. (1999). The effect of metaphor on processing style in a persuasion task: A motivational resonance model. *Journal of Personality and Social Psychology*, 77, 688–697.
- Pexman, P., Ferratti, T., & Katz, A. (2000). Discourse factors that influence on-line reading of metaphor and irony. *Discourse Processes*, 29, 201–222.
- Pexman, P., & Zvaigzne, M., (2004). Does irony go better with friends? *Metaphor and Symbol*, 19, 143–163.
- Pexman, P. M., & Olineck, K. M. (2002). Understanding irony: How do stereotypes cue speaker intent? *Journal of Language and Social Psychology*, 21, 245–274.
- Pfaff, K., Gibbs, R., & Johnson, M. (1997). Metaphor in using and understanding euphemism and dysphemism. *Applied Psycholinguistics*, 18, 59–83.
- Roberts, R., & Kreuz, R. (1994). Why do people use figurative language? *Psychological Science*, 5, 159–163.
- Robins, S., & Mayer, R. (2000). The metaphor framing effect: Metaphorical reasoning about text-based dilemmas. *Discourse Processes*, 30, 57–86.
- Schraw, G. (1995). Components of metaphoric processes. *Journal of Psycholinguistic Research*, 24, 23–38.
- Shinjo, M., & Myers, J. (1987). The role of context in metaphor comprehension. *Journal of Memory and Language*, 26, 226–241.
- Schwobel, J., Dews, S., Winner, E., & Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: Further evidence. *Metaphor and Symbol*, 15(1 & 2), 47–61.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.
- Sperber, D. (1984). Verbal irony: Pretense or echoic mention? *Journal of Experimental Psychology: General*, 113(1), 130–136.

Sopoary, P., & Dillard, J. (2002). The persuasive effects of metaphor: A meta-analysis. *Human Communication Research*, 28, 382–419.

Temple, J., & Honeck, R. (1997). Proverb comprehension: The primacy of literal meaning. *Journal of Psycholinguistic Research*, 28, 41–70.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics*, 32, 1777–1806.

Whaley, B. B., & Holloway, R. L. (1996). 'Rebuttal' analogy: A theoretical note. *Metaphor and Symbolic Activity*, 11(2), 161–167.

This page intentionally left blank

## Chapter 22

### Eye Movements and Spoken Language Comprehension

*Michael K. Tanenhaus and John C. Trueswell*

#### ABSTRACT

This chapter provides an overview of recent research that uses eye movements to investigate spoken language comprehension. We outline the logic of what is now commonly referred to as the “visual world” paradigm and review some of the foundational studies. We then use some sample experiments to review methodological issues, including issues of data analysis, linking hypotheses, and issues that arise when combining language, vision, and action. We conclude with a brief review of some domains within psycholinguistics in which the visual world paradigm is beginning to play a prominent role.

#### 1. INTRODUCTION

Many everyday tasks require people to rapidly interrogate their visual surroundings. Reading a magazine, looking for a friend at a party, and making breakfast, all require people to frequently shift their attention to task-relevant regions of the visual world. These shifts of attention are accompanied by shifts in gaze, accomplished by ballistic eye movements known as *saccades*, which bring the attended region into the central area of the fovea, where visual acuity is greatest. The pattern and timing of saccades, and the resulting fixations, are one of the most widely used response measures in the brain and cognitive sciences, providing important insights into the functional and neural mechanisms underlying attention, perception, and memory (for reviews, see Liversedge & Findlay, 2000; Rayner, 1998). Eye movements are now an important measure in the perception–action literature, especially for studies examining allocation of attention in natural everyday tasks (Hayhoe & Ballard, 2005; Land, 2004). Within psycholinguistics, eye movements have been one of the most widely used response measures in studies of written word recognition and sentence reading for more than two decades, initiated by the classic work of McConkie and Rayner (1976), Frazier and Rayner (1982), and Just and Carpenter (1980). For reviews, see Rayner (1998, this volume).

More recently, eye movements have become a widely used response measure for studying spoken language processing in both adults and children, in situations where participants comprehend and generate utterances that are about a circumscribed “visual world.” Researchers are now using this method to address issues that run the gamut of current topics in language processing. Eye movements are a response measure of choice for studies addressing many classical questions in psycholinguistics, e.g., is the processing of stop consonants categorical (McMurray, Tanenhaus, & Aslin, 2002); does context influence the earliest moments of temporary lexical and syntactic ambiguity resolution (Dahan & Tanenhaus, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002); what is the locus of frequency effects in spoken word recognition (Dahan, Magnuson, & Tanenhaus, 2001a); what factors influence the time course with which anaphoric expressions, such as pronouns, are resolved (Arnold, Eisenbad, Brown-Schmidt, & Trueswell, 2000; Järvikivi, van Gompel, Hyönä, & Bertram, 2005) and, for bilingual speakers, does a word spoken in one language activate the lexical representations of similar sounding words in the other language (Spivey & Marian, 1999; Ju & Luce, 2004).

The visual world paradigm has also opened up relatively uncharted territory in language comprehension, including real-time sentence processing in children (Trueswell, Sekerina, Hill, & Logrip, 1999); the role of common ground in on-line processing (Keysar, Barr, Balin, & Brauner, 2000; Hanna, Tanenhaus, & Trueswell, 2003); how listeners make use of disfluencies in real-time language processing (Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Bailey & Ferreira, 2005; Ferreira & Bailey, *in press*); and how participants in a conversation coordinate their referential domains (Brown-Schmidt, Campana, & Tanenhaus, 2005; Tanenhaus & Brown-Schmidt, *in press*). Finally, the visual world approach has spawned a new family of studies investigating the interface between action and language and between vision and language (Chambers, Tanenhaus & Magnuson, 2004; Spivey et al., 2002; Kamide, Altmann, & Haywood, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005).

Why has the visual world paradigm gained traction so rapidly? First, in contrast to reading, time-locked, relatively natural measures of spoken language processing have been hard to come by. Many of the most widely used tasks for studying spoken language comprehension present only a snapshot of processing at a single point in time, require meta-linguistic judgments, and interrupt the flow of the speech input. In contrast, eye movements provide a sensitive, implicit measure of spoken language processing in which the response is closely time-locked to the input without interrupting the speech stream. Second, the eye-movement paradigm can be used with natural tasks that do not require meta-linguistic judgments. This makes it well suited for studies with young children (Trueswell et al., 1999) and with special populations (Yee, Blumstein, & Sedivy, 2000). Third, the coupling of a visual world with language makes it possible to ask questions about real-time interpretation, especially questions about reference that would be difficult to address, and perhaps would be intractable, if one were limited to measures of processing complexity (e.g., Sedivy, Tanenhaus, Chambers, & Carlson, 1999). It also makes it possible to examine questions at the interface between language, perception, and action (see the chapters by Henderson & Ferreira, 2004 and Trueswell & Tanenhaus, 2005).

Fourth, eye movements can be used to study issues about the relationship between real-time message planning and utterance planning (Bock, Irwin, & Davidson, 2004; Griffin, 2004; Brown-Schmidt & Tanenhaus, in press). Finally, the paradigm allows one to study real-time production and comprehension simultaneously in natural tasks involving conversational interaction. This makes it possible to bridge the two dominant traditions in language-processing research: the “language-as-action” tradition, which has focused on natural interactive conversation while generally ignoring questions about the time course of real-time language processing and the “language-as-product” tradition, which has focused on the time course of processing while being primarily limited to “de-contextualized language” (Clark, 1992; Tanenhaus & Trueswell, 2005).

Our goal in this chapter is to provide an introduction and overview to the rapidly growing literature on eye movements and spoken language processing, focusing on applications to spoken language comprehension. Section 2 focuses on methodological issues. As with any new paradigm, excitement about novel findings and new arenas of investigation must be tempered with concerns about the nature of the paradigm itself, including task-specific strategies, and the assumptions that link the behavioral measure to the hypothesized underlying mechanisms. Major topics include the logic linking eye movements to spoken language processing, how eye-movement data are collected and analyzed, sample applications illustrating some of the paradigms, including comparisons to eye-movement reading studies, and associated experimental logics, and finally, concerns and limitations that arise in examining language in a circumscribed visual world. In addressing these issues, we review results from a number of visual world studies. Section 3 presents a selective review of some of the major lines of research that this method has opened up, focusing on topics in language comprehension, including spoken word recognition, use of referential constraints in parsing, issues that arise in interactive conversation and the development of language processing abilities in children. Before turning our attention to these two major sections, we briefly review some of the foundational studies in the eye-movement literature on spoken language processing.

## 1.1. Some Foundational Studies

### 1.1.1. Comprehension

The use of eye movements as a tool for studying spoken language comprehension was pioneered by Roger Cooper (1974) in a remarkable article, presciently titled *The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory and language processing*. Cooper tracked participant’s eye movements as they listened to stories while looking at a display of pictures. He found that participants initiated saccades to pictures that were named in the stories, as well as pictures associated to words in the story. Moreover, fixations were often generated before the end of the word.

Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) initiated the recent wave of visual world studies, taking advantage of the advent of accurate lightweight head-mounted

eye-trackers. Tanenhaus et al. examined eye movements as participants followed instructions to perform simple tasks with objects in a workspace. They found that varying the number of potential referents for a temporarily ambiguous prepositional phrase (e.g., *Put the apple on the towel...*) determined whether the phrase was initially parsed as a goal argument (where to put the apple) or as a modifier (the location of the apple to be moved), as predicted by Altmann and Steedman (1988). (A more complete report of the Tanenhaus et al. study is presented in Spivey, et al., 2002.)

Trueswell, Skerina, Hill, and Logrip (1999) replicated the Tanenhaus et al. (1995) study with adults, and more importantly extended it to five- and eight-year-old children. They found important developmental differences in how children weight lexical and referential constraints on sentence parsing, laying the foundation for the rapidly expanding field of online sentence processing in preliterate children.

Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) demonstrated that fixations to entities referred to in an instruction are remarkably time-locked to the unfolding utterance. Fixations to a target referent among a display of competitors occurred as soon as continuous integration of constraints provided by both the unfolding speech and the visual display could, in principle, distinguish the referent from its competitors. These results obtained both for simple instructions (*touch the starred red square*) and complex instructions (*Put the five of hearts that's below the eight of clubs above the three of diamonds*). This “point-of disambiguation” logic is now widely used in studies of reference resolution.

Sedivy initiated an influential line of research demonstrating that pre-nominal scalar adjectives, such as *tall*, affect the point of disambiguation of potential referents in referential expressions, such as *the tall glass*. Speakers use, and listeners interpret, scalar adjectives contrastively, that is, to distinguish between two or more objects of the same type (Sedivy, et al., 1999; Sedivy, 2003). For example, in a display with a tall glass, a speaker will typically not use the adjective *tall*, unless the display contains, as a potential contrast, another, smaller glass (Sedivy, 2003). As they hear *tall*, eye movements show that listeners immediately interpret *the tall glass* as referring to the taller of two glasses, even when another taller object, e.g., a *pitcher* is present, whereas in the absence of a potential contrast, fixations to the glass do not begin until after the listener hears *glass* (Sedivy et al., 1999). In addition to being interesting in their own right, the processing of pre-nominal adjectives has become an important methodological tool for addressing a range of issues in language processing.

Building on initial results by Spivey-Knowlton (1996), Allopenna, Magnuson, and Tanenhaus (1998) demonstrated that the timing of fixations to a pictured referent, and competitors with different types of phonological overlap, was sufficiently time-locked to the input so as to trace the time course of lexical access. Allopenna et al. also showed that a simple linking hypothesis could be used to map fixations onto computational models of lexical activation, thus laying the foundation for the growing body of work that uses the visual world paradigm to study spoken word recognition.

Altmann and Kamide (1999) made a seminal contribution to the visual world paradigm by demonstrating linguistically mediated, anticipatory eye movements using a task like Cooper's in which participants listened to a description of an upcoming event involving entities depicted in a display. As participants heard sentences such as, *the boy will eat the cake*, they made anticipatory eye movements to a picture of cake before the offset of *eat*, when the other depicted objects were not edible. Anticipatory eye movements are now widely used as a dependent measure, typically with this so-called, passive listening (non-action-based), variant of the visual world paradigm.

In an ingenious experiment by Keysar and colleagues (Keysar, Barr, Balin, & Brauner, 2000), eye movements were used to evaluate when in the time course of comprehension listeners take into account common ground information, i.e., information that is shared with an interlocutor. A confederate speaker, the director, instructed a naive participant, the matcher, to move objects in a box with cubbyholes. Most objects could be seen by both the speaker and the matcher, and thus were in common ground by virtue of physical co-presence (Clark, & Marshall, 1981). However, some objects were blocked from the speaker's view by an opaque barrier, and were therefore only in the matcher's privileged ground. Nonetheless, the matcher looked at these objects when they, along with an object in common ground, were consistent with the speaker's referential description. This (controversial) study has laid the groundwork for investigations of how interlocutors make use of each other's likely knowledge and intentions in real-time language comprehension.

### 1.1.2. Production

Two studies laid the foundation for using eye movements to study language production. Meyer, Sleiderink, and Levelt (1998) had participant's name sequences of objects. Eye gaze was tightly coordinated with the speech. Participants fixated a to-be-named object about 1 s prior to the onset of naming. This eye-voice lag is similar to the time it takes to initiate naming an object in isolation (Rossion & Pourtois, 2004; Snodgrass & Yuditsky, 1996), suggesting that the eye-voice delay reflects word preparation.

Griffin and Bock (2000) presented participants with a simple event rendered as a line drawing that could be described with either an active or passive sentence, such as a woman shooting a man. The sequence of eye movements reflected the order of constituents in the utterance. Speakers looked at pictured objects about 800 ms to 1 s before naming them. Once speaking began, the sequence and timing of fixations was controlled by the utterance, rather than perceptual properties of the input, suggesting that the speaker had completed message planning prior to beginning to speak (also see Bock, Irwin, Davidson, & Levelt, 2003).

## 2. METHODOLOGICAL ISSUES

These early studies have raised numerous methodological questions, many of which were highlighted by the authors themselves. We now review what we see as the most important of these issues.



### 2.1. Data Analysis and Linking Assumptions

We will use Experiment 1 from Allopenna et al. (1998) to briefly describe how eye-movement data are analyzed. This experiment will also prove useful later for discussing some of the methodological concerns that arise in visual world studies in language comprehension. Allopenna et al. (1998) evaluated the time course of activation for lexical competitors that were cohorts, that is, they shared initial phonemes with the target word (e.g., *beaker* and *beetle*) or that rhymed with the target word (e.g., *beaker* and *speaker*). Participants were instructed to fixate a central cross and then followed a spoken instruction to move one of four objects displayed on a computer screen with the computer mouse (e.g., *Look at the cross. Pick up the beaker. Now put it above the square*).

#### 2.1.1. Data analysis

A schematic of a sample display of pictures is presented in Figure 1 (Panel A). The pictures include the target (the beaker), a cohort (the beetle), a rhyme (speaker), and an unrelated picture (the carriage). The particular pictures displayed are used to exemplify types of conditions and are not repeated across trials. For current purposes, we restrict our attention to the target, cohort, and unrelated pictures. Panel B shows five hypothetical trials. The 0 ms point indicates the onset of the spoken word *beaker*. The dotted line begins at about 200 ms—the earliest point where we would expect to see signal-driven fixations, give the 150–200 ms required to program and launch a saccade (Matin, Shao, & Boff,

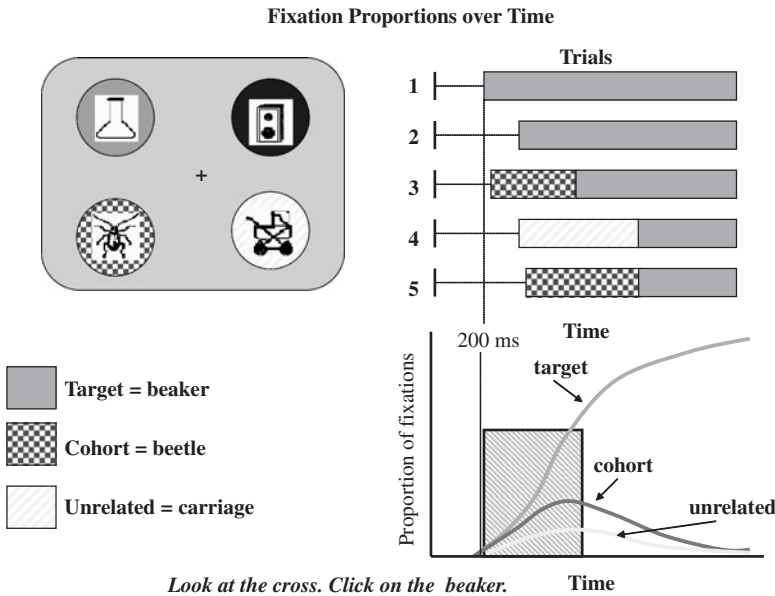


Figure 1. Sample data illustrating display, hypothetical data, proportion of fixation curves and regions of interest, modeled after Allopenna et al. (1998).

1993). On the first trial, the hypothetical participant initiated a fixation to the target about 200 ms after the onset of the word, and continued to fixate on it (typically until the hand brings the mouse onto the target). On the second trial, the fixation to the target begins a bit later. On the third trial, the first fixation is to the cohort, followed by a fixation to the target. On the fourth trial, the first fixation is to the unrelated picture. The fifth trial shows another trial where the initial fixation is to the cohort. Panel C illustrates the proportion of fixations over time for the target, cohort, and unrelated pictures, averaged across trials and participants. These fixation proportions are obtained by determining the proportion of looks to the alternative pictures at a given time slice and they show how the pattern of fixations change as the utterance unfolds. The fixations do not sum to 1.0 as the word is initially unfolding because participants are often still looking at the fixation cross.

Although proportion of fixation curves might seem to imply that eye movements provide a continuous measure it is more accurate to say that eye movements can provide an approximation to a continuous measure. The assumption linking fixations to continuous word recognition processes is that as the instruction unfolds the probability that the listener's attention will shift to a potential referent of a referring expression increases with the activation (evidence for) of its lexical representation, with a saccadic eye movement typically following a shift in visual attention to the region in space where attention has moved. Because saccades are rapid, low-cost, low-threshold responses, a small proportion of saccades will be generated by even small increases in activation, with the likelihood of a saccade increasing as activation increases. Thus, while each saccade is a discrete event, the probabilistic nature of saccades ensures that with sufficient numbers of observations, the results will begin to approximate a continuous measure (see Spivey, Grosjean, & Knoblich, 2005; Magnuson, 2005).

A window of interest is often defined, as illustrated by the rectangle in Panel C. For example, one might want to focus on the fixations to the target and cohort in the region from 200 ms after the onset of the spoken word to the point in the speech stream where disambiguating phonetic information first arrives. The proportion of fixations to pictures or objects and the time spent fixating on the alternative pictures (essentially the area under the curve, which is a simple transformation of proportion of fixations) can then be analyzed. Because each fixation is likely to be 150–250 ms, the proportion of fixations in different time windows is not independent. One way of increasing the independence is to restrict the analysis to the proportion of new saccades generated to pictures within a region of interest. In the future, it will be important for psycholinguists to explore more sophisticated statistical methods for dealing with the temporal dependencies associated with how the linguistic input at time  $t$  effects location of fixations at subsequent temporal intervals.

Figure 2 (Panel A) shows the data from the Allopenna et al. (1998) experiment. The figure plots the proportion of fixations to the target, cohort, rhyme and unrelated picture. Until 200 ms, nearly all of the fixations are on the fixation cross. These fixations are not shown. The first fixations to pictures begin at about 200 ms after the onset of the target word. These fixations are equally distributed between the target and the cohort. These fixations are remarkably time-locked to the utterance: input-driven fixations occurring

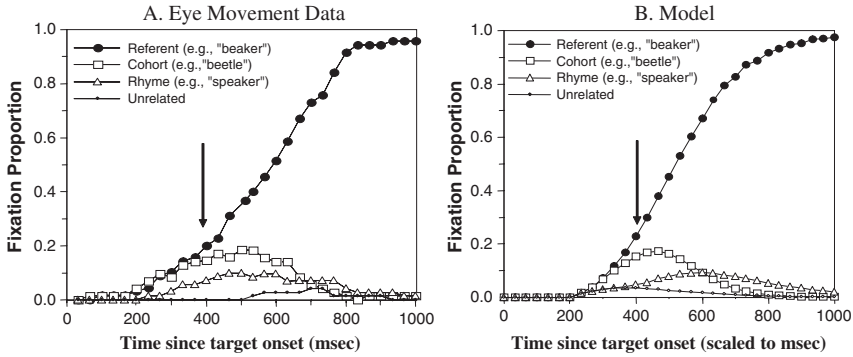


Figure 2. Data from Allopenna et al. (1998) and simulations generated by their linking hypothesis mapping activation in the TRACE model onto predicted proportion of fixations over time.

200–250 ms after the onset of the word are most likely programmed in response to information from the first 50 to 75 ms of the speech signal. At about 400 ms after the onset of the spoken word, the proportion of fixations to the target began to diverge from the proportion of fixations to the cohort. Subsequent research has established that cohorts and targets diverge ~200 ms after the first phonetic input that provides probabilistic evidence favoring the target, including coarticulatory information in vowels (Dahan, Magnuson, Tanenhaus, & Hogan, 2001b, Dahan & Tanenhaus, 2004).

Shortly after fixations to the target and cohort begin to rise, fixations to rhymes begin to increase relative to the proportion of fixations to the unrelated picture. This result supports continuous mapping models, such as TRACE (McClelland & Elman, 1986), which predict competition from similar words that mismatch at onset (e.g., rhymes), but is inconsistent with the cohort model of spoken word recognition and its descendents (e.g., Marslen-Wilson, 1987, 1990, 1993), which assume that any featural mismatch at the onset of a word is sufficient to strongly inhibit a lexical candidate.

2.1.2. Formalizing a linking hypothesis

The assumption providing the link between word recognition and eye movements is that the activation of the name of a picture determines the probability that a subject will shift attention to that picture and thus make a saccadic eye movement to fixate it. Allopenna et al. formalized this linking hypothesis by converting activations generated by a TRACE simulation into response strength, following the procedures outlined in Luce (1959). The Luce choice rule is then used to convert the response strengths into response probabilities.

The Luce choice rule assumes that each response is equally probable when there is no information. Thus when the initial instruction is “look at the cross” or “look at picture X,” the response probabilities are scaled to be proportional to the amount of activation at each time step. Thus the predicted fixation probability is determined both by the amount of evidence for an alternative and the amount of evidence for that alternative compared to the other possible alternatives. Finally, a 200 ms delay is introduced

because programming an eye movement takes  $\sim 200$  ms (Matin et al., 1993). In experiments without explicit instructions to fixate on a particular picture, initial fixations are randomly distributed among the pictures. Under these conditions, the simple form of the choice rule can be used (see Dahan et al., 2001a, 2001b). Note that the Allopenna et al. formalization is only an approximation to what would be a more accurate formalization of the linking hypothesis which would predict the probability that a saccade would be generated at a particular point in time, contingent upon (a) the location of the previous fixation (and perhaps the several preceding fixations; (b) time from the onset of the last fixation and (c) the current goal state of the listener's task—which can be ignored in a simple “click” task like the Allopenna et al. paradigm.

When the linking hypothesis is applied to TRACE simulations of activations for the stimuli used by Allopenna et al., it generates the predicted fixations over time shown in Figure 2 (Panel B). The predictions for the target, the cohort competitor, and a rhyme competitor closely match the behavioral data.

### 2.1.3. Action-contingent analyses

One useful feature of combining eye movements with an action is that the behavioral responses reveal the participant's interpretation. This allows for *interpretation-contingent* analyses in which fixations are analyzed separately for trials on which participants choose a particular interpretation. Two recent applications, illustrate how interpretation-contingent analyses can be used to distinguish among competing hypotheses.

McMurray et al. (2002) used a variation on the Allopenna et al. task to investigate the hypothesis that lexical processing is sensitive to small-within category differences in Voice-Onset Time (VOT). The stimuli were synthesized minimal pairs that differed only in voicing, such as *bomb/palm* and *peach/beach*. VOT varied in 5 ms step sizes from 0 to 40 ms. McMurray et al. found gradient increases in looks to the cross-category competitor as the VOT moved closer to the category boundary. While these results are consistent with the hypothesis that lexical processing is sensitive to within category variation, the results could also be accounted for without abandoning the traditional assumption that within-category variation is quickly discarded by making the following plausible assumption that there is noise in the system. For example, assume a category boundary of  $\sim 18$  ms. For trials with a VOT of 20 ms, given some noise, perhaps 20% of the stimuli might be perceived as having a VOT of  $< 18$  ms. With a VOT of 25 ms, the percentage might drop to 12%, compared to 8% for trials with a VOT of 30 ms and 4% for a VOT of 35 ms, etc. Thus, the proportion of looks to the cross-category competitor might increase as VOT approaches the category boundary because the data will include more trials where the target word was misheard as the cross-category competitor and not because the underlying system responds in a gradient manner.

McMurray et al. were able to rule out this alternative explanation by filtering any trials where the participant clicked on the cross-category picture. For example, if the VOT was 25 ms, and the participant clicked on the picture of the bomb, rather than the palm, then the eye-movement data from that trial would be excluded from the analyses. McMurray

et al. found that looks to the cross-category competitor increased as VOT approached the category boundary, even when all “incorrect” responses were excluded from the analyses, thus providing strong evidence that the system is indeed gradient.

A second illustration comes from recent studies by Runner and his colleagues (e.g., Runner, Sussman, & Tanenhaus, 2003, in press) investigating the interpretation of reflexives and pronouns in so-called picture noun phrases with possessors, e.g., *Harry admired Ken’s picture of him/himself*. Participants were seated in front of a display containing three male dolls, Ken, Joe, and Harry, each with distinct facial features. Digitized pictures of the doll’s faces were mounted in a column on a board directly above each individual doll. The participant was told that each doll “owned” the set of pictures directly above him; that is, the three pictures in the column above Joe were Joe’s pictures, the pictures in the column above Ken were Ken’s pictures, etc.

Binding theory predicts that the reflexive, *himself*, will be interpreted as referring to Ken’s picture of Ken in instructions such as *Pick up Harry. Now have Harry touch Ken’s picture of himself*. Runner et al. found that looks to both the binding-appropriate and inappropriate referents began to increase compared to an unrelated picture in the same row, beginning about ~200 ms after the onset of the reflexive. This result suggests that both binding-appropriate and inappropriate referents are initially considered as potential referents for a reflexive. However, participant’s choices showed frequent violations of classic binding for reflexives: on ~20% of trials with reflexives, participants had Harry touch Ken’s picture of Harry. Thus, one might argue that the early looks to binding-inappropriate referents came from just those trials on which the participant arrived at the “incorrect” interpretation. Runner et al. were able to rule out this interpretation by analyzing just those trials where the participant made the binding-appropriate response, finding that there was still an increase in looks to the inappropriate referent compared to controls.

## 2.2. Task Variables

As the eye-movement literature on spoken language comprehension has developed, researchers have begun to vary the sorts of tasks given to their participants. The effects of these variations is important to evaluate and track from experiment to experiment since as discussed in the opening of this chapter, eye movement patterns are heavily task and goal-dependent (i.e., we shift our attention to *task-relevant* regions of the world). It would be a mistake for instance, to assume that the “task” involved in the studies discussed in this chapter can be monolithically described as “spoken language comprehension” or worse still “use of language.” Very similar issues of task variation arise in reading eye-movement studies; eye-movement patterns over identical sequences of text will differ substantially depending on whether readers are *skimming*, *understanding*, *memorizing*, or *proofing*. Much greater opportunity for task variability appears to be possible in visual world studies because of the wide range of ways that participants can be asked to interact with the world. However, it is precisely this variability that provides experimenters with the leverage to make the visual world paradigm useful for such a wide range of questions.

One important task dimension is whether or not the linguistic stimuli used in the study involve instructions to act on the world. This variable is likely to be crucial because eye fixation plays an important role in visually guided reaching (see Hayhoe & Ballard, 2005). At one extreme, imperative sentences are commonly used, such that participants are required to manipulate the objects (e.g., *Pick up the ball. Put it inside the cup.*) At the other extreme, participants listen to declarative sentences, while looking at visually co-present referents. Here, the reference is intended to be non-deictic. (*The boy picked up the ball. Then he put it inside the cup.*)

Action-based studies offer several advantages in that participants are required in a highly natural way to remain engaged with their referent world; planning to execute a response requires calculating the spatial location of referents and presumably increases the time-locked nature of the relationship between linguistic interpretation and eye fixation. One clear limitation of the action-based paradigm however is that the linguistic stimuli must be embedded in instructions, which can limit the experimenter's degrees of freedom. The non-action-based listening procedure places far fewer constraints on both the experimenter and the participant. Decoupling fixations from action planning may also increase the proportion of anticipatory eye movements, which are extremely useful for inferring expectations generated by the listener.

Indeed, many of the most important applications of non-action-based listening have explored and documented referential expectations, starting with research initiated by Altmann and colleagues who showed that listeners can anticipate upcoming reference based on the semantic requirements of verbs and/or whole predicates (e.g., Altmann & Kamide, 1999; Kamide et al., 2003). Studies building upon this on this work include Boland (2005), who compared verb-based expectations for adjuncts and arguments, and Knoeferle and Crocker (in press) who studied the effects of visually based information on expectation about thematic role assignment.

We should note that this non-action paradigm is sometimes referred to as “passive” listening, and some investigators (e.g., Boland, 2005) have proposed that differences between fixations in action and passive listening tasks might be used to separate fixations that are controlled by language from those that are controlled by action. We are skeptical for several reasons. First, it is becoming increasingly clear that perception and action are inextricably intertwined in most perceptual domains, and we expect that this is also likely to be case for language. Second, interpreting sequences of fixations in the absence of an explicit task are likely to prove problematic for reasons eloquently articulated by Viviani (1990). We note however that many non-action task studies provide listeners with a well-defined task, typically so as to increase engagement with the scene and decrease the variability. For instance, Kaiser and Trueswell (2004) and Arnold et al. (2000) asked listeners to judge whether the depicted image on a trial matched the spoken description/story.

More generally it is important to keep in mind the following considerations. First, all saccadic eye movements involve some attentional overhead (Kowler, 1995). Second, the concept of passive listening leaves the underlying goals of the listener up to the listener.

Thus, each listener may adopt different goals, or worse, all listeners might adopt a pragmatically appropriate goal that was unforeseen by the experimenter. In short, there is no such thing as a taskless task. We therefore consider the notion of passive listening as akin to the notion of the null context, which is problematic for reasons articulated by Crain and Steedman (1985) and Altmann and Steedman (1988). Third, and perhaps most importantly, the difference between action-based (or perhaps more appropriately manipulation-based) and non-action-based variants of the visual world paradigm is really a subset of a more general question about the goal structures that control the moment-by-moment attentional state of the participants. In tasks with complex goal structures, e.g., a task-oriented dialog, multiple layers of goals will contribute to fixations, some of which may be tied to expectations about upcoming linguistic input, some to the current sub-goal, and some to higher-level planning.

Few studies to date have directly compared the action and non-action-based versions of the paradigm with the same materials (but cf., Sussman, 2006). However, to a first approximation, it appears that when anticipatory eye movements are excluded, the timing of fixations to potential referents may be slightly delayed in listening tasks compared to action-based tasks. The data from simple action-based tasks with imperatives (tasks where participants follow a sequence of instructions) is also somewhat cleaner than the data from non-action-based tasks with declaratives, most likely because a higher proportion of the fixations are likely to be task-relevant.

### **2.3. Comparing Visual World and Eye-Movement Reading Studies**

Many of the issues that have been investigated for decades using eye movements in reading, in particular issues in lexical processing and sentence processing are now being investigated using eye movements with spoken language. Although, some aspects of these processes will differ in reading and spoken language because of intrinsic differences between the two modalities, psycholinguists investigating issues such as syntactic ambiguity resolution and reference resolution using eye movements in reading and eye movements in spoken language believe they are testing theoretical claims about these processes that transcend the modality of the input. Thus, the psycholinguistic community will increasingly be faced with questions about how to integrate results from visual world studies with results from studies of eye movements in reading and sometimes how to reconcile conflicting results.

#### *2.3.1. Processing load versus representational measures*

In comparing reading studies to visual world studies it is useful to make a distinction between behavioral measures of language processing that measure processing difficulty and measures that probe representations. The distinction is more of a heuristic than a categorical distinction because many response measures combine aspects of both. Processing load measures assess transient changes in process complexity, and then use these changes to make inferences about the underlying processes and representations. Representational measures examine when during processing a particular type of

representations emerges and then use that information to draw inferences about the underlying processes and representations. Neither class of measure nor its accompanying experimental logic is intrinsically preferable to the other; the nature of the question under investigation determines which type of response measure is more appropriate.

The majority of studies that use eye movements to examine reading make use of eye movements as a processing load measure. The primary dependent measure is fixation duration. The linking hypothesis between fixation duration and underlying processes is that reading times increase when processing becomes more difficult. In contrast, the majority of visual world studies use eye movements as a representational measure. The primary dependent measure is when and where people fixate as the utterance unfolds. We can illustrate these differences by comparing reading studies of lexical and syntactic ambiguity resolution with visual world studies that address the same issues.

### 2.3.2. *Lexical ambiguity*

In a well-known series of studies, Rayner and colleagues (e.g., Duffy, Morris, & Rayner, 1988) have examined whether multiple senses of homographs, such as *bank*, *ball*, and *port* are accessed during reading, and if so, what are the effects of prior context and the frequency with each sense is used. Processing difficulty compared to an appropriate control is used to infer how ambiguous words are accessed and processed. For ‘balanced’ homographs with two more or less equally frequent senses, fixation duration is longer compared to frequency-matched controls—resulting in the inference that the multiple senses are competing with one another. This ambiguity “penalty” is reduced or eliminated for biased homographs when a ‘dominant’ sense is far more frequent than a ‘subordinate’ sense and when the context strongly favors either one of two equally frequent senses or the more frequent sense. Note that while these results do not provide clear evidence about time course per se, the overall data pattern allows one to infer that multiple senses are accessed, with the dominant sense accessed more rapidly. One can get crude time-course information by separately analyzing the duration of the initial fixation and using that as a measure of relatively early processes. More detailed information about time course can be obtained by using fixation duration as a measure, but using variations on the fast priming methods, introduced by Sereno and Rayner (1992).

A study using the visual world paradigm would adopt a similar approach to that used by Allopenna et al. Potential referents associated with the alternative senses would be displayed and the time course of looks to these referents would be used to infer degree of activation and how it changes over time. For balanced homophones, one would predict looks to the referents of both senses. For biased homophones, looks to the more frequent would begin earlier than looks to the less frequent sense. This pattern would be similar to those obtained in classic studies using cross-modal priming from the 1970s and early 1980s (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979; for review see Simpson, 1984; Lucas, 1999). Note that these results would not provide direct information about processing difficulty, though one might infer from them that competing senses would result in an increase in complexity.



Thus, while the eye-movement reading studies do not provide direct information about time course and visual world studies do not provide direct information about processing difficulty, the results from reading studies that use a processing load strategy and visual world studies that probe emerging representations could converge on the same conclusions.

### 2.3.3. *Syntactic ambiguity*

Beginning with the classic article by Frazier and Rayner (1982), eye tracking in reading has been the response measure of choice for psycholinguists interested in syntactic processing. Frazier and Rayner's approach was to examine the processing of temporarily ambiguous sentences, using reading times within pre-defined regions to infer if and when the reader had initially pursued the incorrect interpretation. For a range of syntactic ambiguities, most of which involved disambiguating the phrase that could be "attached" to a verb phrase, thereby introducing an argument, in favor of a noun phrase attachment that modified the head noun, Frazier and Rayner found an increase in fixation duration and an increase in regressive eye movements from the disambiguating region. For current purposes we will focus on fixation duration because it is most clearly a processing load measure. The question of how to interpret regressions is more complex and beyond the scope of this chapter. The increase in fixation duration was interpreted as evidence that processing had been disrupted, thereby leading to the inference that readers had initially chosen the argument interpretation. Frazier and Rayner also introduced several different measures that divided fixations within a region in different ways. For example, 'first pass' reading times include all fixations beginning with the first fixation within a region until a fixation that leaves a region, and are often used as a measure of early processing.

Timing is less straightforward in eye-tracking reading when fixations are divided into multiple word regions. Most of the complexities in inferring time course in reading studies arise because the sequence of fixations need not correspond to the linear order of the words in the text. This is especially the case when one considers that arguments about timing often depend on defining regions of text and then partitioning fixations into categories in ways that separate the measure from when the input is first encountered.

Studies examining syntactic ambiguity resolution with the visual world paradigm use the timing of looks to potential referents to infer, if and, if so, when, a particular analysis is under consideration. For example, in one-referent contexts (an apple on a towel, a towel, a box and a pencil) and instructions such as, *Put the apple on the towel in the box*, Spivey et al. (2002) found that looks to the false goal (the towel without the apple) began to increase several hundred millisecond after the onset of *towel* (see Figure 3). In contrast, in two referent contexts (two apples, one on a towel and one on a napkin) fixations to the apple on the towel begin to increase several hundred millisecond after the onset of *towel*. This pattern of results suggests that the prepositional phrase *on the towel* is initially considered a goal argument in the one-referent context and a noun phrase modifier in the two-referent context. Information about time course is straightforward with the visual world logic because fixations can be aligned with the input, allowing strong inferences about what information in the input was likely to have triggered the fixation. The reason

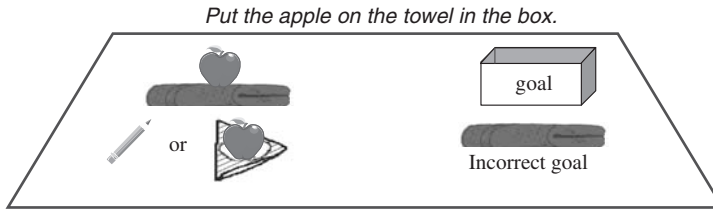


Figure 3. Schematic of the one and two-referent conditions in the Tanenhaus et al. (1995) and Spivey et al. (2002) prepositional phrase-attachment studies.

that one can align fixations and the input is, of course, because the input unfolds sequentially. Note, however, that one cannot use fixations in a straightforward way to draw inferences about processing difficulty. Thus the visual world approach is unlikely to become a paradigm of choice for investigating issues about resource demands, including increasingly important questions about what factors contribute to the complexity of sentences (e.g., Grodner & Gibson, 2005; Hale, 2003; Lewis & Vasishth, 2005).

## 2.4. Effects of Display

The single factor that most complicates the interpretation of visual world studies of language processing is the need to use a display. First, the encoding of the display can introduce contingencies. For example, the timing of looks to a potential referent at point  $t$  could be affected by whether or not that referent has been fixated on during time  $t-x$ , either during preview or as the sentence unfolds. Thus the likelihood of a fixation may be contingent on both the input and the pattern of prior fixations. This, of course, has the potential to complicate inferences about time course, in much the same way that re-reading after a regression can complicate the interpretation of fixation duration data in eye-movement reading studies. Recent studies have begun to examine how having fixated a potential referent during preview affects the likelihood that it will be fixated when it is temporarily consistent with the input (Dahan, Tanenhaus, & Salverda, in press).

Second, use of a display with a small number of pictured referents or objects and a limited set of potential actions creates a more restricted environment than language processing in most natural contexts, while at the same time imposing more demands on the participant than most psycholinguistic tasks. In order to address these closed set issues, we will consider two cases: the first from spoken word recognition; the second from reference resolution.

### 2.4.1. Spoken word recognition

In the Allopenna et al. paradigm, the potential response set on each trial is limited to four pictured items. If participants adopted a task-specific verification strategy, such as implicitly naming the pictures, then the unfolding input might be evaluated against these activated names, effectively bypassing the usual activation process, and leading to

distorted results. Even if participants do not adopt such a strategy, the visual world methodology might be limited if the effects of the response alternatives mask effects of non-displayed alternatives (e.g., neighborhood effects in the entire lexicon). This would restrict its usefulness for investigating many issues in spoken word recognition, in particular issues about the effects of lexical neighborhoods, i.e., the set of words in the lexicon that are similar to the target word. Here, an analogy might be helpful. Researchers often use lexical priming paradigms to probe for whether an exemplar of a particular class of lexical competitor is active, for example, cohorts or rhymes. However, these paradigms are not well suited for asking questions about the aggregate effects of the number and frequency of potential competitors. In order to investigate this class of question, researchers have found it more useful to measure response time to a target word, for example, auditory lexical decision, which more closely approximates a processing load measure.

#### 2.4.2. *Implicit naming*

The issue of implicit naming has been addressed most directly by Dahan and Tanenhaus (2005) in a study that varied the amount of preview time, 300 or 1000 ms, for four-picture displays with minimal phonological overlap between the names of the distractors and the target (Figure 4). On a subset of the trials, two of the pictures were visually similar (e.g., a picture of a snake and a coiled rope) and the instruction referred to one of the pictures (e.g., *click on the snake*). The particular pictures chosen as the two referents shared some features associated with a prototypical visual representation of

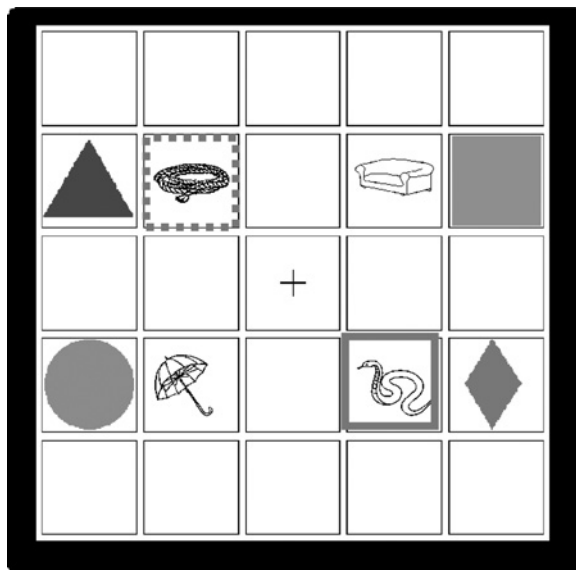


Figure 4. Sample display from Dahan and Tanenhaus (1995), illustrating a display where the visual competitor (the rope) is visually similar to a prototypical snake, whereas the picture of the target referent (snake) is somewhat less prototypical.

one or both words. For example, the pair *snake–rope* was selected because the picture of a coiled rope shares some features with the visual representation most often associated with the concept of a snake. When selecting pictures, Dahn and Tanenhaus (2005) sought to minimize their visual similarity so that the objects could be easily differentiated. For example, we chose a snake in a non-coiled position. Thus, visual similarity was maximized between the prototypical visual representation of one of the concepts, the referent, and the picture associated with the other concept, the competitor, and minimized between the competitor picture and the picture of the referent concept.

Several aspects of the results provide strong evidence against implicit naming. Preview duration did not affect the magnitude of visual similarity effects (looks to visually similar competitors). Moreover, even in the 1000 ms condition, the magnitude of visual similarity effects was not affected by whether or not the competitor was fixated during preview; the naming hypothesis predicts that effects would be eliminated or weakened with preview because the encoded name of the picture would not match the unfolding target. Finally, similarity effects were larger when the target had a competitor that was chosen to share visual features of its prototype representation compared to when that competitor was the referent. Thus visual similarity effects were due to the fit between the picture and the conceptual representation of the picture, not simply surface visual confusability. This suggests that mapping of the word onto its referent picture is mediated by a visual/conceptual match between the activated lexical form of the target and the picture. This hypothesis is further supported by additional analyses of the effects of fixation to a competitor during preview on the likelihood that it will be re-fixated during the speech input and evidence that a spoken word triggers looks to potential referents when the participant is engaged in a visual search task to identify the location of a dot when it appears on a random location within a schematic scene (Salverda & Altmann, 2005).

#### 2.4.3. *Sensitivity to hidden competitors*

Perhaps, the strongest test of the sensitivity of visual world studies comes from studies that look for effects of non-displayed or “hidden competitors.” For example, Magnuson, Dixon, Tanenhaus, and Aslin (in press) examined the temporal dynamics of neighborhood effects using two different metrics: neighborhood density, a frequency-weighted measure defined by the Neighborhood Activation Model (NAM), and a frequency-weighted measure of cohort density. The referent was displayed along with three semantically unrelated pictures, with names that had little phonological overlap with the referent (all names were monosyllabic). Crucially, none of the referent’s neighbors were either displayed or named throughout the course of the experiment. The results showed clear effects of both cohort and neighborhood density, with cohort density effects dominating early in the recognition process and neighborhood effects emerging relatively late.

These results demonstrate that the processing neighborhood for a word changes dynamically as the word unfolds. It also establishes the sensitivity of the paradigm to the entire lexicon. To a first approximation then, when competitors are displayed, the paradigm can be used to probe specific representations, however, the aggregate effects of competitors can be observed in the timing of fixations to the target referent.

Magnuson et al.'s results complement Dahan et al. (2001b) finding that misleading coarticulatory information delays recognition more when it renders the input temporarily consistent with a (non-displayed) word, compared to when it does not. In addition, simulations using the Allopenna et al. linking hypothesis successfully captured differences between the effects of misleading coarticulatory information with displayed and non-displayed competitors. Whether the non-displayed competitor logic can be extended to higher-level sentence processing remains to be seen.

#### 2.4.4. *Sentence processing*

Much trickier issues about the effects of the display come into play in higher-level processing. For example, one could argue that in the Tanenhaus et al. (1995) study displaying an apple on a towel and an apple on a napkin increases the salience of a normally less accessible sense compared to circumstances where the alternative referents are introduced linguistically. One could make a similar argument about the effects of action on the rapidity with which object-based affordances influence ambiguity resolution in studies by Chambers and colleagues (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers et al. 2004). In these studies, the issue of implicit naming seems *prima facie* to be less plausible. However, one might be concerned about task-specific strategies. For example, in Chambers et al. (2002), participants were confused, as indexed by fixations when they were told to, *Pick up the cube. Now put the cube in the can*, and there were two cans. The confusion was reduced or eliminated, however, when the cube would only fit in one of the cans. Because only one action was possible, one might attribute this to problem solving, and not as Chambers et al. argued to the effects of action and affordance on referential domains. However, the manipulation had opposite effects for instructions that used an indefinite article, e.g., *Pick up the cube. Now put it in a can*. Here participants were confused when the cube would only fit in one of the cans. This strategy of pitting linguistic effects against potential problem-solving strategies is crucial for evaluating the impact of strategies due to the display and the task.

Perhaps, the most general caution for researchers using the visual world paradigm in both production and comprehension is to be aware that while the visual world displays entities that can be used to infer the representations that the listener is developing, it also serves as a context for the utterance itself. Note that the fact that information in a display affects processing is not itself any more problematic than the observation that reference resolution, for example, is affected by whether or not potential referents are introduced linguistically in a prior discourse. One sometimes encounters the argument that the visual world paradigm can be informative about language processing only if gaze patterns to a potential referent in a display are not affected by the other characteristics of the display. This argument is no more or less valid than the comparable argument that fixations in reading can only inform us about word recognition or reference resolution if fixations to a word are unaffected by the context in which the fixated word occurs. What is crucial, however, is whether the nature of the interactions with the display shed light on linguistic processing or whether they introduce strategies that mislead or obscure the underlying processes. Thus, far investigations of potential problems has been encouraging for the

approach. However, it will be crucial in further work to explore the nature of the interactions between the display and linguistic processing in much greater detail.

### 3. APPLICATIONS TO ISSUES IN LANGUAGE COMPREHENSION

In this section, we present a brief review of work in three domains where using eye movements is beginning to have a major impact on our understanding of spoken language processing. We begin with issues in speech, spoken word recognition, and prosody that can be addressed by using variations of the procedures we described in presenting the study by Allopenna et al. (1998). The second domain consists of issues in sentence processing, including classic issues about the role of context in syntactic ambiguity resolution, and assorted issues about referential domains. These issues are addressed by taking advantage of various features of the visual world paradigm, including having an implicit measure that can be used with simple tasks and spoken language, having a co-present referential world, and the capability of monitoring real-time processing in paradigms that bridge the language-as-product and language-as-action traditions. We then conclude with a discussion of how eye movements are beginning to provide insights into how real-time language processing develops in infants, toddlers, and young children.

#### 3.1. Spoken Word Recognition and Prosody

##### 3.1.1. *Spoken word recognition*

As we noted earlier, many classic issues about spoken word recognition can naturally be addressed using variations on the procedure used by Allopenna et al. (1998). These include questions about what types of lexical competitors become activated as a spoken word unfolds (Allopenna et al., 1998; Magnuson, 2002) and how lexical competition is modulated by context (Dahan et al., 2001b; Dahan & Tanenhaus, 2004). The Allopenna et al. (1998) procedure has also proved to be extremely useful for addressing questions about how listeners use sub-phonetic information in word recognition. Examples include the McMurray et al. (2002) study described earlier, which used looks to competitors to demonstrate fine-grained sensitivity to within-category variation and work by Dahan et al. (2001a) and Gow and McMurray (in press) on listener's use of coarticulatory information (Dahan et al., 2001a). And, in an important study, Salverda, Dahan, and McQueen (2003) used eye movements to demonstrate that listeners exploit small systematic differences in vowel duration in processing of words such as *captain*, which begin with a phonetic sequence that is itself a word, e.g., *cap* (the vowel in a monosyllabic word is typically longer than the same vowel in a polysyllabic word). Examining looks to cohort competitors to words embedded in utterances has also proved useful for examining spoken word recognition in bilinguals. For example, Spivey and Marian (1999) used looks to cohorts to demonstrate that bilingual speakers following instructions in one language, briefly consider potential referents with names that are cohort competitor in their second language (see also Ju & Luce, 2002). Finally, studies that use eye movements to measure processing of artificial lexicons and languages, initiated by

Magnuson and his colleagues (Magnuson, Tanenhaus, Aslin, & Dahan, 2003) are proving useful for addressing a range of issues in spoken word recognition and learning.

### 3.1.2. *Prosody*

Visual world studies are beginning to have an increasingly large impact on research investigating how listeners process information about prosody, which is carried by the pattern and type of pitch accents and realized acoustically as changes in duration, intensity, and pitch excursion on stressed vowels. Differences in vowel duration between mono- and polysyllabic words vary with the prosodic environment; they are smallest in the middle of a phrase, and largest at the end of a phrase. Salverda (2005) demonstrated that prosodic factors modulate the relative degree to which different members of a neighborhood will be activated in different environments; in medial position a polysyllabic carrier word such as *captain* is a stronger competitor than *cat* for the target *cap*, whereas the opposite pattern obtains in utterance-final position.

Cohort manipulations, in particular, are well suited for examining pitch accents because one can examine effects that are localized to the vowel that carries the pitch accent. Dahan, Tanenhaus, and Chambers (2002) examined the timing of looks to targets and cohort competitors for accented and unaccented words that referred to discourse given and discourse-new entities (e.g., *Put the candle above the triangle. Now put the CANDY/candy...*). Dahan et al. found that listeners use information about pitch accent as the vowel unfolds, initially assuming that nouns in definite referring expression with unaccented vowels refer to the most salient entity (the subject/focus) of the previous sentences, whereas words with accented vowels refer to a non-focused given entity if available, or if not, a new entity.

Arnold and colleagues (Arnold, Tanenhaus, Altmann, & Fagnano, 2004) adapted the Dahan et al. cohort design to evaluate the hypothesis that a disfluent production of a noun phrase (thee uh CANDY) would bias listeners to expect reference to a discourse-new entity. With fluent productions, Arnold et al. replicated Dahan et al.'s finding that an accented noun was preferentially interpreted as referring to a non-focused entity. However, with a disfluent production, the preference shifted to the discourse-new entity. Watson and his colleagues (e.g., Watson, Gunlogson, & Tanenhaus, in press) have also used cohort competitors to test hypotheses about the interpretation of different pitch accents, focusing on potential differences between the H\* (presentational) and L+H\* (contrastive) pitch accents (Pierrehumbert & Hirschberg, 1990).

Ito and Speer (described in Speer & Ito, in press) have also investigated presentational and contrastive accents, combining eye movements with a "targeted language game." The director, a naïve participant, instructs a confederate about how to decorate a Christmas tree using ornaments that need to be placed on the tree in a specified sequence. Ornaments differ in type, e.g., bells, hats, balls, houses, etc. and in color, e.g., orange, silver, gold, blue, etc. Recordings demonstrated that participants typically used a presentational accent (H\*) when a color was new to the local discourse. For example, "orange" typically received a presentational accent in the instruction, "First, hang an orange ball

on the left” when an orange ornament was being mentioned for the first time) for a particular row. However, if the instruction to place the orange ball followed placement of a ball of a different color, e.g., a silver ball, then “orange” was more likely to be produced with a contrastive accent (L+H\*). Ito and Speer showed that the recordings using the preferred pitch accent pattern used by naïve participants facilitated listeners’ time to identify the correct ornament, as measured by eye movements.

## 3.2. Sentence Processing

### 3.2.1. Syntactic ambiguity resolution

In a series of classic papers, Crain (1981), Crain and Steedman (1985), and Altmann and Steedman (1988) argued that many of the systematic preferences that readers and listeners exhibit when resolving temporary syntactic ambiguity are not due to differences in syntactic complexity between the alternative structures, but rather to differences in referential implications. A well-known example comes from prepositional (PP) attachment ambiguities as illustrated in sentences such as *Anne hit the thief with the wart* is one such example. The strong initial preference to consider *with the wart* (erroneously) as the instrument of *hit* rather than as a restrictive modifier of *the thief* could in part be due to the fact that the restrictive modifier is most felicitous in a context in which multiple thieves are present, one of which has a wart. In the absence of such a context, there is little reason for considering the modification analysis. Indeed, some (but not all) eye-movement studies with text have found that this referential factor (i.e., the presence/absence of referential ambiguity) has immediate effects on real-time syntactic ambiguity resolution in reading (e.g., Altmann & Steedman, 1988; Britt, 1994; Sedivy, 2003; Spivey-Knowlton & Sedivy, 1995; Spivey & Tanenhaus, 1998; but for discussion of studies finding weak or delayed effects of referential context see Rayner, this volume, and Rayner & Liversedge, 2004).

Introducing a referential world that is co-present with the unfolding language, naturally highlights these and other questions about reference. Indeed, the initial action-based visual world study (Tanenhaus et al., 1995, described earlier) examined how referential ambiguity (i.e., the presence of multiple apples in a scene) influences the listeners’ initial bias when encountering a sentence with a temporarily ambiguous prepositional phrase (*Put the apple on the towel in the box.*). Recall that the presence of two apples in the scene shifted listeners’ initial preference to interpret *on the towel* from a goal preference to a modifier preference. This study confirms that something like Crain’s Referential Principle is an important factor when listeners interpret spoken language in the context of visually co-present referents.

Subsequent work by Snedeker and Trueswell (2004) confirms the importance of referential context, but importantly establishes that high-level expectations *contribute* to but do not solely *determine* the outcome of ambiguity resolution in visual contexts. A multiple constraint view of sentence processing predicts that lower-level linguistic factors, such as verb argument preferences, contribute simultaneously to the ambiguity resolution process. Snedeker and Trueswell (2004) confirmed this prediction in a study containing



sentences that were globally ambiguous in their structure (not just temporarily ambiguous). College-age adults heard sentences like those in (1a) through (1c).

1. a. *Tickle the pig with the fan.* (Instrument-biased Verb)
- b. *Feel the frog with the feather.* (Equi-biased Verb)
- c. *Choose the cow with the stick.* (Modifier-biased Verb)

Verbs were selected based on a separate sentence completion study, which evaluated how often a *with*-phrase would be used for these verbs as an instrument, allowing verbs to be operationally defined as: Instrument-bias, Modifier-bias, or Equi-bias. As in Tanenhaus et al. (1995), 2- and 1-Referent scenes were compared. Scenes contained, e.g., a Target Theme (a pig holding a small fan); a Competitor Theme (a pig/horse wearing a hat); a Potential Instrument (a large fan); and another object (a large hat). Here, looks to the potential instrument and the ultimate action were analyzed: i.e., participants could pick up the fan and use it to tickle the pig, or they could use their fingers to do the actions. The eye movement and action data revealed simultaneous effects of both the referential context (2-Referent versus 1-Referent) and verb argument preferences; the presence of multiple pigs reduced looks to, and use of, the potential instrument; likewise degree of verb-bias (from Instrument-biased to Modifier-biased) systematically decreased looks and actions involving the Potential Instrument. Crucially, these verb effects were observed in both *1-Referent* and *2-Referent* Scenes, suggesting that the mere presence of multiple referents does not solely determine attachment preferences for listeners.

We note that it remains something of a puzzle why the effects of referential context seem so much stronger in studies examining the PP-attachment ambiguities involving goals versus modifiers (*Put the apple on the towel*) compared to Instrument versus modifiers (*Tickle the frog with the feather*) given that *put* is a verb that has a strong goal-bias. For some speculation about possible explanations, see Snedeker and Trueswell (2004); Spivey et al. (2002); Tanenhaus and Trueswell (2005), and Trueswell and Gleitman (2004).

### 3.3. Circumscribing Referential Domains

The studies reviewed thus far made the simplifying assumption that the referential domain for a linguistic expression comprises all of the salient entities in the environment that are temporarily consistent with the referring expression as it unfolds. However, speakers at least in their own productions consider real-world constraints like the proximity and relevance of potential referents, the relevance of other estimations of the knowledge that the listener has of the world, and several other factors (Clark, 1992; Levelt, 1989; Lyons, 1981; Stone & Webber, 1998). Put more concretely, a speaker's decision to refer to an object as *the ball*, *the red ball*, *the ball closer to you*, *the slightly asymmetric sphere*, *it*, *that one* or *that*, clearly depends on this wide range of spatial, perceptual, social, and cognitive factors.

A central theme of research using the visual world paradigm has been to understand how and when these factors impinge on decisions made by listeners and speakers (Chambers et al., 2001; Sedivy, 2003; Sedivy et al., 1999; Grodner & Sedivy, in press; Keysar et al., 2000; Keysar & Barr, 2005; Brown-Schmidt et al., 2005; Brown-Schmidt & Tanenhaus, in press). For instance, we have already discussed some studies demonstrating that listeners dynamically update referential domains, integrating information from the unfolding utterance in conjunction with the entities in the workspace (Chambers et al., 2002, 2004; Eberhard et al., 1995) and generating expectations about upcoming referents (Altmann & Kamide, 1999; Kamide et al., 2003), especially those that are likely to be realized as arguments (Boland, 2005). And, in an ingenious series of eye movement studies, Altmann and colleagues have recently demonstrated that actions described or implied in a narrative influence expectations about how the location of objects will change in the listener's mental model of the scene, as determined by looks to locations in the scene (Altmann & Kamide, 2004).

A listener's referential domain is also affected by intended actions and the affordances of potential objects that are relevant to those actions (Chambers et al., 2002). These affordances also affect the earliest moments of syntactic ambiguity resolution, challenging the claim that language processing includes a syntactic subsystem (module) that is informationally encapsulated, and thus isolated from high-level non-linguistic expectations (Coltheart, 1999; Fodor, 1983). For example, Chambers, Tanenhaus, and Magnuson (2004) showed that in a two-referent context that includes a liquid egg in a bowl and a liquid egg in a glass, participants will initially treat the PP *in the bowl* as a modifier with an instruction such as *pour the egg in the bowl over the flour*. However, when the egg in the bowl is solid and thus cannot be poured, then participants initially misinterpret *in the bowl* as the Goal. These results cannot be attributed to constraints lexically encoded within the linguistic representation of the verb *pour*; Chambers et al. found the same pattern of results with the verb *put* when the affordances were introduced non-linguistically by handing the participant an instrument.

### 3.3.1. *Scalar implicatures*

Earlier we reviewed Sedivy's finding that listeners assume that the referential domain includes a contrast set when they hear a pre-nominal scalar adjective, such as *tall*. These results are particularly striking because they represent one case in which listeners immediately generate a pragmatic inference based on a generalized implicature. There is an emerging debate about when listeners generate these types of inferences, whether they apply differently to different classes of scales, especially those that involve potential contrasts between a so-called logical interpretation (e.g., logical or inclusive *OR* versus pragmatic or exclusive *OR*) where there are claims that logical *OR* is computed (obligatorily) prior to pragmatic *OR*, and how these inferences are modulated by context (see Noveck & Sperber, 2005). Visual world eye-movement studies are beginning to feature prominently in research in this arena, though this work had not yet begun to appear in the literature as we were preparing this chapter.

Eye-movement research using pre-nominal adjectives is beginning to shed light on inference under other circumstances. Although any adjective can appear post-nominally, either in a restrictive relative clause (*the glass that is tall*) or in a prepositional phrase (*The glass with spots*), some adjectives are typically used pre-nominally (e.g., scalar adjectives and color adjectives), others are nearly always used post-nominally (*the shape with diamonds*), and others occur equally often in pre-nominal and post-nominal positions (e.g., *striped, with stripes*). Using a point of disambiguation logic, Edwards and Chambers (2004) have shown that listeners make rapid use of the **absence** of a pre-nominal modifier to rule out candidate referents. Second, Grodner and Sedivy (in press) have established that listeners rapidly adjust to how reliably a speaker uses scalar adjectives contrastively, including making adjustments based on meta-linguistic information provided by an experimenter. Arnold, J. E. (personal communication) reports similar results with meta-linguistic information provided about a disfluent speaker. These results bear on questions about when in the time course of processing, and under what circumstances speakers and listeners consider the likely knowledge and intentions of their interlocutors, a topic we will return to shortly.

### 3.4. Word-Order Variation, Discourse, and Information Structure

The visual world paradigm has also proved to be a useful tool investigating how discourse-pragmatic factors related to information structure influence reference resolution and parsing. One such area has been an exploration of how sentence processing is achieved in languages that have highly flexible word orders (Kaiser & Trueswell, 2004; Järvi­kivi, van Gompel, Hyönä, & Bertram, 2005). The reason for this interest is that flexible word-order languages of this sort typically use order to communicate the information structure and discourse status (given/new distinctions). Kaiser and Trueswell (2004) used the visual world paradigm to explore how reference resolution in Finnish, a flexible word-order language with canonical SVO order and no articles. The non-canonical order OVS marks the object as given and the subject as new; SVO is more flexible, being used in multiple contexts. In the study, the eye gaze of Finnish listeners was tracked as they heard spoken descriptions of simple pictures, so as to test whether listeners use this knowledge of information structure to their advantage, to increase the efficiency with which visual information is collected. That is, upon hearing an OV... sequence, Finnish listeners should expect the upcoming noun to be discourse-new, whereas an SV... sequence makes no such prediction. The results confirmed these predictions. As compared to SVO, OVS sentences caused listeners to launch anticipatory eye movements to a discourse-new referent at the second noun onset, even before participants had enough acoustic information to recognize this word. The findings illustrate that in a flexible word-order language, a non-canonical order can result in anticipatory processes regarding the discourse status of a yet-to-be-heard constituent.

### 3.5. Pronouns and other Referring Expressions

Relatedly, numerous researchers have begun to use the visual world paradigm to study how syntax and information structure interact with the type of referring form (full noun

phrases, pronouns, etc.) (Arnold et al., 2000; Järvikivi, van Gompel, Hyönä, & Bertram, 2005; Brown-Schmidt et al., 2005; Runner et al., 2003, in press). The visual paradigm is particularly useful for addressing these questions because the looks to potential referents, especially, when combined with a decision, allow for strong inferences about which potential referents are being considered and which referent is selected.

Several studies have examined how the order in which characters in a scene are mentioned influence the interpretation of utterances with both ambiguous and unambiguous pronouns. Arnold et al. (2000) found that English listeners upon hearing a sentence beginning with an ambiguous pronoun (he) preferentially looked to the character that had been mentioned first in the previous sentence. Kaiser and Trueswell (in press) show that this preference, at least in Finnish, reflects a preference for pronouns to refer to the grammatical subject of the previous sentence, not the object (but see also Järvikivi, van Gompel, Hyönä, & Bertram, 2005). Preferences depend though on the type of pronoun used in Finnish, another class of pronouns (demonstratives) preferentially selects referents based on surface word order rather than grammatical role. Brown-Schmidt and colleagues (Brown-Schmidt, Byron, & Tanenhaus, 2005) used eye movements and actions to demonstrate differences in the interpretation of *it* and *that*, following an instruction such as *Put the cup on the saucer. Now put it/that....* Addressees preferentially interpret *it* as referring to the theme (the cup), whereas *that* is preferentially interpreted as referring to the composite created by the action (the cup on the saucer), which does not have a linguistic antecedent (*the cup on the saucer* is not a constituent in the instruction). Finally, as we mentioned earlier, the visual world paradigm is being used to examine the interplay between structural constraints (e.g., binding constraints), discourse, and type of referring expressions for pronouns and reflexives (Runner et al., 2003, in press).

### 3.6. Common Ground, Alignment, and Dialogue

Until recently, most psycholinguistic research on spoken language comprehension could be divided into one of two traditions, each with its own theoretical concerns and dominant methodologies (Clark, 1992; Trueswell & Tanenhaus, 2005). The product tradition emphasized the individual cognitive processes by which listeners recover linguistic representations, typically by examining moment-by-moment processes in real-time language processing, using carefully controlled stimuli scripted materials and fine-grained on-line measures.

In contrast, the action tradition focused on how people use language to perform acts in conversation—the most basic form of language use. Many of the characteristic features of conversation emerge only when interlocutors have joint goals and when they participate in a dialogue both as a speaker and an addressee. Thus, research within the action tradition typically examines unscripted interactive conversation involving two or more participants engaged in a cooperative task, typically with real-world referents and well-defined behavioral goals—conditions that are necessary for many of the characteristic features of conversation to emerge.

Recently, the language-processing community has begun to show increased interest in bridging the product and action traditions (Pickering & Garrod, 2004; Trueswell & Tanenhaus, 2005). However, research that aims to bridge the two traditions has rarely combined on-line measures—the methodological cornerstone of the product tradition, with unscripted cooperative conversation—the central domain of inquiry in the action tradition (see Brennan, 1990, 2005 for a notable exception). The reason is that most on-line measures interfere with dialogue. In contrast, eye movements can be monitored in most of the tasks used by researchers in the action tradition.

We believe that research monitoring eye movements in unscripted conversation is likely to play a central role in addressing at least two fundamental questions that are becoming the focus of much current research. The first is at what temporal grain do interlocutors monitor each other's likely knowledge and intentions. The second is to what degree, and at what temporal grain, do the representations of interlocutors become **aligned** during interactive conversation (Pickering & Garrod, 2004).

With respect to common ground, although keeping track of what is known, and not known, to the individual participants in a discourse would seem to be fundamental for coordinating information flow (Brennan & Hulstijn, 1995; Clark, 1992, 1996), computing common ground by building, maintaining, and updating a model of a conversational partner's beliefs could be memory intensive. (Thus interlocutors may not consider common ground during initial processing; Keysar & Barr, 2005.) Some supporting evidence comes from eye-movement studies showing that addressees often fail to reliably distinguish their own knowledge from that of their interlocutor when interpreting a partner's spoken instructions (Keysar et al., 2000; Keysar, Lin, & Barr, 2003; but cf. Nadig & Sedivy, 2002; Hanna et al., 2003). However, these studies use confederates, which restricts and changes the nature of the interaction (Metzing & Brennan, 2003), the degree to which common goals are negotiated, and perhaps most importantly the types of the constructions that are used in the conversation, each of which can mask effects of perspective taking (for discussion and supporting evidence, see Tanenhaus & Brown-Schmidt, in press).

With respect to alignment, Pickering and Garrod (2004) propose that successful dialogue requires interlocutors to arrive at similar (aligned) representations across multiple linguistic and conceptual levels. They further propose that priming provides a mechanism by which alignment occurs, noting, for example, that syntactic persistence, the tendency for speakers to choose a structure they have previously heard or produced, appears to be particularly robust in dialogue (Branigan, Pickering, & Cleland, 2000). However, even if Pickering and Garrod are correct in identifying priming as an important mechanism for alignment priming will have to be supplemented by real-time measures that probe the representations of interlocutors. Otherwise, priming is being called upon to serve both as a proposed mechanism, and as a diagnostic for alignment, raising concerns about circularity.

Recent research by Brown-Schmidt and colleagues demonstrates that it is possible to use eye movements to monitor real-time processes in task-oriented dialogues with complex tasks and naïve participants (Brown-Schmidt, Campana, & Tanenhaus, 2005;

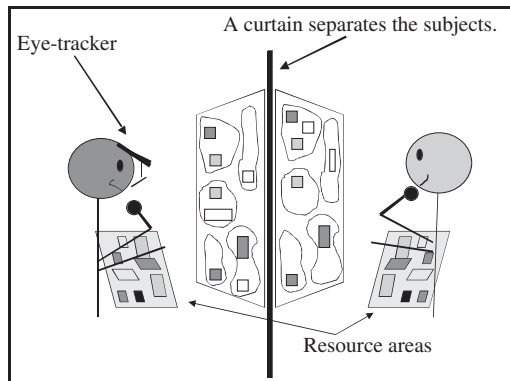


Figure 5. Schematic of the referential communication task used in Brown-Schmidt et al.'s (2005) targeted language games study with naïve participants and unscripted dialogue.

Brown-Schmidt, 2005; Brown-Schmidt & Tanenhaus, 2005) For example, Brown-Schmidt et al. used a referential communication task in which participants separated by a barrier cooperated to replace stickers with blocks to match the placement of the blocks in their respective boards (see Figure 5).

They adopted what they termed a “targeted language game” approach, placing stickers to maximize the likelihood that conditions approximating those that might be incorporated in a standard factorial design would emerge. Despite the complexity of the dialogue, they were able to see point-of-disambiguation effects for referring expressions that mirror effects observed in studies with scripted instructions and simple displays. In particular, as a speaker’s referring expression unfolded, the addressee’s fixations to the referent increased, and fixations to potential competitors decreased, about 200 ms after the place in the speech stream where the input first disambiguated the target from the temporarily consistent competitors.

Additional results strongly demonstrated that the addressee’s referential domains were closely aligned. For example, when proximal competitors that did not match the immediate task goals were not part of the speaker’s referential domain (as inferred by the form of the referring expression), they were also not considered as potential referents by the addressee (as inferred from fixations).

### 3.7. Development of Comprehension Abilities

Eye gaze during listening in studies with infants, toddlers, and young children is proving to be a powerful tool for addressing developmental issues in language processing (e.g., Arnold, Brown-Schmidt, Trueswell, & Fagnano, M., 2005; Song & Fisher, 2005; Fernald, Pinto, Swingley, Weinberg, McRoberts, 1998; Swingley, Pinto, & Fernald, 1999; Swingley & Aslin, 2002; Snedeker & Trueswell, 2004; Trueswell et al., 1999). (For

a review, see Trueswell & Gleitman, 2004.) In these studies, the time course of children's eye movements is established either by inspecting a videotape of the child's face frame by frame (Swingley et al., 1999), or by analyzing the output of a lightweight eye-tracking visor worn by the child (Trueswell et al., 1999). These eye-movement techniques have the potential to revolutionize how we examine the child's emerging understanding of language, because they provide a natural measure of how linguistic knowledge is accessed and used in real-time interpretation.

Many initial studies demonstrate that, like adults, children rapidly access and use their linguistic knowledge in real-time processing, so long as they know the relevant words and structures. For example, Fernald, Swingley, and colleagues have shown that reference to an object with a known name (e.g., *ball*) results in shifts in direction of gaze to that object within 600–700 ms of the name's onset, even in children as young as 24 months (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). More recent research has explored the extent to which there is continuity in lexical processing over the course of development. For instance, the parallel consideration of lexical candidates appears to be a fundamental property of the spoken language comprehension system even at its earliest stages of development. Swingley et al. (1999) provided 24-month olds with spoken instructions to look at a particular object (e.g., *Look at the tree*) in the presence of either lexical cohort competitor (pictures of a tree and a truck) or some other object (pictures of a tree and a dog). Like Allopenna et al.'s (1998) adult subjects, toddlers showed temporary consideration of both the target and the cohort competitor early in the perception of the word, which resolved toward the target soon after the word's offset (also see Swingley & Aslin, 2002). Consideration of the alternative object did not occur when its name and the target name were not cohorts. These results demonstrate that the developing word-recognition system makes use of fine-grained phonemic contrasts, and from the start is designed to interface this linguistic knowledge (how the word sounds, what the word means) with knowledge about how the word might plausibly behave referentially when making contact with the ambient world.

Other work has begun to examine the development of sentence parsing abilities using eye gaze measures. This research began with studies conducted with five- and eight-year-olds, first reported in Trueswell et al. (1999) that were modeled after the adult "apple-on-the-towel" studies described earlier (Tanenhaus et al., 1995; Spivey et al., 2002). Here children's eye movements were recorded using a lightweight visor system as they acted upon spoken instructions that contained temporary ambiguities such as *Put the frog on the napkin in the box*.

The striking finding was that five-year olds showed a strong preference to interpret *on the napkin* as the Goal of *put*, even when the referential scene supported a Modifier interpretation. Upon hearing *on the napkin*, five-year olds typically looked over to a potential Goal in the scene, the empty napkin, regardless of whether there were two frogs present (supporting a modifier interpretation) or one frog present (supporting a Goal interpretation). The timing of these eyemovements were similar to those observed in the 1-Referent condition of adults, i.e., ~600 ms after the onset of the word "napkin," but for children

this pattern of Goal-looks also arose in 2-Referent contexts. In fact, five-year olds' preference for the Goal interpretation was so strong that they showed little sign of revising it; upon hearing *napkin* children looked to the empty napkin as a potential goal, and then frequently moved a frog to that location. In two referent cases, children were equally likely to move the frog that was on the napkin and the frog that was not on the napkin, suggesting they never considered a Modifier interpretation. Importantly, this child-parsing behavior was localized to the ambiguity, and not to the complexity of the sentence. Five-year olds' eye movements and actions became adult-like when the temporary ambiguity was removed, as in the unambiguous modifier form, *Put the frog that's on the napkin in the box*. The nearly perfect performance with unambiguous sentences rules out a potentially mundane explanation of the results, namely that long "complicated" sentences confuse young children. Here an even longer sentence with the same intended structure does not cause difficulty, precisely because the sentence lacks the temporary ambiguity.

Both the Swingley et al. (1999) and Trueswell et al. (1999) results demonstrate that there is considerable continuity in the language-processing system throughout development: lexical and sentential interpretation proceeds incrementally and is designed to coordinate multiple information sources (e.g., linking what is heard to what is seen within milliseconds). However, the differences between five- and eight-year-old children reported by Trueswell et al. (1999) suggest that significant developmental differences do exist. These differences likely pertain to how children learn about sources of evidence pertinent to linguistic and correlated non-linguistic constraints. Highly reliable cues to structure, such as the argument-taking preferences of verbs, are learned earlier than other sources of evidence that may be less reliable or more difficult to discover, because they involve more subtle contingencies. Consistent with this hypothesis, Snedeker and Trueswell (2004) report that young children are more sensitive to verb bias manipulations than to the number of potential referents in the display. Interestingly, children of the same age do appear to be sensitive to referential constraints under some conditions, especially when the discourse guides the child toward the correct referential contrast (see Trueswell & Geltman, 2004; Trueswell, Papafragou & Choi, in press, for further discussion). Moreover, children are also sensitive to at least some aspects of speaker perspective. Nadig and Sedivy (2003) demonstrated that 5-year-old children distinguish between common ground and privileged ground in a simplified version of the task used by Keysar et al. (2000).

#### 4. CLOSING REMARKS

This chapter has provided an overview to the rapidly growing literature on eye movements and spoken language processing, focusing on applications to spoken language comprehension. We have reviewed some of the foundational studies, discussed issues of data analysis and interpretation, and discussed issues that arise in comparing eye-movement reading studies to visual world studies. We have also reviewed some of the major lines of research that are utilizing this method, focusing on topics in language comprehension, including spoken word recognition, use of referential constraints in parsing, interactive conversation, and the development of language processing abilities in children.



It should be clear from this review that the visual world paradigm is being employed in most traditional areas of inquiry within psycholinguistics. And in each of these areas, the visual world approach is encouraging psycholinguists to investigate uncharted theoretical and empirical issues. Within the study of spoken sentence comprehension, issues about reference have taken center stage, in part because the visual world methodology makes it possible to connect research on real-time reference resolution with social and cognitive research on pragmatics and conversation. Within the study of spoken word recognition, the time-locked nature of this measure has allowed researchers to explore phonemic and sub-phonemic and prosodic contributions of word recognition in utterances at a level of detail previously not possible with traditional methods. It is for these reasons and other reasons we are quite optimistic that eye-movement measures will continue rise in interest and use within the psycholinguistics community.

We close by noting that eye-movement measures are likely to be most powerful when combined with other measures. We have seen how combining eye movements with action and structure tasks can shed new light on real-time language processes. We expect that other measures will emerge that provide additional advantages. For instance, other body movements pertaining to gestures and actions are likely to be highly informative when connected to the timing of speech and eye gaze events. Most generally, we see the visual world approach as part of a larger movement toward connecting language and action in rich goal-directed tasks using increasingly rich and complex data arrays to understand the dynamics of comprehension and production in conversation. This approach is likely to have an increasingly important influence on theoretical development in natural language, just as it has begun to enrich theories in other areas of perception and cognition (Ballard, Hayhoe, Pook, & Rao, 1997; Barsalou, 1999; Hayhoe & Ballard, 2005; Land, 2004).

## ACKNOWLEDGMENTS

This work was supported by NIH grants DC005071 and HD 27206 to MKT and HD 37507 to JCT.

## REFERENCES

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't. Mediating the mapping between language and the visual world. In: J. M. Henderson, & F. Ferreira (Eds), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 279–318). New York: Psychology Press.

- Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Arnold, J. E., Brown-Schmidt, S., Trueswell, J., & Fagnano, M. (2005). Children's use of gender and order of mention during pronoun comprehension. In: Trueswell, J. C., & Tanenhaus, M. K. (Eds), *Processing world-situated language: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course for pronoun resolution from eyetracking. *Cognition*, 76(1), B13–B26.
- Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, 15(9), 578–582.
- Bailey, K. G. D. & Ferreira, F. (2005). Don't swim, hop: The timecourse of disfluency processing. Paper to be presented at the 18th annual meeting of the CUNY Conference on Human Sentence Processing, Tucson, AZ.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723–767.
- Barsalou, L. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, 28, 61–80.
- Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In: J. M. Henderson, & F. Ferreira (Eds), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
- Bock, J. K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, 48, 653–685.
- Boland, J. E. (2005). Visual arguments. *Cognition*, 95(3), 237–274.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A., (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25.
- Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding*. Unpublished doctoral dissertation. Stanford University.
- Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In: J. C. Trueswell, & M. K. Tanenhaus (Eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
- Brennan, S. E., & Hulstijn, E. (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, 8, 143–151.
- Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33, 251–283.

- Brown-Schmidt, S. (2005). Language processing in conversation. Unpublished Doctoral dissertation, University of Rochester.
- Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, 53(2), 292–313.
- Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2005). Real-time reference resolution in a referential communication task. In: J. C. Trueswell, & M. K. Tanenhaus (Eds), *Processing world-situated language: Bridging the language-as-action and language-as-product traditions*. Cambridge, MA: MIT Press.
- Brown-Schmidt S., & Tanenhaus, M. K. (2005). Real-time interpretation of referential expressions in unscripted interactive conversations, submitted for publication.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, 54, 592–609.
- Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.
- Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In: A. K. Joshi, B. Webber, & I. Sag (Eds), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, 47(1), 30–49.
- Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Action-based affordances and syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 687–696.
- Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, 3, 115–120.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Crain, S. (1981). *Contextual constraints on sentence Comprehension*. Ph.D. Dissertation, University of California, Irvine.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In: D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320–358). Cambridge, UK: Cambridge University Press.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42(4), 317–367.

- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5–6), 507–534.
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 498–513.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12(3), 453–459.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47(2), 292–314.
- Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (in press). The influence of visual processing on phonetically driven saccades in the “visual world” paradigm. In: R. P. G. van Gompel, Fischer, M. H., Murray, W. S., & Hill, R. L. (Eds), *Eye movements: A window on mind and brain*. Oxford: Elsevier.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27(4), 429–446.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231.
- Ferreira, F., & Bailey, K. G. B. (in press). The processing of filled pause disfluencies in the visual world. In: R. P. G. van Gompel, Fischer, M. H., Murray, W. S., & Hill, R. L. (Eds), *Eye movements: A window on mind and brain*. Oxford: Elsevier.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: Bradford Books.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
- Gow, D. W., & McMurray, B. (in press). Word recognition and phonology: The case of English coronal place assimilation. *Papers in Laboratory Phonology*, 9.
- Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science*, 15, 814–821.
- Griffin, Z. M., & Bock, J. K. (2000). What they eyes say about speaking. *Psychological Science*, 11, 274–279.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29, 261–290.

Grodner, D., & Sedivy, J. (in press). The effect of speaker-specific information on pragmatic inferences. In: N. Pearlmuter, & E. Gibson (Eds), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 33, 101–123.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49, 43–61.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 188–194.

Henderson, J. M., & Ferreira F. (2004). *The interface of language, vision, and action: Eye movements and the visual world*. Edited volume. New York: Psychology Press.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16, 260–264.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15, 314–318.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94, 113–147.

Kaiser, E., & Trueswell J. C. (in press). The referential properties of Dutch pronouns and demonstratives: Is salience enough? In: M. Weisgerber (Ed.) *Proceedings of Sinn und Bedeutung* 8. University of Konstanz linguistics working papers.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in Incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.

Keysar, B. & Barr, D. (2005). Coordination of action and belief in conversation. In: J. C. Trueswell, & M. K. Tanenhaus (Eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32–38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25–41.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30, 481–529.

- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in the depicted events. *Cognition*, *95*(1), 95–127.
- Kowler, E. (1995). Eye movements. In: S. M. Kosslyn, & D. N. Osherson (Eds), *Visual cognition: An invitation to cognitive science* (2<sup>nd</sup> ed., Vol. 2). Cambridge, MA: MIT Press.
- Land, M. (2004). Eye movements in daily life. In: L. Chalupa, & J. Werner (Eds), *The visual neurosciences: Vol. 2* (pp. 1357–1368). Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419.
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*, 6–14.
- Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition*, *27*, 385–398.
- Luce, D. R. (1959). *Individual choice behavior*. Oxford: Wiley.
- Lyons, J. (1981). *Language, meaning and context*. London: Collins/Fontana.
- Magnuson, J. S. (2002). *The microstructure of spoken word recognition*. Unpublished dissertation, University of Rochester.
- Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, *102*, 9995–9996.
- Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (in press). Which words compete? The dynamics of similarity during spoken word recognition. *Cognitive Science*.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). Time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*, 202–227.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 77–102.
- Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In: G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: The MIT Press.
- Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In: G. T. M. Altmann, & R. Shillcock (Eds), *Cognitive models of speech processing: The second Sperlonga meeting*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*, 365–368.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33.
- Noveck, J., & Sperber, D. (Eds). (2005). *Experimental pragmatics*. Oxford: Oxford University Press.
- Pickering, M. J., & Garrod, S. C. (2004). Towards a mechanistic theory of dialog. *Behavioral and Brain Sciences*, *7*(2), 169–190.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In: P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds), *Intentions in communication* (pp. 271–311). Cambridge, MA : MIT Press.
- Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin*, *124*, 372–422.
- Rayner, K., & Liversedge, S. P. (2004). Visual and linguistic processing during eye fixations in reading. In: J. M. Henderson, & F. Ferreira (Eds), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
- Rosson, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.
- Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements. *Cognition*, *89*, B1–B13.
- Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Assigning referents to reflexives and pronouns in picture noun phrases. Experimental tests of binding theory. *Cognitive Science*, *30*, 1–49.
- Salverda, A. P. (2005). Prosodically-conditioned detail in the recognition of spoken words. Unpublished Ph.D. dissertation, Max Planck Institute for Psycholinguistics, Nijmegen.
- Salverda, A. P., & Altmann, G. (2005). Cross-talk between language and vision: Interference of visually-cued eye movements by spoken language. Poster presented at the AMLaP Conference, Ghent, Belgium.

- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.
- Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*, 3–23.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.
- Sereno, S. C., & Rayner, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 173–184.
- Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, *96*(2), 316–340.
- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.
- Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods Instruments & Computers*, *28*(4), 516–536.
- Song, H., & Fisher, C. (2005). Who’s “she”? Discourse prominence influences preschoolers’ comprehension of pronouns. *Journal of Memory and Language*, *52*, 29–57.
- Speer, S. R., & Ito K. (in press). Using interactive tasks to elicit natural dialogue production. In: I. Mleinek (Ed.), *Methods in empirical prosody research*. Berlin: Mouton de Gruyter.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*, 10393–10398.
- Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281–284.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1521–1543.
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M. & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.
- Spivey-Knowlton, M. J. (1996). *Integration of visual and linguistic information: Human data and model simulations*. Ph.D. dissertation, University of Rochester.
- Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267.
- Stone, M., & Webber, B. (1998). Textual economy through close coupling of syntax and semantics. In: *Proceedings of INLG*, (pp. 178–187) Niagara-on-the-Lake, Ontario, Canada.



Sussman, R. S. (2006). *Processing and representation of verbs: Insights from instruments*. Unpublished dissertation, University of Rochester.

Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.

Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71(2), 73–108.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior*, 18(6), 645–659.

Tanenhaus, M. K., & Brown-Schmidt, S. (in press). Language processing in the natural world. In: B. C. M. Moore, L. K. Tyler, W. D. Marslen-Wilson (Eds), *The perception of speech: from sound to meaning. Philosophical Transactions of the Royal Society B: Biological Sciences*.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal Behavior*, 18(4), 427–440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 286, 1632–1634.

Trueswell, J. C., & Gleitman, L. (2004). Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In: J. M. Henderson, & F. Ferreira (Eds), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Trueswell, J. C., Papafragou, A., & Choi, Y. (in press). Syntactic and referential processes: What develops? In: E. Gibson, & N. Perlmutter (Eds), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89–134.

Trueswell, J. C., & Tanenhaus, M. K. (Eds), (2005). *Processing world-situated language: Bridging the language-as-action and language-as-product traditions*. Edited volume. Cambridge, MA: MIT Press.

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects. In: E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes: Vol 4. Reviews of oculomotor research* (pp. 253–393). Amsterdam: Elsevier.

Watson, D., Gunlogson, C., Tanenhaus, M. K. (in press). Online methods for the investigation of prosody. In: I. Mleinek, (Ed.), *Methods in empirical prosody research*. Berlin: Mouton de Gruyter.

Yee, E., Blumstein, S., & Sedivy, J. C. (2000). The time course of lexical activation in Broca's aphasia: Evidence from eye movements. Poster presented at the 13th annual CUNY conference on human sentence processing, La Jolla, CA.

## Chapter 23

### **Perspective Taking and the Coordination of Meaning in Language Use**

*Dale J. Barr and Boaz Keysar*

#### **1. INTRODUCTION**

Many of the words used to talk about language use – e.g., communication, coordination, convention, conversation, cooperation, collaboration, community – come from the Latin stems *co-* or *con-*, which signify “with” and “together”, respectively (Kennedy, 1890/1996). This confirms that our Latin-speaking predecessors thought of language as part of a social activity that people do “together with” other people. Regrettably, throughout the history of modern psycholinguistics, this appreciation for the social dimension of language has often been absent. Instead of studying how people achieve shared understanding in conversation, psycholinguists have traditionally studied the cognitive structures and processes that underlie the production or comprehension of words or sentences in social isolation. Although a worthwhile enterprise, the development of ecologically valid theories requires greater attention to the cognitive activities involved in actual conversation.

While psycholinguistic research has regarded language as a cognitive skill, language use research has focused on the collaborative dimensions of conversation, often ignoring the cognitive basis of speaking and understanding. This tends to result in theories based on descriptive principles rather than explanatory mechanisms. Fortunately, these two traditions are starting to merge, auguring the development of a theory of language processing that is explanatory, cognitively plausible, and ecologically valid. In this chapter, we review research that lays the groundwork for such a theory.

#### **1.1. Overview and Scope**

Language use requires coordination to be successful, because language can be ambiguous. Thus, speakers and listeners will often have to adapt their behavior to their interlocutor in order to avoid misunderstanding. A central question of this chapter is how such adaptations are achieved during the production and comprehension of language.

The specific hypothesis that we investigate in this chapter is the *audience design hypothesis*, or the *design hypothesis* for short. This hypothesis assumes that speakers and listeners achieve success in communication because they maintain detailed models of what the other person knows, and speak and understand against these models (Clark & Murphy, 1982). Given the quantity of evidence currently available, it seems undeniable that speakers adapt some aspects of their speech to characteristics of their listeners. Adults often speak to small children using “motherese,” a form of speech containing exaggerated prosody, a higher pitch, and simplified syntax (Snow, 1972). Speakers change their speaking register or style depending upon the social identity of their addressees (Bell, 1984). Casual observation reveals that bilinguals will often mix languages when talking to other bilinguals, but tend to use a single language when speaking to monolinguals. And developmental psychologists such as Piaget (1926/1955) have observed that the speech of children becomes less egocentric and more listener-centered as they mature (see also Flavell, Botkin, & Fry, 1968).

There is extensive evidence from social psychology that people’s speech varies according to their assumptions about the recipient (for a review, see Krauss & Fussell, 1991). When speakers describe an abstract figure so that another person will be able to identify it, they use longer descriptions that are easier for others to understand than the descriptions that they produce for themselves (e.g., Danks, 1970; Fussell & Krauss, 1989a; Krauss, Vivekananthan, & Weinheimer, 1968; Werner & Kaplan, 1963). When they are asked to describe a figure for a specific friend, the friend is slightly better at interpreting that description than a stranger for whom the message was not intended (Fussell & Krauss, 1989b). Speakers also have some ability to convey a message to an intended audience while concealing it from overhearers (Clark & Schaefer, 1987; Fleming & Darley, 1991). Finally, it has been noted that the lower the perceived identifiability of a stimulus among the members of a social group, the more information speakers will provide when describing that figure to a member of that group (Fussell & Krauss, 1992). All of these findings indicate that speakers take characteristics of the listener into account when they speak.

Therefore, the question is no longer *whether* speakers design their speech for their listeners, but *how* and *when* they do so. In this review, we focus on research that addresses the following questions:

- How fully integrated into the language-processing system is perspective taking? Which levels of processing does it constrain, and which not?
- Does audience design figure into the initial design of a speaker’s utterance, or do its effects arise through monitoring?
- How do listeners exploit the assumption of audience design when they comprehend language?
- Can mechanisms other than perspective-taking mimic the effects of audience design?

Our review begins by introducing the notion of *common ground*, a form of knowledge that has been assumed to provide the basis for audience design. Then, we provide

background for our review in terms of a set of theoretical and methodological preliminaries. The review itself is divided into three main sections: one on language production, one on language comprehension, and one on repeated reference. We conclude by discussing the extent to which the body of evidence supports various models for the use of common ground in language processing.

## 1.2. Establishing Common Ground

Sometime in your life, you might find yourself seated in a plane next to a guy who you recognize as member of the same secret society as you are. As joint members of the society, you would both have extensive shared knowledge: the secret handshake, the mission and philosophy of your group, the time and location of your annual meeting, the identity of the society's president, and so on. However, unless you knew that your companion also recognized you as a member of this secret community, there would be no reason for you to use this knowledge in interpreting what he says. Thus, when your companion starts a conversation by saying *I hear that the president will give a speech today*, you might misunderstand him if you interpret *the president* to mean the president of your secret society. Because you lack evidence that he knows that you are also a member of the same secret society, the knowledge is not mutual. As a consequence, you might be better off assuming that he is referring to the president of the country you are both citizens of.

Clark and colleagues (e.g., Clark, 1992; Clark & Marshall, 1981) have proposed that a special kind of mutual knowledge, or *common ground*, provides the critical background against which speakers produce utterances and listeners comprehend them. In this framework, speakers strive to optimally design their speech with respect to the common ground: "The speaker designs his utterance in such a way that he has good reason to believe that the addressees can readily and uniquely compute what he meant on the basis of the utterance along with the rest of their common ground" (Clark, Schreuder, & Buttrick, 1983, p. 246). Likewise, it is assumed that listeners make the assumption that speakers' utterances are optimally designed, and use this assumption to guide how they interpret the speech (Clark & Carlson, 1981; Clark et al., 1983). Although mutual knowledge per se may never be attainable in practice, common ground can be inferred based on the following types of evidence (Clark & Marshall, 1981):

*Physical copresence.* Information that is physically copresent is perceptually available to both interlocutors, who can also perceive that they can both perceive it. Physical or perceptual copresence is regarded as the strongest evidence for common ground. The standard example is of a candle that is placed on a table between two interlocutors. Both can see the candle, and both can also see that they both see it.

*Linguistic copresence.* Things that are said during the course of a conversation become part of a "discourse record" that is a subset of interlocutors' common ground. Thus, if I tell you *Yesterday, I met a man who is an Amway salesman*, I can later refer to him using a simpler description such as *the Amway guy* or simply the pronoun *he*, and you can retrieve the appropriate referent from the discourse record.

*Community membership.* As in the “secret society” example, interlocutors can infer common ground based on their joint membership in certain communities, once there is something to make this joint membership mutually evident.

### 1.3. Using Common Ground during Language Processing

A theory of language use in dialogue must describe not only how interlocutors establish common ground, but also how that knowledge is used in processing language. Here are four main contending models:

*Full constraint.* This model adopts the strongest stance regarding the use of common ground in language processing; thus, it lies at one extreme of the spectrum of possible theories. Under full constraint, all levels of language production or comprehension have full access to common ground, and common ground imposes an immediate, obligatory, and complete constraint on the operation of each level. This account predicts that information that is available to the speaker or listener but that is not part of their common ground would have no impact on language processing.

*Partial constraint.* This model assumes that common ground does not fully constrain processing, but is one of many cues integrated in parallel during production or comprehension (Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002). The probability that common ground will be used will depend on its salience and cue validity relative to other cues. Although the constraint of common ground on processing is partial rather than full, the constraint applies to all levels of production or comprehension from the earliest moments of processing. Partial constraint predicts that private information will have some effect on processing; however, all other things being equal, it should have less of an effect than information in common ground.

*Multiple-systems accounts.* Multiple-systems accounts assume that the use of common ground is a controlled, optional process, and that it is therefore subject to restrictions on cognitive capacity. The two main existing multiple-systems accounts make different assumptions about how common ground is integrated into language production and comprehension.

One of these accounts, the *perspective adjustment* model of Keysar and colleagues (Epley, Keysar, van Boven, & Gilovich, 2004; Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Paek, 1998) assumes that common ground is a functionally distinct process that belongs to an “adjustment” stage of processing, but that it imposes no constraint on production or comprehension processes per se. Instead, it assumes that these core stages of language processing are conducted egocentrically, that is, by using information available to the self regardless of whether it is part of common ground. The adjustment stage, which is optional, and slow to take effect, uses common ground to detect and correct violations of common ground. For production, perspective adjustment assumes that utterances are initially designed egocentrically, and

common ground is used to bring utterances into line with the listener's perspective. For comprehension, it contends that listeners initially interpret utterances egocentrically and only later monitor for, and correct, violations of common ground.

Another multiple-systems account is what we call the *dual-process* model, formulated for language production by Bard et al. (2000) and Rossnagel (2000). Instead of assuming that the core set of language processes operate egocentrically with a common-ground-sensitive monitor, these accounts make a distinction between automatic and controlled processes in the language-processing system. The model assumes that some language processes operate automatically in a manner that is insensitive to common ground, while others are controlled, and therefore, can be influenced by common ground, given sufficient cognitive resources. For instance, Bard and Aylett (2005) argue that whether a speaker articulates a word more or less intelligibly is determined by an automatic process (priming), while the decision of whether to mark a reference using the indefinite article *a* versus the definite article *the* is a controlled process that is constrained by common ground.

*Ordinary memory view.* What we call the ordinary memory view (Gerrig & McKoon, 1998; Horton & Gerrig, 2005) is less a theory about the use of common ground than a denial of the psychological reality of common ground. In this model, the effects of common ground are fully reduced to everyday memory processes. As Horton and Gerrig (2005) state: "In general, we wish to 'normalize' audience design as a straightforward consequence of how partner-specific information is encoded and retrieved during routine utterance planning ... people will perform utterances that show evidence of audience design whenever the memory representations that permit audience design become accessible with the appropriate time course." (p. 129). Thus, this theory assumes that knowledge of mutual information is not "consulted" during conversation, but instead, successful coordination emerges naturally from low-level memory processes.

## 2. THEORETICAL AND METHODOLOGICAL PRELIMINARIES

For an explanatory theory of audience design, it is essential to view the problem of coordinating mutual understanding at three levels of analysis: the cultural, the interactional, and the cognitive levels (Barr & Keysar, 2005). We suggest that this three-level framework is important because it highlights the possibility of multiple explanations for the same phenomenon. In this section, we define these three levels and discuss their implications for the study of audience design.

*The cultural level.* Successful communication depends on the existence of overlapping cognitive representations among a community of language users. This overlap exists because of the evolution and diffusion of conventions. Conventions can come into existence as a by-product of language users' efforts to achieve mutual understanding with specific partners (Barr, 2004; Garrod & Doherty, 1994; Hutchins & Hazlehurst, 1995; Steels, 1996). This is consistent with findings that suggest that meaning coordination

during conversation result in enduring changes to how people represent and/or talk about referents (Garrod & Doherty, 1994; Malt & Sloman, 2004; Markman & Makin, 1998). Social network theory suggests that even very large communities can be organized in the form of “small worlds,” where the paths between any two arbitrary individuals can be traversed in a small number of steps (Travers & Milgram, 1969; Watts, 1999). Such short paths enable the rapid diffusion of information throughout the community.

*The interactional level.* Although sharing representations goes some way toward reducing ambiguity, language users will rarely find perfect overlap in their cognitive representations. Thus, interlocutors will often find it necessary to adapt prevailing linguistic conventions to suit their current needs. At the interactional level of analysis this adaptation is achieved via collaborative processes. This may include consulting a model of the interlocutor’s knowledge, monitoring feedback from one’s interlocutor, and making adjustments in response to this feedback. The collaborative model of Clark and colleagues (Clark, 1992, 1996) provides an elaborate theory of the interactional level. Face-to-face conversations provide multiple channels of communication beyond words – including gesture, eye gaze, and prosody – that can reduce the ambiguity of speech. These channels form part of a “collateral track” of communication through which listeners can provide concurrent evidence of understanding without disrupting the current speaker’s turn at talk (Clark, 1996).

*The cognitive level.* The cognitive level of analysis focuses on the mental machinery that underlies language processing in conversation, which includes processes specific to language (e.g., grammatical and phonological encoding), as well as more general cognitive processes (e.g., memory, attention, decision-making). Interlocutors’ ability to successfully communicate will be constrained by the nature and limitations of these underlying processes.

We have highlighted this three-tiered framework because it implies that there will typically be more than one explanation for the behavior of speakers and listeners. Explanations couched at the cultural level will invoke the existence of overlapping representations, dominant language practices, frequency of usage, and so on. They may also invoke language users’ mental models of the capabilities and preferences of the typical community member. Explanations at the interactional level will invoke collaboration, feedback, and the use of common ground. Finally, explanations at the cognitive level will invoke cognitive availability, access to semantic or episodic representations, priming, attention, cognitive capacity, and so on.

The existence of multiple explanations makes language use in dialogue a notoriously difficult subject to study. A good example of how more than one explanation could be provided for the same phenomenon can be seen by considering structural repetition in discourse (Levelt & Kelter, 1982). When speakers say something that follows the same syntax as that of a previous speaker, they may do this because: (1) it is the dominant way to say this in their language (cultural level); (2) they wish to reduce ambiguity for the listener or express agreement with the previous speaker (interactional level); or (3) because

structural priming simply made this particular structure temporarily more available to the production system (cognitive level). Researchers tend to give priority to certain levels of explanation over others, consistent with their pretheoretical commitments. Thus, a psycholinguist would be most likely to explain this phenomenon in terms of structural priming, whereas a researcher working in a collaborative framework would view it as a way of expressing agreement or reducing ambiguity for one's interlocutor.

In the study of audience design, priority has been given to the interactional level. This has led to the following misconception: if a speaker produces an utterance that is consistent with the common ground with the listener, then that speaker must have consulted common ground in the course of designing the utterance. That speakers modulate their speech depending on context, and that these modulations tend to make their speech easy for listeners to understand, would seem to provide excellent evidence for audience design. However, changes in common ground also change the information available to the speaker (Keysar, 1997). For example, speakers might use the description *circle* to refer to the round object in a display containing one circle and one square, but would tend to call it *small circle* in a display containing a larger circle (Olson, 1970). Such behavior is in line with what one would expect from a cooperative speaker (Grice, 1975). However, it is also consistent with the idea that people categorize objects differently in different contexts. Even without any intention to speak, a person looking at such displays may mentally categorize the object as a circle in one display and as a small circle in the other display.

Dell and Brown (1991) were the first to observe that audience design is not the only explanation for speakers' adaptations. They distinguished between two kinds of adaptations: *generic-listener adaptations* and *particular-listener adaptations*. Generic-listener adaptations would benefit comprehension for a generic listener. These adaptations can be made by consulting a model of the generic listener in the language community, or could simply be a by-product of parallelism between language production and language comprehension. In contrast, particular-listener adaptations are based on common ground with the listener.

Dell and Brown illustrate this with an example of the choice of the word *lie* versus *prevaricate* to refer to the telling of an untruth. Speakers might choose *lie* over *prevaricate* because they estimate that their listener would be unlikely to know the meaning of *prevaricate*; this choice would reflect a particular-listener adaptation. However, speakers use the word *lie* over *prevaricate* simply because it is more accessible to them due to its greater frequency of usage. Precisely because *lie* is more frequent, it would also be easier for the listener to comprehend. Thus, this choice would be a *generic-listener adaptation*; specifically, in this case it would be a by-product of how the lexicon is organized rather than the outcome of an assessment of the listener's knowledge.

Given the possibility of multiple explanations for a given phenomenon, it is necessary to deconfound effects that are due to the speaker's perspective from the effects that are due to the speaker's beliefs about the listener's perspective (Keysar, 1997). To achieve



such separation of perspectives, at least two strategies are appropriate: (1) manipulate the speaker's perspective while holding constant the speaker's beliefs about common ground; (2) manipulate the speaker's beliefs about common ground while holding constant the speaker's perspective. The first of these two options is often accomplished through the introduction of "private" speaker knowledge. The second option can be implemented by manipulating a speaker's beliefs about what the listener knows. However, it could also inadvertently affect the availability of information to the speaker (Keysar, 1997).

A final caution in the study of conversational coordination involves a distinction between *self-prompted* and *other-prompted* adaptations. Self-prompted adaptations are those that speakers generate spontaneously, even without the addressee's intervention. In contrast, other-prompted adaptations are due to the intervention of the addressee. Sometimes in experiments on language production, speakers converse with a naïve participant. While this is intended to safeguard ecological validity, uncontrolled listener feedback can cloud interpretation by making it difficult to distinguish between self- and other-prompted adaptations. The only way to distinguish between them is to control for the behavior of the addressee, either by using a trained confederate (Hanna et al., 2003; Keysar et al., 2000; Metzger & Brennan, 2003), a "virtual partner" (Barr & Keysar, 2002), or an imaginary partner (Schober, 1993). Experiments that lack such a control might gain ecological validity, but cannot inform us about the nature of speaker adaptations.

Although the above concerns apply to language production, they are also relevant to language comprehension. Listeners may interpret utterances in line with common ground not because they consulted their common ground with the speaker, but because they used available information that happened to be in common ground. Thus, it is important to design comprehension experiments that either manipulate the information available to the listener while holding common ground constant, or manipulate the common ground while holding the information available to the listener constant. Furthermore, listeners' adaptations may be self-generated or prompted by speaker feedback. The remedies suggested above also apply to comprehension studies.

### **3. CURRENT STATUS OF THE DESIGN HYPOTHESIS IN LANGUAGE PRODUCTION**

Speakers have many opportunities to take a listener's needs into account – when they decide what to say, when they formulate their linguistic utterance that they will produce, and when they monitor their utterance plans or their overt speech. In this section, we discuss the evidence for and against audience design at each of these stages.

#### **3.1. Determining What to Say**

When listeners comprehend narratives they fill in missing information by accessing long-term memory structures such as scripts, schemata, or frames (Minsky, 1974; Schank

& Abelson, 1977). For example, in comprehending a passage describing a stabbing incident, comprehenders might draw the inference that a knife was used. A study by Brown and Dell (1987) examined whether speakers provide information precisely when the comprehender is less likely to infer it from background knowledge. In their study, participants read and then retold to a confederate, short narratives that described a character performing some action (e.g., stabbing). The action involved an instrument that was either typical (e.g., knife) or atypical (e.g., icepick) for that action. Brown and Dell found that speakers were about twice as likely to mention an atypical than a typical instrument in their retellings, suggesting a possible listener adaptation.

However, the results could also be a by-product of how lexical knowledge is represented and accessed during production. For example, in accessing the verb *stab*, the production process might note that the instrument, an icepick, conflicted with the instrument conventionally used for this action, and therefore might tag the instrument to be explicitly mentioned. Supporting this idea, the majority of explicit mentions of atypical instruments (52%) were linguistically packaged in the same syntactic clause as the verb, in an immediately following prepositional phrase (e.g., *he stabbed the man with an icepick*). In contrast, only 33% of the explicit mentions of typical instruments were in this location. The explicit mention of atypical instruments in the verb clause was not sensitive to the needs of the listener: even when the listeners had already been informed about the instrument, speakers mentioned the atypical instrument at the same rate. However, whether or not the listener was informed did have an impact on the likelihood of explicit mention in a separate clause after the verb (e.g., *The robber stabbed the man. He used an icepick*). Notably, when the listener was uninformed about the instrument's identity, speakers were simply more likely to mention *any* instrument – typical or atypical – in this location, which might be considered a relatively coarse particularized adaptation. Because the effects occur late in the sentence (in a separate clause after the verb), they are likely to result from a monitoring process – that is, as a kind of afterthought that is appended to the original preverbal message and which did not figure into the initial design of the utterance (Brown & Dell, 1987).

Lockridge and Brennan (2002) challenged this conclusion, arguing that the informational needs of actual addressees would have been different from those of the confederates in Brown and Dell (1987). The confederates, who had heard the stories many times over, may have provided feedback that indicated that they were already knowledgeable about the narratives. Indeed, Lockridge and Brennan found that when naïve listeners lacked pictures, speakers were about 15% more likely to mention atypical than typical instruments within the verb clause. This suggests that speakers are able to make use of listener knowledge in their production of speech. However, it is still possible that explicit mention of the atypical instrument was not included in the original design of the utterance, but was edited into the verb clause by the monitor. Given that speakers and addressees interacted freely, speakers would have had repeated opportunities to learn about their addressees' specific needs, and may have begun to monitor their speech more carefully before articulation. These results might be reconciled with those of Brown and Dell by assuming that the mention of atypical instruments is a spontaneous, generic

adaptation, but may also constitute a particularized adaptation when the speaker receives adequate listener prompting.

This possibility seems viable given evidence from Horton and Keysar (1996) that certain adaptations may be accomplished via monitoring and adjustment rather than through incorporation into the initial utterance plan. In their study, speakers described a target object to a listener. The target object (e.g., circle) appeared in the presence of a “context” object (e.g., a larger circle) that was either shared with the listener or private to the speaker. Adapting to the listener’s needs, speakers’ target object descriptions were more likely to include adjectives related to the context object (e.g., *small circle*) when the context object was shared compared to when it was private (75% versus 24%). However, when placed under pressure to begin their descriptions quickly, speakers showed no such sensitivity to the listener’s perspective. Assuming that time pressure obstructs later, monitoring-type processes, this result demonstrates that information about the interlocutor’s perspective is not taken into account during the design of an utterance, but during an additional monitoring stage. (See Polichak & Gerrig, 1998, for a dissenting view, and Keysar & Horton, 1998, for a rebuttal.)

Rossnagel (2000) argued that such adaptation in speaking might be the outcome of both automatic and controlled processes. Automatic processes of message planning make use of cognitively available information, while controlled processes allow for the tailoring of the plan to the addressee’s perspective. As with perspective adjustment, this automatic-controlled account predicts that utterances will be less adapted to the listener’s needs when the speaker is under cognitive load. In Rossnagel’s study, speakers described the assembly of a machine model to an adult university student or to a seven-year old boy, whose limited knowledge would require special tailoring. It was found that cognitive load impaired speakers’ ability to take the addressee’s perspective into account. While speakers varied their instructions depending upon the identity of the addressee in the low-load condition, they gave similar instructions to the student and to the young boy under high load. These results complement those of Horton and Keysar (1996) in support of the monitoring and adjustment hypothesis.

### 3.2. Avoiding Syntactic Ambiguity

A related question is whether or not speakers take steps to make their speech easier for comprehenders to parse. Speakers can avoid syntactic ambiguity by using optional words, by ordering the constituents in a sentence, or by marking syntactic breaks using prosody.

#### 3.2.1. Structural disambiguation

Most people would initially experience difficulty in parsing the string *the coach knew you ate* upon arriving at the word *ate*. This is because the comprehension system prefers to interpret *you* as the direct object of the verb *knew* rather than as the subject

of an embedded sentence. Speakers could circumvent such “garden path” difficulties by including optional words such as the complementizer *that*, as in *the coach knew that you ate*. Ferreira and Dell (2000) investigated whether the inclusion of such optional words reflected audience design. They contrasted sentences where insertion of the word “that” would prevent an ambiguity (e.g., *the coach knew you ...*) with sentences where there was no ambiguity to prevent (e.g., *the coach knew I ...*). Ferreira and Dell found that speakers are just as likely to include optional words when it would prevent ambiguity as when it would serve no such function (for a similar result, see Kraljic & Brennan, 2005). They found that speakers were only more likely to use optional words when material in the embedded clause was less readily available in memory, a speaker-oriented factor that is independent of the listener’s needs.

The ordering of syntactic constituents is another device that speakers could use to avoid ambiguity. Arnold, Wasow, Asudeh, and Alrenga (2004) investigated whether speakers would use constituent ordering to disambiguate how a prepositional phrase should be attached to the syntactic structure. Such attachment ambiguities arise in sentences containing more than one prepositional phrase, e.g., *The judge sent the letter to the president to the committee*. Only at the end of this sentence does it become clear that the letter was not sent to the president, but to the committee. Speakers could help listeners avoid such garden paths by positioning the goal argument directly after the verb, as in, *The judge sent to the committee the letter to the president*. Arnold et al. found that speakers did not use reordering to avoid syntactic ambiguity: they were not more likely to include the goal early in sentences that would otherwise be ambiguous compared to in control sentences that included no such syntactic ambiguity. However, Arnold et al. did find effects of the speaker-oriented factors of syntactic weight and lexical bias on constituent ordering.

### 3.2.2. Prosodic disambiguation

Speakers can also use prosody to disambiguate syntax, but this does not appear to be a particularized adaptation. Allbritton, McKoon, and Ratcliff (1996) had speakers read aloud syntactically ambiguous sentences that were disambiguated for them by context. They found that speakers – even professionals such as actors and broadcasters – did not reliably use prosody to disambiguate the sentence. Speakers did so only with heavy prompting, such as when they viewed the two different interpretations side by side (see also Fox Tree & Meijer, 2000). However, it is possible that reading out loud is fundamentally different from speaking. To test this, Schafer, Speer, Warren, and White (2000) had participants play a collaborative game in which they used prescribed sentences to communicate about the locations of game pieces. These sentences had a temporary syntactic ambiguity following the main verb, e.g., *When that moves the square will ...* versus *When that moves the square it ...*. In contrast to Allbritton et al., they found that speakers did prosodically disambiguate the utterances. Curiously, the speakers used disambiguating prosody even when the context already disambiguated the meaning. Therefore, this could not be a listener-oriented adaptation.

Only one study found that speakers might spontaneously use prosody to disambiguate syntax. Snedeker and Trueswell (2003) had speakers play a game with a listener involving actions on a set of objects. The experimenter demonstrated an action for the speaker, and the speaker's task was to read a sentence from a card to lead the listener, who did not witness the action, to perform the same action on the objects. The utterances were syntactically ambiguous (e.g., *Tap the frog with the flower*) in that the prepositional phrase could be taken as specifying the referent (e.g., which frog to be tapped), or the instrument to be used in the action. When the speaker's referential context supported both interpretations, speakers provided disambiguating prosody; however, they did not do so when the context allowed only one interpretation (e.g., only one frog). These findings support the idea that speakers only produce prosodic disambiguation when they are aware of an ambiguity, and other disambiguating cues are absent (see also Straub, 1997).

However, Kraljic and Brennan (2005; see also Schafer, Speer, & Warren, 2005) could not replicate this result. They found that the likelihood of prosodic disambiguation depended only on the syntactic structure, and did not depend on the speaker's perception of situational ambiguity. Kraljic and Brennan did not prescript speakers' utterances, but used a task that would naturally induce utterances containing prepositional phrase-attachment ambiguities, such as *put the dog in the basket on the star*. Speakers disambiguated sentences using prosodic cues, but they did so even when the referential context already disambiguated them. Furthermore, interacting speakers were no more likely to produce prosodic disambiguation than speakers who spoke alone. These findings show that speakers use prosody for syntactic disambiguation, but not with a listener's needs in mind.

Another prosodic dimension of speech that might appear to be related to audience design is the clarity with which speakers pronounce individual words in ongoing speech. According to the design hypothesis, speakers should modulate their articulatory effort depending upon the predictability of the word in context (Lindblom, 1990). When a word is not predictable, speakers should expend extra effort to pronounce it clearly, because the listener would have to rely mainly on bottom-up information to identify the word. In support of this view, listeners tend to articulate the word *nine* more clearly in the sentence *The next number you will hear is nine* than in *A stitch in time saves nine* (Lieberman, 1963). In addition, words used to refer to "given" entities in discourse tend to be shorter in duration than words used to refer to "new" entities, with vowels often taking on a phonologically reduced form, e.g., [i] to [ɨ], (Bard et al., 2000; Fowler & Housum, 1987). At first blush, such findings seem to be evidence for audience design, because they seem well-adapted to listeners' perceptual needs. However, Bard et al. (2000) found that speakers pronounced words less clearly whenever they were repeated and coreferential, regardless of whether they were given or new for the listener. This result suggests that articulatory reduction is the result of speaker-centered priming processes, and is not a true listener adaptation.

### 3.3. Marking Referent Accessibility

Interestingly, one area in which researchers have found evidence for audience design is in the speaker's marking of the accessibility of referents (Bard & Aylett, 2005; Hupet &

Chantraine, 1992; Lockridge & Brennan, 2002). Linguistic theory suggests that speakers mark the accessibility of referents using variations in referential form (Ariel, 1988; Chafe, 1976; Gundel, Hedberg, & Zacharski, 1993). New referents – referents not yet introduced into the discourse – are generally of low accessibility. Speakers tend to mark new referents using the indefinite article *a* accompanied by an elaborate noun phrase (*a man I met yesterday on the plane*). Referents that have already been introduced (“given” referents), in contrast, tend to be more accessible; accordingly, they are often marked using the definite article *the* accompanied by a short descriptive phrase (*the man*). The choice between indefinite versus definite article per se need not indicate a particularized adaptation; speakers might simply mark the accessibility of a referent based on its availability in their own model instead of their beliefs about its accessibility in the listener’s model. However, studies that have independently manipulated the speaker’s and the listener’s knowledge suggest that the choice of definite versus indefinite marking is indeed a particularized adaptation (Bard et al., 2000; Bard & Aylett, 2005; Hupet & Chantraine, 1992).

### 3.4. Monitoring and Awareness of Ambiguity

Referential ambiguity is non-linguistic in nature: it derives from the mapping between words and the world. In contrast, linguistic ambiguity arises out of linguistic structure and linguistic performance. There are many different sources of linguistic ambiguity that could cause miscomprehension, such as homophony (words that sound alike but that have different denotations, e.g., *bat* as in flying mammal versus *bat* as in baseball or cricket bat), ambiguity in phonological segmentation (e.g., *alone* versus *a lone*), or syntactic ambiguity (e.g., attachment ambiguities, such as *the girl saw the boy on the bike*). Furthermore, there are many production variables that can affect the comprehensibility of utterances, such as speech rate, loudness, and clarity of articulation. For these reasons, speakers need to monitor the comprehensibility of what they say.

Ferreira, Slevc, and Rogers (2005) demonstrated that linguistic and non-linguistic ambiguity are monitored by different mechanisms. Speakers described targets such as a vampire bat in the context of a foil object. The foil object was either from the same category (e.g., a larger vampire bat), thereby creating a non-linguistic ambiguity, or from a different category covered by a homophonous term (e.g., a baseball bat), resulting in a linguistic ambiguity (i.e., since the word *bat* could be used for both types of object). In either case, speakers should avoid the bare homophone and include disambiguating information (e.g., *small bat* or *vampire bat*). Although speakers nearly always avoided non-linguistic ambiguity, they rarely avoided linguistic ambiguity except when they had already used the homophone to describe the foil. For example, they were only likely to call the flying mammal a *vampire bat* when they had just referred to the baseball bat as *the bat*. This suggests that when speakers monitor their speech before they begin articulation, they will reliably detect non-linguistic ambiguities but not linguistic ambiguities. The difference in the likelihood of detection supports the idea of distinct monitoring mechanisms.

Another mechanism that must be involved in audience design is one that allows the speaker to evaluate how well an utterance is designed for an addressee. Keysar and Henly

(2002) investigated how accurate speakers are in gauging the effectiveness of their utterances. In their study, speakers read aloud syntactically or lexically ambiguous sentences, such as “the man is chasing a woman on a bicycle,” trying to convey one of the meanings to a naïve listener that was seated opposite of them. The pair was then presented with the two paraphrases. Speakers selected which of the two meanings they thought the listener understood, while listeners selected the meaning they thought the speaker had intended. The majority of speakers overestimated the likelihood that their listeners had understood them: they assumed that their meanings were successfully conveyed in 72% of the cases, while listeners actually understood the intended meanings in only 61% of the cases. Although speakers overestimated their effectiveness, overhearers who knew the speakers’ intentions did not. This indicates that the overestimation arises out of the very act of speaking. Such overestimation puts limits on speakers’ ability to engage in audience design.

### **3.5. Audience Design in the Production of Written Text**

Writing differs from speaking in important ways. Unlike authors, speakers produce language under circumstances that leave little time for reflection, leading to greater difficulties in formulating what to say. Speakers also are likely to have immediate feedback from the recipient, which authors lack. Thus, successful communication in writing might demand deeper engagement in perspective taking, but may also offer authors more time to do so.

Traxler and Gernsbacher (1992) demonstrated the difficulty of the lack of feedback for writers by showing that even minimal feedback will improve writers’ effectiveness. They had writers describe a set of abstract figures. One group of writers were later told how many readers were able to identify the correct figure based on their description. This procedure was repeated three times. Writers who received feedback improved the effectiveness of their descriptions, while writers who did not receive feedback showed no such improvement. In a follow-up study, Traxler and Gernsbacher (1993) found that experience in perspective taking – by performing the selection task as a reader – helped writers improve the effectiveness of their messages. Envisioning the addressee’s perspective, then, is a barrier to effective communication, and even minimal feedback can help people communicate more effectively.

## **4. CURRENT STATUS OF THE DESIGN HYPOTHESIS IN LANGUAGE COMPREHENSION**

The audience design hypothesis that was posed for language production can also be posed for comprehension: do people interpret utterances as if they were optimally designed with respect to their common ground with the speaker or writer? Comprehension, like production, also proceeds incrementally through multiple stages, although in the opposite direction: from sound or print into a meaning represented in the mind. The comprehension system must identify words within the perceptual input and then access

syntactic and semantic information from the lexicon in order to build a sentence frame and construct an initial interpretation. Referents are identified within a discourse model or established as new. World-knowledge (scripts and schemata) may need to be consulted to draw inferences and fill in missing information. The ultimate goal of this process is to recover the speaker's or writer's communicative intention. Finally – though often overlooked – listeners, like speakers, monitor the coherence and adequacy of the emerging representation (Keysar et al., 1998; Markman, 1977).

At any or all of these stages, the comprehension system might consult a model of the speaker's knowledge. As was the case for production, the interesting question is not whether common ground constrains comprehension, but at what stage. To date, research has focused on the use of common ground in finding referents for spoken expressions as well as in comprehending written text. In this section, we review these findings and their implications.

#### 4.1. Comprehending Spoken References

Clark et al. (1983) sought to demonstrate that listeners use common ground in interpreting references. They pointed out that demonstrative references as in *that man* are used felicitously even when the term could apply to more than one referent in the context, suggesting that this is so because listeners can use common ground to find a unique referent. For instance, a speaker could point toward two men, a fat and a thin one, and say *George will look like that man very soon*, so long as one of the men is salient with respect to the common ground of the speaker and the listener. If the interlocutors had just been discussing George's obsession with weight loss, then the listener might take the thin man as the referent. If they had just been discussing his eating binges, then the listener might instead consider the fat man as the referent. In a series of experiments, they found that this is precisely what listeners do. However, listeners in these studies may have responded based on what was salient to them at the moment, rather than on their common ground with the speaker (Keysar, 1997).

A growing number of studies have sought to address listeners' use of common ground to identify referents using eyetracking. Eyetracking is a powerful technique that provides a moment-by-moment record of the comprehension process (e.g., Cooper, 1974; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). This makes it possible to observe how listeners progressively narrow the domain of reference as speech unfolds (see chapters in Trueswell & Tanenhaus, 2005, for discussion). Typically, in such experiments listeners take part in a referential communication task in which they follow a speaker's instructions, which involve looking at or manipulating objects.

In general, eyetracking studies of spoken language comprehension find strong, immediate effects of various kinds of contextual constraint on language processing. For instance, Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) found that listeners interpreted complex descriptions containing multiple pronominal modifiers (e.g., *touch the large starred rectangle*) by incrementally constraining the domain of reference as the



speech unfolded. For instance, upon hearing the adjective *large*, they tended to look at objects in the display that contrasted in size (e.g., a large and a small rectangle). Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) reported that when the visual context supported a nonpreferred syntactic structure, listeners did not appear to experience a garden path. Similarly, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) found that when listeners interpreted instructions such as *put the cube inside the can*, they restricted attention to cans that would be big enough to accommodate the cube. Finally, Hanna and Tanenhaus (2004) found that when listeners interpreted a speaker's references, they were attentive to whether the speaker was momentarily able to physically reach various objects.

The above studies suggest that various kinds of contextual constraint can immediately influence how listeners process language. The fact that listeners are able to so flexibly adapt comprehension to context might be viewed as evidence that listeners expect speakers to design their utterances with respect to common ground. For example, a listener might only expect a speaker to refer to a square using a modifier such as *large* if there is more than one square in the listener's common ground with the speaker. However, none of these studies is designed to permit such an inference, since they do not separate the information available to the listener from the listener's beliefs about the speaker's knowledge. Processing language with respect to the available context – whether the context is instantiated in a visual display or in a prevailing discourse model – could be a generic adaptation that would tend to promote successful comprehension whenever information that is available to a listener is also available to the speaker (Barr & Keysar, 2005; Pickering & Garrod, 2004).

The critical question is to what degree is a listener's search for referents restricted to copresent information. The full constraint model predicts that listeners would only consider referents that are in their common ground with the speaker (Clark and Carlson, 1981). In contrast to the full constraint model, perspective adjustment predicts that listeners would consider referents whether or not they are common. Evidence against full constraint and in favor of perspective adjustment was provided in a series of studies (Keysar et al., 2000, 1998; Keysar, Lin, & Barr, 2003). For example, Keysar et al. (2000) tracked participants' eyes as they played a communication game with a speaker. In the game, the speaker was a confederate who instructed the participant to move objects around a grid that was placed between them. To keep common ground constant while manipulating the listener's private knowledge, some objects were mutually visible (i.e., copresent), while others were visible only to the listener (i.e., private). For instance, in one item, the speaker told the listener to *move the small candle* in a display containing two copresent candles that varied in size. While the target was obviously the smaller of the two copresent candles, the listener could also see an even smaller, private candle (the "competitor"). In contrast to the full constraint hypothesis, listeners considered the competitor object as the referent: they were more likely to look at it than at a control object, and the presence of the competitor also delayed the identification of the target object. The most surprising result was predicted only by perspective adjustment: about a quarter of

the time, listeners attempted to move the private competitor, with the majority of listeners attempting this at least once during the experiment.

Such egocentric errors have been found to persist despite extensive attempts to eliminate them. Keysar et al. (2003) had listeners hide the private competitor object from their own view in an opaque brown bag, so that they knew of its existence in the bag but could not see it. Despite this, listeners still considered the private competitor in the bag as the referent. To make it even more obvious that the speaker did not know the identity of the competitor, another experiment led listeners to believe that the speaker had a false belief about the identity of the private competitor in the bag. This manipulation made the experiment similar in logic to the false belief task commonly used with children to assess Theory of Mind (Astington & Gopnik, 1988; Wimmer & Perner, 1983). Surprisingly, listeners showed the same interference when they believed the speaker had a false belief as when they believed she was ignorant. This shows that the cognitive system for reasoning about another's beliefs is not fully incorporated into the comprehension system, since it appears unable to constrain the search for referents.

Epley, Morewedge, and Keysar (2004) found further evidence for perspective adjustment in a study of perspective taking that directly compared the performance of adults with that of children. The fact that adults are less egocentric than children could be explained in two ways: (1) that adults are better at using mutual knowledge to avoid generating egocentric interpretations; or (2) that egocentric interpretations are generated automatically by the comprehension system, and adults are better at using common ground to monitor and reject these interpretations. Epley et al. found that although adults and children were equally likely to consider the private competitor, adults were faster to recover and were less likely to erroneously select the private object. These results support the idea that language is processed against an automatic egocentric default, and that the main difference between children and adults lies in adults' ability to detect and correct this egocentric interpretation.

Nadig and Sedivy (2002) and Hanna et al. (2003) argued that the results presented in support of perspective adjustment were also consistent with the partial constraint hypothesis, which assumes that common ground is but one of many cues integrated in parallel by a fully interactive comprehension system. Under this view, the effects of common ground would be immediate but only partial, since other cues might activate information that is available to the listener but is not in common ground. To show that comprehenders are initially egocentric, these researchers suggested, it is necessary to show that they would be equally likely to fixate a private competitor as they would a shared competitor that was an equally good fit to the speaker's description. Partial constraint would predict that comprehenders would be less likely to fixate a private competitor than a shared competitor during the early moments of comprehension, thereby demonstrating an early, albeit partial, effect of common ground. Perspective adjustment would predict that listeners would initially be equally likely to fixate private and shared competitors.

To test these hypotheses, Nadig and Sedivy (2002) had five-year-old children play a communication game with an adult confederate speaker while their eye movements were monitored. As in Keysar et al. (2000), Nadig and Sedivy found interference from private competitors, but they also found greater facilitation for the target with a private competitor than with a shared competitor, suggesting that common ground imposes a partial constraint on comprehension from the earliest moments. Hanna et al. (2003) found similar results in an eyetracking study with adults. From the earliest moments of comprehension, listeners were more likely to look at a target shape that was in common ground than a matching private shape. This provides further evidence against complete egocentrism, and supports the partial constraint hypothesis.

At present, the literature on common ground use in the comprehension of referential expressions suggests the existence of partial effects of common ground from the earliest moments of comprehension. At the same time, the egocentric element in comprehension has proven surprisingly resilient to manipulations that highlight the difference in perspective between speaker and listener. An important question that has not yet been addressed is whether the early effects of common ground reflect use of common ground by the comprehension system, or a strategic, non-linguistic use of common ground. Listeners may have been less likely to look at private objects than shared objects even in the absence of any linguistic input, as Keysar et al. (2000) found. Thus, copresent and private objects may have different baseline probabilities of fixation at the initial moment of the referring expression. Experiments that involve direct comparisons between shared and private objects do not control for this baseline difference, and thus leave open the possibility that the “partial” effects of common ground are due to task-based strategic effects of common ground mixed with automatic “egocentric” effects.

Although the debate over the use of common ground in the interpretation of spoken references is ongoing, it has already produced substantial insight into the integration of linguistic and contextual information in language comprehension. The use of common ground in non-conversational contexts, such as in text processing, has also increased our understanding of how multiple sources of information are brought together during language comprehension.

## **4.2. Using Common Ground in Text Processing**

### *4.2.1. Difficulty in using common ground*

Effective story comprehension requires readers to keep track of what characters believe and to use this information in interpreting their statements and actions (Graesser, Bowers, Bayen, & Hu, 2001; Graesser, Bowers, Olde, & Pomeroy, 1999). One potential difficulty for readers might be in considering how a given character would interpret the statement of another character, because this requires simultaneously keeping in mind both characters’ points of view. Keysar (1994) investigated such perspective taking with a variety of scenarios. For example, people read that David asked June for a restaurant

recommendation, and that he either had an excellent dinner there (positive experience) or a miserable one (negative experience). He then left her the message “About that restaurant, it was marvelous, just marvelous.” Readers were to assess how June would interpret the message. Given that she was uninformed about the experience, the assessment should have been the same in the two conditions. However, readers were more likely to attribute the perception of sarcasm to June when they knew that David was being sarcastic (negative experience). This phenomenon was termed the “Illusory transparency of intention.”

Weingartner and Klin (2005) showed that the illusion of transparency is not restricted to explicit judgments but occurs during normal reading. They presented readers with the original scenarios that ended with a target sentence that demonstrated that June either understood David’s intention or misunderstood him. Readers were slower to understand the target sentence when it was inconsistent with David’s intention, suggesting that readers indeed inferred that June understood what he meant even though she did not have sufficient information. Keysar (1998) showed that it is the *communicative intention* of the speaker that appears transparent, and not the speaker’s true attitude: when David’s experience was miserable, but he wanted to conceal this negative information from June, readers no longer thought that June would perceive sarcasm (this result was also replicated in Gerrig, Ohaeri, & Brennan, 2000, and in Weingartner & Klin, 2005).

The phenomenon of the illusory transparency of a speaker’s intention is striking, given that listeners can readily differentiate between what different characters know (Graesser et al., 2001). Thus, it suggests a dissociation between what readers know about what characters know, and how they use that knowledge to assess the perception of intentions. Researchers have offered two explanations for the phenomenon. The first of these explanations assumes that readers, like interlocutors, observe the principle of audience design (Gerrig et al., 2000; Keysar, 1994; see also Keysar, 2000; Gerrig et al., 2000, for discussion). Gerrig et al., (2000) argued that readers will go to great lengths to preserve the notion that speakers are being cooperative. Readers, then, might have assumed that speakers would not have made sarcastic utterances unless they had sufficient reason to believe that the addressee would have some way of perceiving their underlying intention. Thus, readers believe that the “message gets through” because they assume that characters strive to be cooperative, and proceed on their best estimate of the common ground between speaker and listener. Although this view is in line with standard pragmatic theory, there is no direct evidence to support it yet.

The second explanation for the phenomenon assumes perspective adjustment: readers compute the actual intention of the speaker and then try to make allowances for the perspective of the addressee. But as with other anchoring and adjustment phenomena, they adjust insufficiently (Epley et al., 2004; Keysar, 1994). Epley et al. (2004) demonstrated that illusory transparency is exacerbated when people are under conditions that hinder the adjustment process, such as time pressure. The difficulty readers have in taking the perspective of characters, then, might be a special consequence of the difficulty of taking perspective in general.

#### 4.2.2. *When is common ground used?*

Some studies offer a more optimistic assessment of readers' effectiveness in taking characters' common ground into account. Gibbs, Mueller, and Cox (1988) investigated the role of common ground in asking and comprehending questions. Participants read scenarios in which the common ground between characters was manipulated. After each scenario, participants were asked to select the most appropriate question for a speaking character to ask a listening character. There were three alternatives, reflecting different presuppositions about the common ground between the protagonists. The most common selection was in accordance with the common ground that had been established between the speaker and listener. In addition, in a reading comprehension study listeners took less time to read questions that appropriately specified the common ground between the interlocutors. These results indicate that readers are sensitive to common ground in developing expectations about how characters will interact.

The correct interpretation of many kinds of ambiguous utterances depends upon what communities the interlocutors jointly belong to (Clark & Marshall, 1981). For instance, consider how you would interpret a comment regarding a university sports event, such as *that game was a disaster*. If the speaker is a member of your university, then you would have evidence that your university's team lost. In contrast, if the speaker is a stranger, then the interpretation would be less certain. Gerrig and Littman (1990) examined how readers exploit protagonists' community membership in order to interpret their statements. Participants read a set of scenarios, each of which ended with an ambiguous statement. They were asked to interpret such statements as if they were addressed to them. Gerrig and Littman found that readers were more likely to interpret such ambiguous statements in line with community expectations when the speaker was a friend, compared to when the speaker was a stranger.

Greene, Gerrig, McKoon, and Ratcliff (1994) suggested that common ground affects reading by modulating the accessibility of information in memory. They argued that as characters come together, their common ground becomes more accessible in the reader's mind. In their experiment, participants read scenarios about two characters discussing a third character; one of the two characters then left, and then, after a short interlude, was reunited with the first character. Using a recognition probe task, Greene et al. found that the reunion of the characters could restore the accessibility of information related to the third character, who was in common ground. They suggested that, "upon the reunion of the two characters, the reader makes accessible the common ground that the characters share, in preparation for understanding their future interaction." (p. 524)

However, Lea, Mason, Albrecht, Birch, and Myers (1998; see also Keysar, 1997) demonstrated that this *reunion effect* reflects general memory associations rather than the use of common ground *per se*. Memory-based models (e.g., Gerrig & McKoon, 1998) stipulate that information can be reactivated through a memory "resonance" process that is based on the overlap between concepts. Thus, in the above example, the return of one of the characters may have reactivated concepts associated with her, regardless of

whether they were part of the common ground between the characters. Indeed, Lea et al. found that the reunion effect extended equally to concepts not in common ground between the protagonists. This supported the idea that the reunion effect was due to memory-based resonance processes, and not to the use of common ground.

In contrast to Greene et al., who argued that common ground modulates memory processes, Gerrig and McKoon (1998) rejected altogether the psychological reality of common ground in language use: “our account denies that common ground is something that speakers (or addressees) can or cannot take into account – merely memory processes acting on representations.” (p. 82) and “to the extent that ‘common ground’ seems to work pretty well, we suggest that it is because people serve as highly valid memory cues” (p. 81). However, this view is inconsistent with several observations in the literature. Lea et al. found that though concepts were reactivated independently of common ground, when a protagonist said something that violated common ground with the other, readers were surprised. Clearly, they should not have noticed a violation if they did not take common ground into account. It has also been observed that readers will *not* use information associated with a speaker when interpreting that speaker’s utterance, if they believe that the speaker wishes to keep that information concealed from the addressee (Gerrig et al., 2000; Keysar, 1998). Finally, it is inconsistent with results from eyetracking studies, which clearly show that making a decision about whether or not to use available information during comprehension involves the use of common ground, whether through monitoring (Keysar et al., 2000) or through some other mechanism that is more integral to the comprehension system (Hanna et al., 2003). Against the background of such a broad range of evidence for the use of common ground in interpreting language, attempts to deny the psychological reality of common ground are not compelling.

## 5. REPEATED REFERENCE IN DISCOURSE

Patterns of language use tend to evolve over the course of a conversation as interlocutors adapt the conventions of their language to current conversational needs. Most research on such adaptation has focused on how speakers adjust the speech they produce, although there is evidence that listeners adjust their expectations about the speech they are likely to hear (Barr & Keysar, 2002). The majority of these studies focus on repeated reference, i.e., on the changes that occur when speakers repeatedly make reference to an object during conversation. The main phenomena associated with repeated reference are reductions in the number of words used to describe a referent, stability in lexical choice, and changes in articulatory quality. In this section, we review these phenomena and discuss whether they reflect generalized or particularized adaptations.

### 5.1. Abbreviation of Description

Zipf (1935) noted an inverse relationship between the length of a word and its frequency of use, a relationship known as “Zipf’s Law.” A similar inverse relationship

exists between the length of a description of an object and the number of times the speaker has referred to it within the conversation (Krauss & Weinheimer, 1964, 1966). Typically, when speakers refer to the same referent multiple times, the overall length of their descriptions, as measured in number of words or in speaking turns, declines. Over time, the description takes on “name-like” properties, functioning more as a cue to memory rather than an actual description of the object (Carroll, 1980). In most studies observing this phenomenon, speakers describe abstract figures. The number of words they use is plotted against frequency of use, yielding an exponentially decaying function.

One explanation for this phenomenon is in terms of the accumulation of common ground between participants. This “collaborative model” assumes that interlocutors strive to minimize their collaborative effort in establishing reference through the establishment of mutually accepted descriptions (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). This view predicts that as common ground accumulates, speakers can shorten their descriptions because addressees will need less information. Clark and Wilkes-Gibbs (1986) observed this process of shortening in a referential communication task where speakers described a set of abstract figures. However, their experiment lacked a control condition to show that speakers who described the figures to different addressees each time would not shorten their descriptions. This left open the possibility that the length reduction is due to speaker-related factors.

Hupet and Chantraine (1992) tested this possibility with two groups of participants performed a referential communication task. Both groups of participants spoke into a tape recorder with no physically present addressee. Participants in one group were told that they would keep talking to the same addressee as they did on the previous trial, while those in the other group were told that they would be talking to a different addressee each time. Hupet and Chantraine found that speakers who spoke to different addressees each time actually increased the number of words they used, while those speaking to the same addressee did not decrease the number of words they used. These results suggest that the presence of addressee feedback may be critical for the reduction in length of referential descriptions (see also Krauss & Weinheimer, 1966). However, in a later study, Hupet and Chantraine (1994) compared groups of participants who described abstract figures to the same addressee in either a monologue or a dialogue, and found that a subset of the participants in the monologue condition behaved the same as the participants in the dialogue condition. This suggests that the reduction effect is only partly due to collaboration.

Isaacs and Clark (1987) looked for evidence for the collaborative model using a referential communication task in which participants sought to communicate the ordering of pictures of New York landmarks. This made it possible to observe when speakers opt for a description of a referent (e.g., the peaked building) versus a proper name (e.g., the Citicorp Building). According to the collaborative model, speakers should use proper names alone only when they have evidence that their listeners are familiar with New York landmarks. A main finding was that experts who spoke to novices often started out using proper names and descriptions, and over time, increased their use of descriptions alone. Novice directors who spoke to expert matchers started out using descriptions and

increased their use of proper names, which they could only have learned from the matcher. These results indicate that, during conversation, speakers and listeners are able to discover their common ground and adjust their language use accordingly. Similar results were reported by Nohara-LeClair (2001) in a series of experiments in which participants made reference to international flags. Participants' shared knowledge of flag names increased during the referential communication task, as well as their accuracy in their assessments of their partner's knowledge. In addition, speakers tended to use the name of the country in referring to the flag only when they believed that their partner shared that knowledge; otherwise they tended to use descriptions.

## 5.2. Conversational Precedents

Another phenomenon of repeated reference is that speakers' lexical choice tends to stabilize over time. Such "lexical entrainment" (Brennan & Clark, 1996; Garrod & Anderson, 1987) is interesting because it suggests that a speaker's decision regarding how a referent is to be described depends not only on the current context, but also on the history of usage within a given conversation. When speakers initially select a label for a referent, they have multiple options that vary in their specificity: a certain car can be called *car*, *sportscar*, *fancy car*, *vehicle*, etc. One factor that will influence a speaker's initial choice is the set of objects from which the referent is to be distinguished (Olson, 1970). Thus, in the context of a flower, speakers might refer to a car by its basic-level term *car*. However, in the context of a second car, a speaker would need to choose a more specific term, such as *sportscar*. Yet, once speakers use a particular term for a referent, they tend to continue to use it on subsequent turns. One important consequence of this is that it can lead speakers to "overspecify" the referent; that is, to refer to a car as the sportscar when there is no other car in the context (Brennan & Clark, 1996). On the surface, such usage would seem to violate Grice's Maxim of Quantity (Grice, 1975), according to which speakers should provide no more information than necessary. However, such usage could be considered cooperative, in that it exploits established agreements or "precedents" on how a referent is to be conceptualized, or what Brennan and Clark (1996) called "conceptual pacts."

Brennan and Clark (1996) argued that speakers use conversational precedents in a partner-specific manner; that is, in a manner that is sensitive to common ground. They used a task that induced speakers to entrain on subordinate-level descriptions for pictures of everyday objects (e.g., *sportscar* instead of *car*, *loafer* instead of *shoe*), and in a subsequent test phase manipulated whether the speakers continued on with the same partner or with a new partner. In the test phase, the pictures appeared in displays lacking same-category objects, so speakers could simply use basic-level descriptions (e.g., *car* and *shoe*). The prediction was that speakers would be more likely to do so with a new partner than with an old partner, because the precedents of using subordinate terms are specific to the old partner. Although this prediction was confirmed, in the first trial of the test phase, speakers used subordinate terms with new partners just as much as with old ones. This suggests that the speaker's abandonment of the precedents with the new addressee may have been other-prompted rather than self-prompted; that is, speakers might have



used the precedents regardless of whether the partner was new or old, but this might have confused the new addressee, prompting speakers to adapt to them.

Horton and Gerrig (2002) provided further evidence for the role of other-prompted adjustments in audience design. They suggested that audience design is only sometimes necessary for references to be successful (which means that listeners will often find speakers' egocentric descriptions adequate). They hypothesized that the need for audience design will depend on the current conversational task, and that speakers can discover the appropriate circumstances through experience. Horton and Gerrig had participants play a referential communication task with two independent matchers who had different subsets of knowledge. In initial rounds of the game, directors independently established precedents for referring to specific objects with each of the two matchers. There were two later rounds that immediately followed a partner switch, confronting directors with the task of describing referents for which precedents existed with the previous partner, but not the current one. Horton and Gerrig found more evidence of audience design after the second partner switch than after the first one, suggesting that the feedback directors' received after the first switch prompted awareness of the need to take listeners' knowledge into account. Thus, speakers did not deploy audience design in an absolute manner, but did learn to detect cases in which it would be necessary.

Bromme and colleagues (Bromme, Jucks, & Runde, 2005; Bromme, Jucks, & Wagner, 2005; Jucks, Bromme, & Becker, 2005; Jucks, Bromme, & Wagner, 2005) investigated audience design where sensitivity to the audience could have serious consequences – in doctor–patient communication. Jucks et al. (2005) investigated entrainment in how medical experts use language when answering e-mail queries about medical problems with the help of a “concept map.” They found that when a query used technical medical terms such as “arteriosclerosis,” the expert was more likely to answer with technical language compared to when the query used everyday language such as “vascular hardening.” Such entrainment could reflect audience design, as the patient who uses technical language is probably more knowledgeable about medicine than the one who does not. But Jucks et al. showed that it actually reflects the availability of the term for the speaker: such entrainment occurred even when the technical term came from the concept map, and not from the patient. Interestingly, though experts' reflections showed a sensitivity to patient terminology, their answers to the queries did not take advantage of this knowledge. Therefore, even when the stakes are higher than in the harmless activity of identifying cars and candles, speakers use precedents because they are there, not because they are shared with the addressee.

If conversational precedents are used strictly in a partner-specific manner, then speakers should not carry over the precedents from one partner to the next. However, such carry-over could be a cultural-level adaptation that contributes to the diffusion of conventions within a language community. Malt and Sloman (2004) tested whether the precedents established by a speaker in one conversation would be used in a distal conversation. A confederate director used one of two conventional labels (e.g., *trash can* or *waste basket*) to describe common objects to a participant matcher in a referential

communication task. The matcher then became a director for a new participant matcher, and the task was repeated. Finally, the new matcher became a director with a new participant. Even though none of the participants in this third conversation interacted with the confederate who originally established the precedent terms, those terms were more likely to show up in this third conversation than the alternate terms. This lends some support to the idea that speakers do not necessarily use precedents in a partner-specific manner.

Barr and Keysar (2002) investigated whether precedents benefit language comprehension, and if so, whether this benefit is partner-specific or partner-independent. They monitored listeners' eye movements in a referential communication task. Unsurprisingly, listeners were faster to identify referents when referring expressions had precedents, compared to when they were new. This comprehension benefit could be the result of the accumulation of common ground, or could be due to the availability of the precedent. A common ground account would predict that if the second use of the term was made by a new, uninformed speaker, the benefit would be reduced, while the availability account would predict an equal benefit. Barr and Keysar found that listeners benefited equally from precedents, regardless of whether or not the speaker was the one who established the precedent or a new speaker. This result is surprising because the objects used in the experiments were highly unconventional, and therefore it would be highly unlikely for two independent speakers to refer to such objects in precisely the same way.

But perhaps listeners inferred common ground between the two speakers precisely because those speakers choose the same names for an unfamiliar object. For this reason, Barr and Keysar conducted an additional experiment in which a speaker broke a precedent instead of following it. If precedents are used in a partner-specific manner, then the breaking of a precedent should lead to greater interference when it was broken by the same person who established it, as compared to when it was broken by a new speaker. Barr and Keysar found equally strong interference when precedents were violated by a new or old speaker. Thus, listeners appear to use precedents when comprehending speech because they are cognitively available, not because they are part of their common ground with the speaker.

However, in this latter experiment, speakers referred to familiar, everyday objects, and broke precedents by reverting from subordinate-level descriptions (*sportscar*) to basic-level descriptions (*car*). Metzger and Brennan (2003) sought to replicate Barr and Keysar's findings of partner-independence, but used unconventional objects, for which the violation of a precedent might be more jarring. They replicated Barr and Keysar's finding that listeners were fast to identify referents when precedents were maintained, regardless of speaker. However, they also found that listeners were slower to identify the old-referent target when speakers broke their own precedents than when a new speaker broke a previous speaker's precedents.

Although the results of Metzger and Brennan do support the idea that listeners use common ground to deal with broken precedents, they leave open at least two important questions. First, since Metzger and Brennan did not provide detailed time-course

information, it is not clear whether their results reflect “early” effects of common ground, as would be predicted by the partial constraint hypothesis, or “late” effects of common ground, as would be predicted by perspective adjustment. Second, the results might reflect a more general expectation that speakers would use terms consistently, independently of their addressees. That would render such results speaker-specific, but without involving common ground.

## 6. SUMMARY AND PROSPECTUS

We have made the case that the existence of multiple explanations is intrinsic to the study of language use, since it is an activity that can be analyzed at the cultural, interactional, and cognitive levels. Researchers often overlook these multiple explanations due to the entrenchment of theoretical frameworks (whether interactional or psycholinguistic) that give priority to certain levels of explanation over others. We proposed that to properly interpret experimental findings, it is necessary to observe distinctions between generic and particularized adaptations as well as between self-prompted versus other-prompted adaptations. In light of these distinctions, we reviewed evidence for audience design in language production, language comprehension, and repeated reference. In this section, we attempt to assess the degree of support this body of evidence provides for the various models of the use of common ground in language processing.

### 6.1. Summary of Findings and Evaluation of the Models

When participating in a conversation, language users are beholden to the maxims of cooperative communication (Grice, 1975). Thus, when deciding what one wishes to say to an interlocutor, what kind of speech act to use, what topic to converse about, and what language to use, speakers might extensively consult their knowledge about common ground. By comparing the speech produced by adults to the speech produced by children, one can see that adults pay attention to the interlocutor’s perspective in making such high-level decisions. Researchers have noted that children’s speech tends to not only be strikingly more egocentric than that of adults (Flavell et al., 1968), but children’s dialogues also tend to take the form of “serial monologues,” in which children take turns at speaking, but often fail to produce speech that follows the thread of the previous speaker (Piaget, 1926/1955).

Although many high-level decisions speakers make may be governed by common ground, our review has suggested that many of their lower-level decisions are surprisingly insensitive to common ground. Speakers do include information in their speech that a listener could not readily infer, such as mentioning an atypical instrument, but this appears to be a by-product of the way concepts are accessed during language production (Brown & Dell, 1987). Such adaptations may be made in regard to the needs of the particular listener (Lockridge & Brennan, 2002), but these particularized adaptations appear to be other-prompted and not self-prompted. Perspective adjustment could accommodate this latter finding by assuming that speakers who failed to mention atypical instruments

received negative feedback from addressees, which made them monitor their speech more closely. Furthermore, a multiple-systems approach seems necessary to explain the findings that, if individuals are placed under cognitive load, they fall back on a default form of speaking that is not tailored to the addressee (Horton & Keysar, 1996; Rossnagel, 2000).

Although speakers do appear to monitor aspects of their speech for audience design, they are not very proficient at detecting ambiguity in their own speech, even when they are not under load (Ferreira et al., 2005; Keysar & Henly, 2002). This distinction between use of common ground in the conceptualization versus formulation of an utterance plan poses problems for both full and partial constraint models, and supports multiple-systems approaches. Speakers do not use optional words or constituent ordering to avoid garden paths. In spoken dialogue, speakers do appear to adopt a prosodic marking that correlates with the syntactic structure, and these cues do appear to help listeners (Kraljic & Brennan, 2005; Schafer et al., 2000; Snedeker & Trueswell, 2003). But given that speakers use the cues independently of the listener's needs (Kraljic & Brennan, 2005), the cues are not part of audience design. In addition, the articulatory reduction of repeated words in dialogue does not depend upon the listener's perceptual needs, but appears to be the result of repetition priming (Bard et al., 2000; Bard & Aylett, 2005).

The finding that speakers use the definite and indefinite article to mark accessibility for the listener (Bard & Aylett, 2005; Hupet & Chantraine, 1992; Lockridge & Brennan, 2002) reflects a particularized adaptation. To explain this, perspective adjustment would have to assume that the selection of a definite versus indefinite form is first done egocentrically (i.e., on the basis of the given/new status of the referent for the speaker) and then adjusted during a later-monitoring stage. The dual process model provides a more parsimonious explanation by assuming that the marking of definiteness is a controlled process that figures into the initial design of an utterance. It is not clear how these two possibilities could be distinguished, since both theories would predict an "egocentric" result under cognitive load.

For language comprehension, the results clearly reject the full constraint model, since effects of private knowledge on comprehension have now been widely replicated in many studies by various laboratories, using a variety of tasks. The ordinary memory view also cannot explain the majority of the findings, because comprehenders appear to honor the distinction between private and common information. Instead, the bulk of evidence that we reviewed on language comprehension could be construed as supporting either perspective adjustment or partial constraint. Notably absent from the literature to date have been discussions of a dual process model for comprehension, although it could also explain the body of findings. More fine-grained, process-level investigations are needed to distinguish between these three competing models.

The abbreviation of description is another case in which particularized adjustments appear to be made; but here, too, multiple explanations exist. Abbreviation depends upon the presence of feedback (Hupet & Chantraine, 1992, 1994; Krauss & Weinheimer,

1966), but little is known regarding just how this feedback makes a difference. Feedback might enable the orderly accumulation of common ground (Clark & Brennan, 1991), and the shortening may be due to the speaker's self-prompted use of this common ground. Alternatively, the speaker's reduction in description length might be other-prompted, falling out naturally from the speaker's sensitivity to evidence of listener understanding: speakers may terminate their descriptions as soon as they have evidence that the listener understands, which would come earlier and earlier at each reference turn.

Evidence regarding the use of conversational precedents in production and comprehension does not unequivocally support partner-independence or partner-specificity. Some of the partner-specific effects might be interpreted as other-prompted adaptations, such as speakers' reversion to basic-level terms in Brennan and Clark (1996). Others, such as listeners' spontaneous use of common ground when precedents are broken (Metzing & Brennan, 2003), would seem to be self-generated, posing problems for perspective adjustment. In contrast, the robust finding of no effect of common ground when precedents are maintained poses problems for both full and partial constraint models. In sum, no current model is able to explain the totality of the findings on conversational precedents.

Our verdict is that the multiple-systems approaches fare best in explaining the full set of findings reviewed in this chapter. The balance of findings would appear to tip slightly in favor of the dual process model over perspective adjustment, only because of a single finding: that speakers use the definite versus indefinite article according to the accessibility of referents for the listener, a finding that is explained more parsimoniously by the dual process model. Otherwise, the two models can account for the data equally well. Partial constraint can explain some of the findings – especially those pertaining to language comprehension – but encounters difficulties explaining how manipulations of cognitive load might affect audience design processes in language production, while leaving other aspects of production relatively intact. Finally, the full constraint and ordinary memory models can be rejected; the former based on the observation of effects of private knowledge on all aspects of production and comprehension, and the latter based on the observation that decision processes clearly honor the distinction between private and common information.

## 6.2. Interactive Alignment

The majority of the work reviewed here appears to undermine the proposal that speakers and listeners solve coordination problems by processing language against their common ground. There may still be important roles for common ground in very high levels of planning during language production, in the detection and correction of errors in both production and comprehension, and in knowing how to interpret feedback from one's interlocutor. However, audience design does not appear to play a role in many of the routine processes involved in speaking and understanding. This might seem theoretically problematic, since egocentric processing would seem to provide a poor foundation for

cooperative communication. However, audience design is not the only possible explanation for why conversation succeeds. We close this chapter by considering an alternative proposal from Pickering and Garrod (2004).

According to the *interactive alignment* approach (Pickering & Garrod, 2004), listeners tend to understand speakers because mechanisms exist that promote the generation of cognitive representations that are shared, or in “alignment.” This alignment is not serendipitous, nor is it explicitly negotiated by interlocutors, but emerges as a by-product of the many individual acts of coordination that take place during conversation. Importantly, these shared representations are different from common ground in that they do not involve an explicit model of the other, nor are they necessarily “known to be shared.” The alignment process is facilitated by automatic priming processes, both within the individual language user as well as across individual language users. For instance, a speaker who produces a particular syntactic structure not only becomes more likely to repeat it in the future, but the comprehender who recovers that structure during the course of interpretation will also be more likely to produce it in the future. In this way, interlocutors’ representations converge over the course of a dialogue. Alignment is made possible by the assumption of *representational parity*: that the production and comprehension of language draw upon the same cognitive representations, a notion similar to the “parallelism” proposed by Dell and Brown (1991). As in perspective adjustment, common ground has a limited role in the coordination of shared understanding by enabling a process of “interactive repair” that can correct misunderstandings caused by insufficient alignment.

The findings reported in this review are largely consistent with the claims of the interactive alignment model. The model also enjoys a broad range of empirical support from studies showing spontaneous alignment between interlocutors in the absence of explicit negotiation, at many levels of processing, including speech rate (Webb, 1969), syntax (Bock, 1986; Branigan, Pickering, & Cleland, 2000; Levelt & Kelter, 1982), lexical choice (Garrod & Anderson, 1987), and conceptualization (Barr, 2003; Garrod & Anderson, 1987; Garrod & Doherty, 1994; Markman & Makin, 1998). Thus, interactive alignment offers a viable alternative for explaining coordination in language use. However, it is currently lacking a description of how this convergence of representations takes place. Although many aspects of alignment may be supported by automatic mechanisms such as priming, it may be necessary to invoke other mechanisms to explain certain aspects of language use (Brown-Schmidt & Tanenhaus, 2004; Kaschak & Glenberg, 2004; Krauss & Pardo, 2004; Markman, Kim, Larkey, Narvaez, & Stilwell, 2004; Shintel & Nusbaum, 2004).

### 6.3. Concluding Statement

The main lesson of our review is simple but important: when communication succeeds in the face of ambiguity, it does not necessarily entail audience design. Even if a speaker says something that seems designed for a listener, it does not mean that the speaker

actually designed it for that listener. Likewise, just because a listener understood an utterance as intended does not mean that she/he consulted her/his common ground with the speaker. Interlocutors may understand one another very well when neither of them uses common ground. After all, the cognitive representations that speakers and listeners tend to be working with are representations that are likely to be shared (Barr & Keysar, 2005; Pickering & Garrod, 2004).

Developing a theory of how people adapt their language behavior to meet the informational needs of other individuals is one of the central goals of research on language use. To meet this goal, researchers must surmount the challenges posed by the complexity of language use, a complexity that derives from its simultaneous cultural, interactional, and cognitive nature. Although the research of the past few decades has made great strides, understanding the joint activity of conversation is a task that will require the joint activities of researchers for many decades to come.

## REFERENCES

- Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 714–735.
- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*, 65–87.
- Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*, 55–70.
- Astington, J. W., & Gopnik, A. (1988). Knowing you've changed your mind: Children's understanding of representational change. In: J. W. Astington, & P. L. Harris (Eds), *Developing theories of mind* (pp. 193–206). New York: Cambridge University Press.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.
- Bard, E. G., & Aylett, M. P. (2005). Referential form, word duration, and modeling the listener in spoken dialogue. In: J. C. Trueswell, & M. K. Tanenhaus (Eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173–192). Cambridge, MA: MIT Press.
- Barr, D. J. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin and Review*, *10*, 462–467.
- Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, *28*, 937–962.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418.

- Barr, D. J., & Keysar, B. (2005). Making sense of how we make sense: The paradox of egocentrism in language use. In: H. L. Colston, & A. N. Katz (Eds), *Figurative language comprehension: Social and cultural influences* (pp. 21–41). Mahwah, NJ: Erlbaum.
- Bell, A. (1984). Language style as audience design. *Language and Society*, 13, 145–204.
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75, B13–B25.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–1493.
- Bromme, R., Jucks, R., & Runde, A. (2005). Barriers and biases in computer-mediated expert-layperson-communication: An overview and insights into the field of medical advice. In: R. Bromme, F. W. Hesse, & H. Spada (Eds), *Barriers and biases in computer-mediated knowledge communication – and how they may be overcome* (pp. 89–118). New York: Springer.
- Bromme, R., Jucks, R., & Wagner, T. (2005). How to refer to “diabetes”? Language in online health advice. *Applied Cognitive Psychology*, 19, 569–586.
- Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19(4), 441–472.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2004). Priming and alignment: Mechanism or consequence? *Behavioral and Brain Sciences*, 27, 25–26.
- Carroll, J. M. (1980). Naming and describing in social communication. *Language and Speech*, 23, 309–322.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: C. N. Li (Ed.), *Subject and topic* (pp. 26–55). New York: Academic Press.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, 47, 30–49.
- Clark, H. H. (1992). *Arenas of language use*. Chicago, IL, USA: University of Chicago Press.
- Clark, H. H. (1996). *Using language*. Cambridge, England, UK: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In: L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC, USA: American Psychological Association.
- Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In: J. Long, & A. Baddeley (Eds), *Attention and performance IX* (pp. 313–330). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.



- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In: A. K. Joshe, B. L. Webber, & I. A. Sag (Eds), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In: J. F. Le Ny, & W. Kintsch (Eds), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.
- Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, 26, 209–225.
- Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22(2), 245–258.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Danks, J. H. (1970). Encoding of novel figures for communication and memory. *Cognitive Psychology*, 1, 179–191.
- Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the “model of the listener”. In: D. J. Napoli, & J. A. Kegl (Eds), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila Gleitman*. Hillsdale, NJ: Erlbaum.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87, 327–339.
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40, 760–768.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4), 296–340.
- Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96, 263–284.
- Flavell, J. H., Botkin, P. T., & Fry, C. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.
- Fleming, J. H., & Darley, J. M. (1991). Mixed messages: The multiple audience problem and strategic communication. *Social Cognition*, 9, 25–46.

- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26, 489–504.
- Fox Tree, J. E., & Meijer, P. J. A. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research*, 63, 1–13.
- Fussell, S. R., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, 25(3), 203–219.
- Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6), 509–525.
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62(3), 378–391.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218.
- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53, 181–215.
- Gerrig, R. J., Brennan, S. E., & Ohaeri, J. O. (2000). What can we conclude from speakers behaving badly? *Discourse Processes*, 29(2), 173–178.
- Gerrig, R. J., & Littman, M. L. (1990). Disambiguation by community membership. *Memory and Cognition*, 18(4), 331–338.
- Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26(2–3), 67–86.
- Gerrig, R. J., Ohaeri, J. O., & Brennan, S. E. (2000). Illusory transparency revisited. *Discourse Processes*, 29, 137–159.
- Gibbs Jr., R. W., Mueller, R. A., & Cox, R. W. (1988). Common ground in asking and understanding questions. *Language and Speech*, 31, 321–335.
- Graesser, A. C., Bowers, C. A., Bayen, U. J., & Hu, X. (2001). Who said what? Who knows what? Tracking speakers' knowledge in narrative. In: W. van Peer, & S. Chatman (Eds), *Narrative perspective: Cognition and emotion* (pp. 255–272). Albany, NY: SUNY University Press.
- Graesser, A. C., Bowers, C. A., Olde, B., & Pomeroy, V. (1999). Who said what? Source memory for narrator and character agents in literary short stories. *Journal of Educational Psychology*, 91, 284–300.
- Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, 33, 511–526.

Grice, H. P. (1975). Logic and conversation. In: P. Cole, & J. Morgan (Eds), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28, 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49, 43–61.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing *when* and knowing *how* to adjust utterances to addressees. *Journal of Memory and Language*, 47, 598–606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96, 127–142.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59(1), 91–117.

Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of Psycholinguistic Research*, 21, 485–496.

Hupet, M., & Chantraine, Y. (1994). Efficiency of the addressee's contribution to the establishment of references: Comparing monologues with dialogues. *Cahiers de Psychologie Cognitive*, 13, 777–796.

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: The development of shared symbols in interaction. In: N. Gilbert, & R. Conte (Eds), *Artificial societies: The computer simulation of social life* (pp. 157–189). London: UCL Press.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26–37.

Jucks, R., Bromme, R., & Becker, B. -M. (2005). Lexical entrainment – is expert's word use adapted to the addressee? Under editorial review.

Jucks, R., Bromme, R., & Wagner, T. (2005). Choice of words in doctor-patient-communication: An analysis of health-related Internet sites. Under editorial review.

Kaschak, M., & Glenberg, A. (2004). Interactive alignment: Priming or memory retrieval? *Behavioral and Brain Sciences*, 27, 33–34.

Kennedy, J. (1890/1996). *Word stems: A dictionary*. New York: Soho Press.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, 26(2), 165–208.

- Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, 24(2–3), 253–270.
- Keysar, B. (1998). Language users as problem solvers: Just what ambiguity problem do they solve? In: S. R. Fussell, & R. J. Kreuz (Eds), *Social and cognitive psychological approaches to interpersonal communication* (pp. 175–200). Mahwah, NJ: Erlbaum.
- Keysar, B. (2000). The illusory transparency of intention: Does June understand what Mark means because he means it? *Discourse Processes*, 29, 161–172.
- Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13, 207–212.
- Keysar, B., & Horton, W. S. (1998). Speaking with common ground: From principles to processes in pragmatics: A reply to Polichak and Gerrig. *Cognition*, 66(2), 191–198.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38.
- Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, 39(1), 1–20.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25–41.
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50, 194–231.
- Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, 9(1), 2–24.
- Krauss, R. M., & Pardo, J. S. (2004). Is alignment always the result of automatic priming? *Behavioral and Brain Sciences*, 27, 35–36.
- Krauss, R. M., Vivekananthan, P. S., & Weinheimer, S. (1968). "Inner speech" and "external speech": Characteristics and communication effectiveness of socially and nonsocially encoded messages. *Journal of Personality and Social Psychology*, 9, 295–300.
- Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1(5), 113–114.
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3), 343–346.
- Lea, R. B., Mason, R. A., Albrecht, J. E., Birch, S. L., & Myers, J. L. (1998). Who knows what about whom: What role does common ground play in accessing distant information? *Journal of Memory and Language*, 39, 70–84.
- Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14, 78–106.

- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172–187.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In: W. J. Hardcastle, & A. Marchal (Eds), *Speech production and speech modeling*. Dordrecht: Kluwer Academic Publisher.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review*, 9, 550–557.
- Malt, B. C., & Sloman, S. A. (2004). Conversation and convention: Enduring influences on name choice for common objects. *Memory and Cognition*, 32, 1346–1354.
- Markman, A. B., Kim, K., Larkey, L. B., Narvaez, L., & Stilwell, C. H. (2004). One alignment mechanism or many? *Behavioral and Brain Sciences*, 27, 36–37.
- Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, 127, 331–354.
- Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development*, 48, 986–992.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49, 201–213.
- Minsky, M. (1974). A framework for representing knowledge. In: P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329–336.
- Nohara-LeClair, M. (2001). A direct assessment of the relation between shared knowledge and communication in a referential communication task. *Language and Speech*, 44, 217–236.
- Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 77, 257–273.
- Piaget, J. (1926/1955). *The language and thought of the child*. New York: Meridian.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 1–22.
- Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). *Cognition*, 66(2), 183–189.
- Rossnagel, C. (2000). Cognitive load and perspective taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, 30, 429–445.
- Schafer, A. J., Speer, S. R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In: J. C. Trueswell,

- & M. K. Tanenhaus (Eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 209–225). Cambridge, MA: MIT Press.
- Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169–182.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Schober, M. F. (1993). Spatial perspective taking in conversation. *Cognition*, 47, 1–24.
- Shintel, H., & Nusbaum, H. C. (2004). Dialogue processing: Automatic alignment or controlled understanding? *Behavioral and Brain Sciences*, 27, 42–43.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130.
- Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 43, 549–565.
- Steels, L. (1996). Self-organizing vocabularies. In: C. G. Langton, & T. Shimohara (Eds), *Artificial life V: Proceedings of the fifth international workshop*. Cambridge, MA: MIT Press.
- Straub, K. A. (1997). *The production of prosodic cues and their role in the comprehension of syntactically ambiguous sentences*. Unpublished doctoral dissertation University of Rochester, Rochester, NY.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 32, 425–443.
- Traxler, M. J., & Gernsbacher, M. A. (1992). Improving written communication through minimal feedback. *Language and Cognitive Processes*, 7, 1–22.
- Traxler, M. J., & Gernsbacher, M. A. (1993). Improving written communication through perspective taking. *Language and Cognitive Processes*, 8, 311–334.
- Trueswell, J. C., & Tanenhaus, M. K. (Eds). (2005). *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
- Watts, D. J. (1999). Networks, dynamics and the small-world phenomenon. *American Journal of Sociology*, 105, 493–527.
- Webb, J. T. (1969). Subject speech rates as a function of interviewer behavior. *Language and Speech*, 12, 54–67.

Weingartner, K. M., & Klin, C. M. (2005). Perspective taking during reading: An on-line investigation of the illusory transparency of intention. *Memory and Cognition*, 33, 48–58.

Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought*. New York: Wiley.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.

## Chapter 24

### **Comprehension Disorders in Aphasia: The Case of Sentences that Require Syntactic Analysis**

*David Caplan and Gloria Waters*

#### **1. INTRODUCTION**

Sentences convey information that goes beyond the meanings of the words they contain, and that indicates how the items referred to by these words interact. For example, sentence (1) does more than simply mention a boy, an apple, and eating; it conveys the information that the boy did the eating and that what was eaten was an apple.

1. The apple the boy is eating is red.

This information is known as the *thematic roles* played by items in the sentence: in (1), the boy is the agent of the verb “eat;” apple is the theme of this verb. Thematic roles such as agent, theme, and others are some of the semantic features conveyed by sentences. Sentences convey other information of this type, including the relation of modifiers to nouns and to each other (e.g., that the apple in (1) is red), the fact that the items referred to by nouns, verbs, pronouns, and other elements are the same, and other types of relations among words. Collectively, this type of information is often termed the *propositional content* of a sentence.

Propositional information can sometimes be inferred from the meanings of the words in a sentence combined with knowledge about the way the world operates. A person who knows what the words in (1) mean can assign the role of agent to boy and theme to apple because boys eat apples and not vice versa. We have called this a “lexical-pragmatic” basis for understanding sentences. Not all sentences can be understood lexico-pragmatically; however, only “semantically irreversible” sentences, in which only one way of arranging words is compatible with world events, can be. As far as the world is concerned, either boys or girls can chase each other and be tall. Thus sentence (2), a “semantically reversible” sentence, would be ambiguous, if all that we could use to understand it were a lexical-pragmatic mechanism.



2. The girl the boy is chasing is tall.

Sentence (2) is not ambiguous, however. It derives its meaning from its syntactic structure. Syntactic structures relate words to one another. They consist of complex hierarchically organized sets of syntactic categories (nouns, verbs, noun phrases, etc.) over which specific relationships (such as “subject of”) are defined, and these relationships determine aspects of propositional meaning. In (2), “the boy” is the subject of “is chasing” and therefore its agent; “the girl” is the subject of the matrix sentence and therefore the predicate “is tall” is assigned to *the girl*. All sentences have syntactic structure, and this structure determines their propositional content. Assigning syntactic structure and using it to determine propositional content is a second mechanism that underlies sentence comprehension, which we shall call a *syntactic* mechanism.

Syntactic representations are complex, and there is evidence that listeners and readers take short cuts in the process of assigning these structures, often basing their interpretations of sentences on common and simple structures. We will call this a “heuristic” mechanism. A famous example of such a mechanism is illustrated in (3).

3. The horse raced past the barn fell.

Readers not familiar with this sentence often think it is ungrammatical. They understand it to mean that the horse raced past the barn, and then are stuck with the final verb “fell” and cannot form a grammatical sentence. In fact, (3) is perfectly grammatical: it has the same structure as the sentence *The horse racing past the barn fell*. (3) is short for *The horse that had been raced past the barn fell* (see McKoon & Ratcliff, 2005, for a different view of these structures). The problem for readers confronted with (3) is that “raced” is ambiguous in form: it can be a past tense verb or a past participle. Readers prefer to take it as a past tense verb, and assign “the horse” as its agent. Bever (1970) referred to this as a “noun–verb–noun (NVN) mapping onto agent–verb–theme” heuristic. The NVN sequence is one of several simplified structures that listeners and readers use to assign meaning. Sometimes the meanings derived from these simplified structures are the same as those derived from a complete syntactic analysis; sometimes they are not.

To summarize, we have identified three mechanisms that can be used to determine sentence meaning: a lexico-pragmatic mechanism, a syntactic mechanism, and a heuristic mechanism. For ease of illustration, we depict these three mechanisms graphically in Figure 1.

While these three routes to meaning appear to play a role in normal comprehension, the syntactic route is the final determinant of meaning. In fact, one of the reasons language has the power it does have is that we can express unexpected and counterfactual thoughts, which would be impossible if the lexical-pragmatic route was the final determinant of propositional meaning.

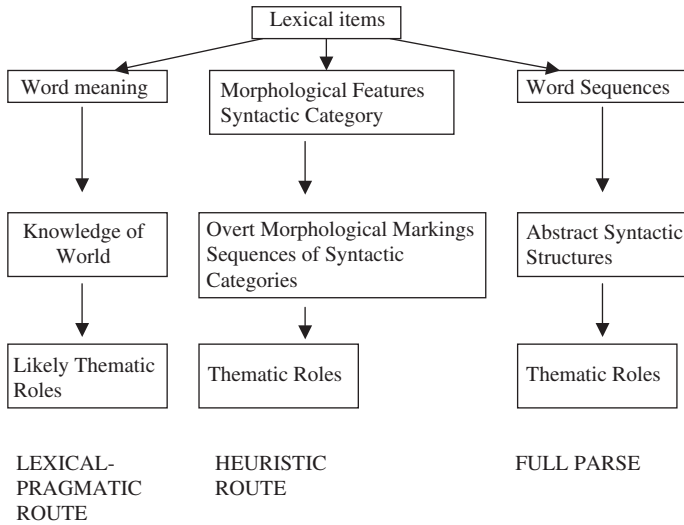


Figure 1. Schematic of three routes (mechanisms) whereby propositional information is constructed in the process of sentence comprehension.

We should mention that this tripartite division of sentence comprehension mechanisms is not universally accepted. Some researchers do not think of syntactic representations as hierarchically structured sets of categories; rather, they believe that surface cues to meaning, such as the order of words in a sentence, the presence of function words such as relative pronouns and prepositions, morphological markers on nouns and verbs, and other such features, are all the syntax that there is (e.g., Dick et al., 2001). In the minds of these researchers, what we are calling a heuristic mechanism is really the only syntactic mechanism that exists. Even more radical models deny the existence of heuristics and maintain that comprehension is entirely based on the probability that one specific word will follow another (e.g., MacDonald & Christiansen, 2002). (We consider these more radical models inadequate at present.) Finally, among researchers who believe complex syntactic structures are constructed, there is considerable disagreement as to what their exact nature is. We will continue to refer to a tripartite model of sentence comprehension, in part because it specifies the largest number of basic mechanisms and in part because it has some utility in approaching aphasic disorders. We will mention a few specific hypotheses about syntactic structures, where they are relevant to researchers' models.

The reader will appreciate that sentence structure and processing are quite complex, and will not be surprised to learn that sentence comprehension is not immune to disturbance after neurological injury. In fact, however, the appreciation that sentence comprehension could be affected without major disturbances of comprehension of single words came relatively late in the history of aphasiology. The existence of patients with disturbances of the ability to understand single words was first reported by Karl Wernicke

in 1874; the identification of patients who have problems understanding sentences despite good word comprehension came 100 years later. Caramazza and Zurif (1976) described patients who could match sentences to pictures when the thematic roles in the sentence could be inferred from lexical meaning and real world knowledge. Thus, these patients could match sentences such as (1) to pictures, showing that they did extract thematic roles from spoken sentences. They could not, however, do sentence – picture matching when the thematic roles could not be inferred on the basis of word meaning and knowledge of the world and the sentences were syntactically complicated, as in (2). These patients also systematically misunderstood nonsensical sentences, assigning thematic roles according to real world likelihood rather than syntactically dictated meaning in sentences such as (4).

#### 4. The boy the apple is eating is tall.

Caramazza and Zurif (1976) concluded that these patients had lost an *algorithmic* method of interpreting sentences. By the term algorithmic, they had in mind the syntactic mechanism we have discussed, that assigns the syntactic structure of a sentence and determines thematic roles by integrating lexical meanings into that structure. They claimed that the patients retained *heuristic* methods of interpreting sentences. By heuristic, they were referring to what we have called the lexical-pragmatic comprehension mechanism, that assigns thematic roles according to the likely interactions in the real world of the items referred to by lexical items.

The interpretations of these data have stood essentially intact for over 30 years. It is now commonplace to use the combination of a deficit in understanding semantically reversible, syntactically complex sentences and the retained ability to understand semantically irreversible sentences with the same syntactic structures as the basis for diagnosing a disorder of syntactic comprehension.

We shall spend most of this chapter exploring what researchers have said about these syntactically based sentence comprehension disturbances, but before we do, we note that, if the tripartite model of sentence comprehension we have outlined above is correct, other disturbances of sentence comprehension would be expected to occur. We will briefly outline some of them before going on to syntactic disorders.

The most severe disturbance of sentence comprehension would be one in which a patient could not appreciate the semantic values that are associated with sentences. None of the three routes to meaning would be working. A patient like this would know that a sentence like (1) mentioned a boy, eating and a cake, but would not know that the boy was the agent of eat and the cake was its theme. One way this could happen would be for a patient to have lost the most fundamental knowledge about the way the world works; s/he might not know that boys eat cakes, or might not be able to access that knowledge. However, the patient's disturbance might not be so severe; s/he might know such things about the world but not appreciate that sentences convey this type of information. Such deficits have not been described.

A second type of deficit might consist of the retention of only the lexico-pragmatic route to meaning. A patient with this type of deficit would be able to understand semantically irreversible sentences such as (1) but have no systematic way to assign meaning in semantically reversible sentences, such as (2). Sentences such as (2) would be ambiguous; even a simple active sentence such as “John pushed Mary” would be ambiguous. There may be patients with this sort of problem. Caplan, Baker, and Dehaut (1985) examined aphasic patients’ abilities to enact the meanings of semantically reversible sentences with a variety of syntactic structures, and clustered the patients into groups that turned out to contain patients with roughly equally severe disorders of this ability. The worst performing patients were only able to enact sentences with very simple structures, such as active sentences, and were as likely to assign the reverse set of thematic roles as the correct set. These patients appear to have retained the ability to understand that sentences convey thematic roles but not to use either heuristics or a full syntactic analysis to assign these aspects of meaning. Whether they use the lexico-pragmatic route to meaning is, however, not known. If they do, they should behave like the patients described by Caramazza and Zurif (1976); i.e., they should be able to understand semantically irreversible sentences regardless of their syntactic structure (as in sentence (1)) and they should systematically misinterpret sentences whose structure expresses an unlikely or impossible even (as in sentence (4)). We do not know if this is the case, since virtually no patients who have been tested for syntactically based comprehension disorders using semantically reversible sentences have been examined for the ability to understand semantically irreversible sentences since Caramazza and Zurif’s (1976) first report.

Thus, the full range of disorders of sentence comprehension remains to be explored. Researchers have focused on patients’ abilities to construct syntactic structures and use them to assign aspects of propositional meaning. This has been done because syntactically based sentence comprehension disorders have a number of potential implications for two types of theories: theories of syntactic representations and theories of sentence analysis (parsing). We will draw out these implications as we proceed.

Before turning to disorders affecting syntactic comprehension, we make one final note; namely, that issues relating to sentence comprehension have theoretical implications have been explored, that we will not discuss in depth here. Chief among these is the relationship between the ability to understand sentences and the integrity of various short-duration memory systems. Very briefly, it is widely believed that humans have a specialized memory system that allows them to retain small amounts of information for short periods of time and to operate on this information in the service of performing a task. This “working memory” is made up of several capacities, including a “central executive” that accomplishes operations as well as having some storage functions and various “slave systems” that maintain information in a phonological or visual form (see, e.g., Baddeley, 1986). Another short-duration memory system, possibly related to the working memory described by Baddeley and others, that maintains information in a semantic form, also seems to exist (Martin & He, 2004). A considerable amount of research has gone into trying to understand what happens to sentence comprehension when these memory systems are affected by neurological disease. In

the opinion of the authors, though reductions in these short-duration memory capacities do lead to disturbances of comprehension, none of these systems is required for syntactic comprehension. This is a long and technically somewhat detailed argument, and, since we believe these systems are not relevant to the main focus of this chapter, we will spare the reader. The issues can be reviewed in Caplan and Waters (1990, 1999) and the ensuing discussion, as well as in Just and Carpenter (1992), Waters and Caplan (1996), Caplan and Waters (2002), and Martin and He (2004). In general, the disorders of syntactically based comprehension we will discuss do not appear to be secondary to other impairments, though this has yet to be unequivocally established.

## 2. SYNTACTIC COMPREHENSION DISORDERS: IMPAIRMENTS TO SPECIFIC STRUCTURES AND OPERATIONS

As we indicated above, in the first description of syntactic comprehension disorders, Caramazza and Zurif (1976) suggested that the patients they had studied had lost the “algorithmic” basis for sentence comprehension. These authors thought that all syntactic operations had been affected in these patients. Other authors expressed similar views at about the same time (Berndt & Caramazza, 1980; Caplan, 1985). However, alternatives to this analysis of the deficit in these patients were quickly expressed, which postulated more limited impairments of syntactic processes.

The best known of these hypotheses is now known as the “trace deletion hypothesis.” It was first expressed by Yosef Grodzinsky in 1986 and has been developed in many of his subsequent publications (for a summary, see Grodzinsky, 2000). Prior to Grodzinsky’s (1986) paper, several research groups had found that certain aphasic patients performed at chance in sentence–picture matching when presented semantically reversible passive sentences (5) and sentences with object relative clauses (6), and normally or above chance when presented with active sentences (7) and sentences with subject relative clauses (8):

5. The man<sub>i</sub> was pushed *t*<sub>i</sub> by the woman.
6. It was the man<sub>i</sub> who(m) the woman pushed *t*<sub>i</sub>.
7. The woman pushed the man.
8. It was the woman<sub>i</sub> who *t*<sub>i</sub> pushed the man.

We have annotated (5) – (8) with the symbols [*t*<sub>i</sub>] and [<sub>i</sub>] to illustrate the structures relevant to the trace deletion hypothesis. This hypothesis argues that difficulty with passives and object relatives stems from a single deficit affecting an aspect of the syntactic structure common to both sentences types—an element known as a “trace” (denoted with the symbol [*t*]) according to Chomsky’s (1986, 1995) theory of syntax.

Chomsky pointed out that sentences (5) and (7) conveyed the same thematic roles, as did sentences (6) and (8). Chomsky has argued that the way to explain why these pairs of sentences express the same thematic roles is to say that the syntactic structures of the

members of each of these pairs of sentences are the same in critical respects. Specifically, Chomsky has argued that thematic roles are always assigned the same way: the “external argument” of the verb (in these cases the agent) is assigned to the subject of the verb; the first “internal argument” of the verb (here the theme) to its object. To preserve this universal feature of the way syntax maps onto semantics, Chomsky postulated that there is more than one level of syntactic structure. According to this theory, in sentences such as in (5), (6), and (8), noun phrases in the surface form of the sentence have moved from their original positions, leaving “traces” of themselves in their former positions. These traces are shown in their “underlying” positions in (5) – (8) as [*t*]. The rules of sentence interpretation we outlined above apply to these traces, assigning the correct thematic roles to them. The noun phrases that have moved are connected to (“co-indexed with”) their traces (shown with the subscript symbol [<sub>*i*</sub>]). Grodzinsky suggested that some aphasic patients lose the ability to co-index traces and their related noun phrases. When this happens, a heuristic that takes the first noun of the sentence as the agent of the first verb in the sentence applies, leading to errors in (5) and (6). (7) is interpreted correctly by the syntactic route, and (8) by the application of this heuristic.

Other researchers have also proposed that specific syntactic operations are subject to individual disruption in aphasia. Caplan and Hildebrandt (1988) presented a series of case studies in which they looked for deficits affecting all the different types of co-indexation operations involving referentially dependent noun phrases postulated in Chomsky’s model of syntactic structure. These include the traces discussed above, reflexives (*himself*, *herself*), pronouns (*him*, *her*), and, in English, another type of “empty” noun phrase that occurs as the subject of infinitives, shown in (9).

9. John promised Bill **PRO** to help.

Caplan and Hildebrandt (1988) argued that all of these different co-indexation operations could be damaged in relative isolation in individual patients. The damage was “relatively” isolated in some cases, because all patients who had trouble with either reflexives or pronouns, or both, always had trouble with both *trace* and *PRO*. Caplan and Hildebrandt suggested that this was because reflexives and pronouns are physically present in the acoustic signal, whereas *trace* and *PRO* are not and their presence has to be derived by rules, making them harder to process. This is one feature of the results that led Caplan and Hildebrandt to suggest that there were two main sources of aphasic impairments in syntactic processing: deficits in the ability to process specific representations, and some sort of limitation on how much processing could go on. They called the latter impairment, which affects all sentences that have a certain level of complexity or more, a “resource limitation.” We shall discuss this aspect of the problem below.

Assuming that there are structure-specific impairments in syntactic processing, how many of them are there? No one knows. Syntax consists of many different types of representations, and, in principle, any one could be affected. Perhaps some patients cannot construct complex noun phrases (such as *the father of the bride*), others cannot construct co-ordinate noun phrases (such as *the father and the bride*), and yet others cannot

construct other aspects of the hierarchical structure of sentences. Also, perhaps there are higher-order disturbances that affect an entire class of structures. Grodzinsky's trace deletion hypothesis is actually a disturbance of this sort, since, technically, the type of trace found in a passive sentence (an "NP-trace") is different from that found in a relative clause (a "wh-trace"), and the trace deletion hypothesis suggests that both are affected. Another example of a proposed high-level deficit is a deficit in the ability to build any hierarchical construction at all, suggested by Caplan (1985). If many deficits of this sort could be shown to exist, the patterns of deficits in aphasics would bear on many important questions about the syntactic structure of sentences and how that structure is processed. For instance, if there really are patients whose deficits affect all hierarchical structures, their performances could be used to test the theory that syntactic structures are only collections of surface cues to meaning. It is therefore of great interest to find patients with such isolated deficits. Unfortunately for the use of aphasia to develop theories of normal syntactic structure and processing, this turns out to be much harder to do than most researchers seem to think.

### 3. STRUCTURE-SPECIFIC IMPAIRMENTS: PROBLEMS AND NEW DATA

To demonstrate that a patient's deficit is restricted to a specific structure or operation, it is necessary to show several things. First, the ability to understand sentences that do not contain that structure or require that operation must be intact. Second, all sentences that do contain that structure or require that operation must be affected. Third, all these sentences must be affected on all tasks that require comprehension. These requirements for documenting structure-specific deficits have not been met in most studies in the literature on the topic.

Let us first consider the integrity of the ability to understand sentences that do not contain a particular structure or require the operation in question. Many studies compare comprehension of active and passive sentences, and object and subject relatives. However, other critical sentence types have only rarely been tested in the same patients as sentences with traces that cannot be understood using heuristics. While there is no end to the number of sentences that could be tested, and proponents of any particular theory could at some point legitimately complain that skeptics are just continually raising the bar, some baseline testing is reasonable. For instance, sentences with reflexives and pronouns are obviously important to test to show that a deficit affecting co-indexation is restricted to traces and spares sentences with other types of co-indexations. To our knowledge, however, Caplan and Hildebrandt (1988) and Caplan, Hildebrandt, and Makris (1996) are the only studies in which sentences with reflexives and pronouns have been presented to the same patients as were tested on passives and object relatives. The specificity of the deficit in all other patients said to have a deficit in co-indexing traces is not well established.

Further, all sentences that contain the structure should be affected. That means that a patient with a trace deletion deficit should have difficulty with passives with a *by*-phrase

and those without a *by*-phrase, as well as cleft-object sentences, sentences with object relative clauses, and sentences with object questions that begin with *wh*-words. Again, no individual patients have been tested on all these sentence types.

Finally, a patient with a deficit affecting one structure or operation should fail on all comprehension tasks when presented with sentences that require processing that structure or applying that operation. If a patient can enact the meaning of a sentence but not match its meaning to a picture, s/he has a deficit, but the deficit cannot be one that affects the construction of a particular structure that is essential to the comprehension of the sentence. Once again, patients have very rarely been tested on more than one task requiring comprehension. When this has been done, they often do not perform the same way on both tasks. Cupples and Inglis (1993) reported two cases who performed poorly on passive sentences in sentence–picture matching but well in an enactment task. Caplan, Waters, and Hildebrandt (1997) reported a correlation of about 0.6 on average for performance on the same sentences in enactment and picture matching in 17 aphasic patients – a significant degree of correlation but one that leaves plenty of room for patients to do very differently in one task and another.

We have recently completed a study of 42 aphasic patients and 25 age and education matched controls in whom we considered these issues (Caplan, Waters, DeDe, Michand, & Reddy in press a). The ability to structure and understand three syntactic constructions – passives, relative clauses, and sentences with reflexive pronouns – was tested by having participants respond to pairs of sentences in which the baseline sentence did not contain the construction/element in question or could be interpreted on the basis of a heuristic and the experimental sentence contained the structure/element and required the assignment of a complex syntactic structure to be understood. Each construction was tested with two experimental/baseline contrasts, with 10 examples of each sentence type. Subjects were tested in object manipulation (enactment), sentence–picture matching, and grammaticality judgment tasks, the latter two with both whole sentence and self-paced listening presentation conditions, using digitized computer-delivered auditory stimuli. Accuracy and end-of-sentence reaction time (RT) (in sentence–picture matching and grammaticality judgment) were measured. We considered performance on the object manipulation and sentence–picture matching tasks, which require comprehension (as opposed to grammaticality judgment, which may not), with whole sentence presentation, which is the most natural form of language presentation in a hearing subject.

We identified good and poor performance in three ways. (1) “Normal” vs. “abnormal” accuracy: normal accuracy was not less than 1.74 SDs below the normal mean; abnormal accuracy was below this level. (2) Above chance vs. at or below chance performance. (3) “Normal speed” vs. “slow:” RTs that were 1.74 SDs longer than the normal mean RT were considered “slow.” We considered a deficit to be present if a patient had poor performance on an experimental (test) sentence and good performance on the corresponding baseline sentence by one of these criteria, and did not show evidence for a possible speed-accuracy trade-off. We searched the individual case data for four types of deficits:



*Task-specific, sentence-specific deficits:* a deficit for one experimental sentence of a structural type in only one task.

*Task-specific, structure-specific deficits:* a deficit for both experimental sentences of a structural type in only one task.

*Task-independent, sentence-specific deficits:* a deficit for one experimental sentence of a structural type in both tasks.

*Task-independent, structure-specific deficits:* a deficit for both experimental sentences of a structural type in both tasks.

One patient had a task-independent, structure-specific deficit (for passives); 17 had task-independent, sentence-specific deficits (8 by the criteria for normal accuracy, and 9 by “chance” criteria); and 33 had task-specific sentence- or structure-specific deficits. No patient showed a stable pattern of performance across sentence types and tasks that could be clearly interpreted as due to a deficit in the ability to co-index traces.

These data point to two important features of aphasic performances.

First, they indicate that task-independent structure-specific deficits are rare. This finding casts doubt on the conclusion that is often drawn on the basis of accuracy of performance in a single task that an individual patient has a deficit in a particular parsing or interpretive operation. Because of the theoretical importance of deficits of particular parsing or interpretive operations, the rarity of task-independent structure-specific impairments needs to be confirmed (or shown to be somehow an artifact of our study). Task-independent sentence-specific deficits are more common, and some can be explained in terms of specific operations in relation to specific tasks (e.g., poor comprehension of truncated passives in sentence–picture matching can be due to inability to infer an agent in a picture-matching task; see Caplan, DeDe and Michand, in press, for discussion).

Second, these results point to the ubiquity of task-specific, sentence- and structure-specific impairments. We therefore turn to these disorders.

#### 4. TASK-SPECIFIC IMPAIRMENTS

We have described dissociations in performance in two tasks that require comprehension – enactment and sentence–picture matching – but dissociation in performance on two tasks on the same sentences has been known for over two decades. Dissociations in performance on sentence–picture matching and grammaticality judgment – chance performance on sentence–picture matching and good performance in making judgments about grammaticality – were first reported by Linebarger Schwartz, and Saffran (1983). This pattern – good grammaticality judgment and poor sentence–picture matching – has been attributed to retention of the ability to construct syntactic structures but not to use them to assign meanings, a deficit that has been termed one of “mapping” syntactic structures onto meaning. The pattern of performance provides evidence that

processing of syntactic structure is to some extent separate from the processing of propositional information, though it does not resolve controversies that exist in the literature about how different types of information, including information about the likelihood and plausibility of thematic roles, interact in the normal assignment of sentence and discourse meaning.

Dissociations in processing one structure in two comprehension tasks (object manipulation and sentence–picture matching) are far more difficult to explain. If a patient can understand a sentence as shown by, say, reliably matching its meaning to a picture, and is capable of enacting the same meaning derived from a different structure, it is not immediately clear why s/he cannot enact the meaning of that sentence. The form of a sentence itself is not needed to demonstrate the meaning of the sentence; all that is needed is that the sentence be understood. It might be the case that one task is harder than another, leading to difficulty with complex sentences in only the harder task. However, task-specific, structure- and sentence-specific deficits occur in both enactment and picture matching, ruling out task complexity as the sole reason for the occurrence of these performances. We can imagine two reasons why such patterns might occur (see Caplan, DeDe and Michand, *in press*, for fuller discussion).

The first is related to the fact that syntactic processing and the performance of a task operate in cascade on-line (Tanenhaus et al., 1995). If listeners are planning actions or matching thematic roles depicted in pictures to sentences as a sentence unfolds, the syntactic structure of a sentence, as well as its meaning, is being activated as the task is being planned and/or performed. Task-specific, sentence- and structure-specific deficits could arise from an inability to integrate the demands of one task with sentence structure building and semantic interpretation. One can conceive of this explanation as postulating that comprehension in the service of a task creates a dual-task/divided attention situation and individual patients have difficulty with particular task pairings (a sort of limitation on their ability to divide attention).

The second account of task-specific sentence- and structure-specific deficits is based on the observation that syntactic structure determines not only propositional semantic values such as thematic roles, but discourse-level representations, such as focus, presuppositions, etc. as well. While discourse level representations are logically irrelevant to the performance of the tasks we used, which only require consideration of thematic roles and co-indexation, they may affect performance, as is well documented for many other implicit processes, and may do so differently in different tasks. For instance, passives and cleft-object sentences place the theme in the discourse focus. A patient who has difficulty formulating an action in which the focused element is the theme, but who can match a focused theme to depictions of thematic roles, would have trouble with passive and cleft object sentences only in enactment. On this account, task-specific sentence- or structure-specific deficits arise in the course of mapping the combination of discourse and propositional meanings onto tasks.

These – or other – accounts of these deficits need to be tested.

## 5. RESOURCE REDUCTION

The results we have reviewed show that performances on many sentence types were often affected in one patient. In fact, in our study, the average of the correlations of accuracy of performance for different sentences across pairs of tasks (the two tasks that require comprehension – enactment and picture matching, and the whole sentence and auditory moving windows versions of grammaticality judgment and picture matching) was almost the same as the average of the correlations of accuracy on the same sentences across the same tasks. In addition, when sentences were randomly divided into two sets in each task, the average of the correlations of accuracy of performance for different sentences within each task was about the same as the average of the correlations of performance on the two sets of the same sentences (mean split-half reliability) in the task. All these results point to the fact that patients' performances were highly correlated on different sentence types both within a task and across tasks. It does not seem that what determines a patient's performance is a specific deficit in assigning a particular syntactic structure or using that structure to comprehend a sentence. Rather, something more general, that affects many sentences, is operating in most cases.

One way to gather clues as to what that something might be is to utilize an approach to data analysis known as factor analysis. Factor analysis has not been used very often in aphasia. Factor analysis looks for how the variance in performance on the part of each participant is best captured by "factors" that are weighted for performance on all sentence types. The number of factors that are needed to get to the point that additional factors no longer account for significant amounts of variance, and the weight of each sentence type on each factor, are cues as to what processes are the major determinants of the variance in participants' performances. Factor analysis can be both exploratory and confirmatory; in exploratory factor analysis, no assumptions are made about the data, while in confirmatory factor analysis, assumptions about how sentence types will load on factors and/or about how many factors will be retained are tested. Exploratory factor analysis is just that – exploratory, and there are different approaches to constructing factors. In "unrotated" factor analysis, factors are constrained to capture the most variance possible and to not interact (they are "orthogonal"). In "rotated" factor analysis, factors are constrained to maximize the weights of individual sentences as well as to capture the most variance possible, and may or may not interact. Rotated factor analyses thus are more likely than unrotated factor analyses to group together sentences with similar structures, if indeed such groups of sentences account for much of the variance in performance.

We performed both unrotated and rotated exploratory factor analyses and confirmatory factor analyses on the patient accuracy scores on each task separately and in various combinations, and on the patient and control RT scores for the tasks with RT measurements (Caplan et al., in press a; DeDe, Caplan, & Waters, 2006; DeDe, Caplan, Waters, Michand, & Reddy, 2005b). Controls' accuracy scores could not be analyzed by this technique because of the uniformly good performance ("ceiling effects") and consequent lack of variation, leading to a statistical property known as multiple colinearity that makes factor analysis inappropriate.

The unrotated factor analyses of patient accuracy and of both patient and control RT data had the same basic structure. In all cases (including the analyses of accuracy in all combinations of tasks), using standard guidelines for how many factors to include and how to consider that a sentence type loaded on a factor, these analyses resulted in first factors that were roughly equally positively weighted for all sentence types that accounted for the great majority of the variance in performance (always above 70%). In only a few cases was even a second factor retained in the analysis. This structure of unrotated factor analysis of accuracy in enactment tasks has been previously reported (Caplan et al., 1985, 1996). Confirmatory factor analyses showed that one-factor solutions were preferred to two-factor solutions for these analyses, that the structure of the patient and control factor analyses of RT data did not differ substantially, and that imposing a factor structure that reflected the groupings of sentence types that corresponded to the syntactic structures that were varied in the creation of the stimuli did not account for more variance than the structure that emerged from the unconstrained analysis. Rotated factors, of course, returned more factors, but, again, none of the factors in these analyses contained groupings of sentence types that corresponded to the syntactic structures that were varied in the creation of the stimuli.

One question we might ask is what types of features load on first factors in factor analyses of comprehension performances of the sort we have reported. Interactive activation models of sentence comprehension (e.g., MacDonald, Pearlmuter, & Seidenberg, 1994) maintain that knowledge about the frequency of occurrence of particular forms of words, the frequency of occurrence of sequences of words, the likelihood of thematic roles, and other aspects of knowledge about language and the world all interact in a single process to determine sentence meaning. If so, one might expect that all such information might load on first factors in analyses of the sort reported here. The data that would show this does or does not happen have yet to be collected. We do know that not all factor analyses of aphasic performance have the same structure as that described above. When we performed factor analyses over both performances on the screening tasks that the patients had taken and performances on the syntactic tasks, the unrotated factor analysis was quite different and returned several factors that reflected performance on these different tasks. Whether the different types of information that some researchers believe is used together in the process of comprehension would load on a single factor, like the different structures we have studied in the comprehension tasks, or on different factors, like different tasks we used, remains to be determined.

The factor analyses we have performed do not support the idea that individual patients' losing the ability to perform particular syntactic operations determines much of aphasic patients' performance on these tasks. What these analyses show is that a single factor, roughly equally weighted for all sentence types, accounts for the great majority of variance in performance. What might this factor be due to?

One way to find out is to compute what is known as a "factor score" for each participant, in which the weights for each sentence on a factor are multiplied by the performance of each participant on that sentence and the results summed for each participant. These

scores give a measure of how each participant has performed on the processes that underlie that factor. Patients' "first factor scores" correlated extremely highly with their overall accuracy on each task, and less well with the difference between their performances on the experimental, syntactically complex sentences and the baseline, syntactically simple sentences. These first factors thus reflect a process that applies to all sentences and that is related to how well patients perform on each task.

There are two possible mechanisms that could produce such factors. One is a disturbance in some process that affects how word meaning is activated, because disturbances of activating word meaning could affect performance on all sentence types. There are three reasons this is not likely to be the mechanism that produced these first factors. First, patients were screened to eliminate cases with significant disturbances of lexical processing. Second, the vocabulary in the stimulus sentences was kept very simple, and was within patients' grasps. Third, the same factor structure was seen in the factor analyses for RTs in control subjects, and this is not easily attributable to lexical access effects.

The second possible mechanism that could have produced these factors is what we may call a reduction in the "processing resources" available for sentence comprehension, syntactic processing, and performance of these tasks, in a patient. Resources allow a task to be performed. For instance, most readers of this chapter will be able to multiply  $7 \times 14$ , but not  $39 \times 458$ , in their heads; they run out of something when they try the second of these tasks. Detailed analyses may one day make it clearer what the limitation of processing is in the second of these tasks, but, for now, the best we can say is that they have insufficient processing resources to simultaneously hold the intermediate products of computation in their minds and continue to perform the necessary mathematical operations.

In addition to the results of the factor analyses, other aspects of aphasic performance suggest processing resource reduction as a mechanism underlying comprehension disorders.

At the level of single case analyses, Caplan and Hildebrandt (1988) reported that patients often were able to show comprehension of a sentence that required one type of syntactic operation, but failed when a second operation was also needed. Thus, patients could understand sentences such as *John said that Bill helped **him***, that requires the pronoun **him** to be co-indexed with **John**, and *John promised Bill **PRO** to help*, that requires the abstract category **PRO** to be co-indexed with **John**, but not *John promised Bill **PRO** to help **him***, that requires both the pronoun **him** the abstract category **PRO** to be co-indexed. This suggests these patients can accomplish one, but not two, co-indexation operations, consistent with a reduction in processing resources.

At the level of performances of patient groups, if resource reduction is a major determinant of performance, patients who perform less well would be ones with the fewest resources. They should perform least well on the most complex sentences, since these require the most resources. This should lead to interactions of patient groups defined by overall performance level and sentence type, such that there are greater effects of sentence complexity in less well performing patients. Such interactions have been found by Caplan

et al. (1985) and in the study we have been reporting on in this chapter (Caplan et al., in press a), though not by all researchers (cf. Dick et al., 2001, whose patients may have performed too well to show these effects).

The concept of resources is one of most maligned in experimental psychology, and, given how important we think a reduction in some capacity that affects all sentence types across the board is in aphasia, it is worthwhile to take a moment or two to relate it to serious models of cognitive functioning. We can define processing resources as features of mental functioning (of a “functional cognitive architecture”) that allow cognitive operations to proceed and that affect their operating characteristics, but that are not themselves a computational operation or a type of knowledge. Most computational models have such features (though not all; see, e.g., Meyer & Kieras, 1997). Examples are the look-ahead buffer in an L/Rk grammar (Marcus, 1980), the activation units in a hybrid procedural model such as CC-READER (Just & Carpenter, 1992), the hidden units in a Boltzman machine (Seidenberg & McClelland, 1989), and the recurrent units in a simple recurrent network model (Elman, 1991). These features of these models support all the operations of the model (they are not specific operations or computations), enabling computations to proceed and imposing limits on their operation. In some cases they affect the details of aspects of processing (Seidenberg & McClelland, 1989); in others, they affect the types of computations that can be successfully accomplished (e.g., Minsky & Papert, 1969; Rumelhart, Hinton, & Williams, 1986).

While the notion of processing resources is thus not “vague,” it does face numerous problems. One problem is that there are many well-defined models of resources, each embedded in a specific model of function, and choosing between models of resources cannot be done without choosing between the processing models themselves. For instance, with respect to sentence processing, the resource limitations in Just and Carpenter (1992) and MacDonald and Christiansen (2002) are both well-defined aspects of the architecture of those models and which, if either, is true of human parsing will depend on whether parsing is procedure-based or connectionist in nature. Caplan and Hildebrandt (1988) argued that, given the uncertainty regarding models of parsing and sentence interpretation, it was premature then to commit to a model of resources; we feel that, since 1988, the range of plausible parsing/interpretation models has increased (cf. MacDonald et al., 1994; MacDonald & Christiansen, 2002; Tabor & Tanenhaus, 1999), making a commitment to a particular model of resources even less appropriate now.

A second problem is to determine what tasks and functions are supported by a single processing resource system. We have argued that there are “specialized” and “general” pools of processing resources that support verbal tasks (Caplan & Waters, 1999). The “specialized” pool supports the initial assignment of meaning, including syntactic analysis, and the “general” pool supports the use of the meaning of a sentence to accomplish a task. As the discussion of task effects above suggests, this characterization of a division of resource systems is likely to be too simple, although some division of these systems seems justified. Aphasics seem to have reductions in both systems (assuming more than one system exists).

Finally, we can try to relate this global reduction in the ability to process sentences to abnormalities in particular processes. We mentioned above that many researchers have tried to relate sentence comprehension disturbances to limitations of one or another short-duration memory system; while we doubt that the reduction in “resources” that we are attempting to understand is simply a reduction in such memory systems, this remains an open question. Some researchers have suggested that there are impairments in the speed of activation or decay of syntactic representations (e.g., Haarman & Kolk, 1991; Frazier & Freiderici, 1991) in aphasic patients. Such impairments could have the same effect as a reduction in processing resources, since they would affect the construction of more complex representations more than simple representations (Haarman, Just, & Carpenter, 1997). In other areas of psychology, such as the study of cognitive aging, efforts have been made to separate the effect of reductions in processing speed from reductions in processing resources (Salthouse, 1991). This is another line of investigation that needs to be undertaken in aphasic patients to understand these types of results.

## 6. ON-LINE PROCESSING

We have thus far considered patients’ performances at the end of a sentence. These end-of-sentence measures are important, because they tell us what patients finally decide a sentence means. However, they are not necessarily a reflection of what goes on while patients perceive each word in a sentence and integrate that word into the structure they are building. Such “on-line” processes could show the same picture as end-of-sentence measures, or they could differ from end-of-sentence measures in either direction. They could be abnormal in cases where end-of-sentence measures are normal – suggesting that patients have more subtle problems (such as delayed activation) than are seen in end-of-sentence performances. This would hardly be surprising, but important to document for clinical purposes and because such abnormalities could inform theories of processing (by providing evidence for structure-specific, task-independent deficits, for instance). Less intuitively, on-line performances could also be normal in cases where end-of-sentence measures are abnormal. This would suggest that patients run into difficulty with using what they have understood to perform a task.

There are very few studies in which patients have been tested for on-line performance. One such series of studies uses the cross-modal lexical priming (CMLP) task. In this task, a subject listens to a sentence and, at some point in the sentence, a series of letters appears on a computer screen; the subject must push a button to indicate if the letter string is or is not a word (a lexical decision task). Unbeknownst to the subject, in half the cases in which the letter string is a word, it is semantically related or associated with a word that was previously presented in the sentence. In these cases, reaction times are typically faster than when the real-word stimuli are not related to previously presented words.

Swinney and his colleagues (Swinney & Zurif, 1995; Swinney, Zurif, Prather, & Love, 1996; Zurif, Swiney, Prather, Solomon, & Bushell, 1993) capitalized on this “priming” phenomenon to test for the co-indexation of traces. The normal finding is that

cross-modal lexical decisions are speeded for words related to the head noun of a relative clause at the point of the trace in the relative clause and not immediately before that point (for issues regarding this technique, see McKoon, Ratcliff, & Ward (1994), McKoon & Ratcliff (1994); Nicol et al., 1994). Two studies of aphasics found that Broca's aphasics did not show such priming, and were interpreted as consistent with the TDH hypothesis (Swinney & Zurif, 1995; Swinney et al., 1996; Zurif et al., 1993). These data are not convincing evidence for a disturbance affecting processing traces, however, for several reasons. One is that the patients' deficit may not have been restricted to traces; performance on sentences with other referentially dependent items, such as reflexives, was not tested (this is the same issue regarding reflexives as was raised above for off-line data). Second, the disturbance that was documented in these patients may have arisen at the lexical, not the syntactic, level. The Broca's aphasics tested in these two studies were not tested for word-to-word priming effects, and the absence of any priming effects may have been due to a failure of these patients to show lexical priming (Milberg & Blumstein, 1981; Blumstein, Milberg, Shrier, 1982).

The second of these problems leads us to the beginnings of a discussion about how to interpret average performance of groups of patients selected according to classical clinical syndromes, such as Broca's aphasics. Some Broca's aphasics have been shown to have delayed, not absent, priming (Prather et al., 1992, 1997), and the authors of the CMLP studies cited above have argued that the time course of activation of the antecedent of the trace is delayed in Broca's aphasics. This was confirmed by later studies from Swinney's lab, which have shown priming for Broca's aphasics several words after the trace appears in these sentences (Love, Swinney, Zurif, 2001). However, delayed priming may only occur in some Broca's aphasics; as noted, in some studies, no priming was seen in Broca's aphasics, and some of these studies used long intervals between the prime and the target. In addition, not all priming studies of Broca's aphasics' sentence processing have shown delayed priming at the point of co-indexation of traces; using a purely auditory paradigm, Blumstein et al. (1998) found normal CMLP effects at the point of traces in such patients. It is dangerous to assume that a patient will perform in a particular way on a task on which s/he has not been tested, just because other patients with the same clinical diagnosis perform in one way or another on that task. The problem is that, more often than not, patients with the same clinical diagnosis perform one way *and* another in different studies; variability in performance of patients with clinically defined syndromes is as much of a problem as variability in performance of single cases.

Two other laboratories have also studied on-line sentence processing in aphasia.

Dickey and Thompson (2004) used an auditory anomaly detection task, in which 8 agrammatic aphasics and 24 young normal participants listened to sentences and pressed a button when the sentences stopped making sense. Critical sentences contained an anomaly in object relative clauses or conjoined clauses. Controls were able to reject anomalous sentences on-line. Patients who had not received remediation directed at syntactic comprehension were unable to discriminate anomalous and non-anomalous sentences; patients who had received speech therapy directed at the comprehension of



these sentences performed as did normal controls. The authors took these results as evidence that individuals with agrammatic aphasia “retain some gap-filling capacity” and that treatment can improve their ability to make use of this capacity. The second of these conclusions seems justified; the first can be questioned (perhaps the untreated cases would never have developed this capacity, even with treatment).

Thompson, Dickey, and Choy (2004) and Dickey, Thompson, and Choy (2005) also reported the results of a study using eye tracking to examine the time course of comprehension. Tanenhaus and his colleagues (Tanenhaus et al., 1995) pioneered the use of eye tracking as a means of exploring this process. They showed that, when subjects listen to sentences while viewing arrays of objects, the pattern of looks to objects is revealing about what structures the subjects are building and what they are comprehending on-line. Thompson et al. (2004) reported the results of this type of eye fixation monitoring in four Broca’s aphasics, and reported differences between the eye tracking performance of patients and controls when they were presented with the critical verbs in object-extracted structures. However, Thompson et al. (2004) found that their patients made abnormal fixations only in one structure, and suggested that the complexity of the object-extracted structure affects performance. In addition, in a follow-up report, Dickey et al. (2005) reported that a larger group of 12 patients showed normal eye fixations at the critical points in all object-extracted sentence structures, despite making many more errors than controls. When they made errors, patients tended to look at objects that normals did not look at later in the sentences.

Dickey interpreted this pattern as showing that these patients initially always processed the sentences normally, to the point of initially understanding the thematic roles in object extracted structures in the same way as normals. In his view, errors were due to patients’ initial understanding being overridden at a later point in processing by an alternative interpretation of the sentences. Such interpretations could result from the application of heuristics, and appear to be present in normals as well as patients (Ferreira, 2003). The difference between aphasics and normals, on this view, is not that aphasics’ initial comprehension processes are disturbed, but that aphasics are less capable than normals of determining that the syntactically derived meaning, as opposed to the heuristically derived meaning, is correct. A similar possibility was suggested by Caplan and Waters (2003). This deficit would be a failure of some sort of control, or perhaps a labeling, process, not of assigning syntactic structure or using it to determine sentence meaning *per se*.

This conclusion may be correct, but it seems unlikely to be the whole story. How likely is it that brain damage spares initial on-line syntactic processing and that comprehension disorders are all due to failures in adjudication between syntactically licensed and heuristically and pragmatically derived meanings? Our studies suggest this is not the case. In the auditory moving windows version of the sentence–picture and grammaticality judgment tasks, we found that, when patients made correct responses, their listening times (corrected for spoken word duration and word frequency) showed the normal effects of syntactic structure, but that self-paced listening times were abnormal at critical points in complex sentences when patients made errors (Caplan et al., in press a). This suggests that on-line processing fails intermittently, resulting in end-of-sentence errors.

The difference between the results of the Dickey et al. (2005) study and ours needs to be explained. Dickey interprets every look to the correct object as an indication that a subject has understood the sentence correctly, which may not be the case. Looks may reflect the current hypothesis formed by the listener, which may be held with variable degrees of certainty; aphasics may be less certain about these structures than normals. At the same time, taking self-paced listening times as reflecting only initial syntactic processing and comprehension may also not be correct. Self-paced listening times were longer for patients in our study for words in the sentence–picture matching than in the grammaticality judgment task, suggesting that these responses reflect the time it takes to begin to accomplish the task, while patients are self-pacing themselves through a sentence word-by-word.

All told, it is becoming apparent that it is possible to measure aphasic patients' performance on-line, while they are in the process of constructing syntactic form, using it to assign sentence meaning, and accomplishing a task, and that doing so may be a very profitable way to determine what is going wrong with the comprehension process. Advocates of various versions of the trace deletion hypothesis have argued that the CMLP results support their position, but, as we have seen, the results are far from definitive. Results of visual world tracking and self-paced listening suggest a remarkable degree of preservation of on-line processing in aphasia, which would imply that errors arise in the course of deciding which of several meanings is the correct one for the sentence or in mapping the correct meaning onto a task. As we said above, if this were the case, it would have both clinical and theoretic implications, and it is a high priority to pursue such studies.

## 7. APHASIC SYNDROMES AND NEUROLOGICAL CONSIDERATIONS

We close this chapter with a brief discussion of two topics that arise in connection with the study of aphasic disturbances of syntactic processing. The first is the question of whether particular aphasic syndromes are associated with particular types of disorders in this domain of function. The second is whether lesions in particular areas of the brain are associated with particular disorders. To anticipate, the answer to both questions is: "no."

If we go back to the first study of aphasic disturbances of syntactic comprehension, Caramazza and Zurif (1976), these authors studied two groups of patients: Broca's aphasics and mixed anterior aphasics. Both showed syntactic comprehension disorders. Caramazza and Zurif, however, focused on Broca's aphasics, because these patients also had agrammatic speech. The co-occurrence of these two problems seemed like it was more than a co-incidence, and, in the 1976 paper and other publications, the authors suggested that these patients had lost the ability to use syntax in both speech production and speech comprehension (Berndt & Caramazza, 1980; Zurif, 1982).

History repeats itself. Almost 100 years earlier, Wernicke (1874) had made exactly the same theoretical move about single word comprehension. Confronted with a patient who made phonemic errors in speech and did not understand spoken language, he argued that

the two performances arose because of a single impairment – a loss of the long-term store of the sounds of words, which he argued would produce exactly these two deficits. However, as Nobert Hornstein remarked, in neuropsychology, “Association proposes; Double dissociation disposes.” An analysis of two abnormal performances being due to a deficit in a single underlying function has trouble withstanding the observation that each of the abnormal performances occurs in isolation. Unfortunately for both Wernicke’s theory and the theory that agrammatic speech and “asyntactic” comprehension have their origins in a single deficit of syntactic processing, disturbances of each function have been observed in isolation. Some agrammatic patients have normal comprehension (to the extent they have been tested; Berndt, Mitchum, & Haendiges, 1996; Caramazza, Capitani, Rey, & Berndt, 2001); some patients without Broca’s aphasia or agrammatism have disorders of comprehension that are indistinguishable from those seen in these syndromes (Balogh & Grodzinsky, 2000). As far as is known, the same or at least similar disorders of comprehension occur in all aphasic syndromes; Wernicke’s and Conduction Aphasics make errors on picture-matching and act-out tasks that are similar to those made by Broca’s aphasics (Berndt et al., 1996; Caplan et al., 1985, 1996, 2005a).

One reason researchers have been interested in aphasic syndromes is that they have some localizing value. Broca’s aphasics and agrammatic aphasics tend to have lesions in Broca’s area. It was a major revelation in 1976 that patients with presumably anterior perisylvian lesions had any comprehension disturbances at all, as the main models of language and the brain at that time posited that comprehension was located behind the central sulcus, in association cortex adjacent to the primary cortical receptive areas of audition (Geschwind, 1965). However, it quickly became well known that the lesions in Broca’s aphasia were usually extensive and affected posterior as well as anterior perisylvian association cortex (Mohr et al., 1978). Which parts of these lesions were responsible for the syntactically based comprehension deficits in these patients – and in patients with other syndromes with similar deficits – needed to be studied further.

There are very few studies of this topic that actually analyze brains or neuroradiological images of brains. To our knowledge, there are only five studies in the literature in which radiological images have been analyzed and related to sentence comprehension in aphasics, and all five have significant limitations.

Two of these studies (Karbe et al., 1989; Kempler, Curtiss, Metter, Jackson, & Hanson, 1991), which measured cerebral hypometabolism with FDG PET, used tests of general sentence comprehension not well suited to characterize deficits in a linguistically or psycholinguistically specific way (the Western Aphasia Battery), or that completely confound syntactic processing with short-term memory requirements (the Token Test). Dronkers, Wilkin, Van Valin, Redfern, and Jaeger (2004) analyzed CT and MR scans, but their study also suffers from the ambiguity of the behavioral measurements. Dronkers et al. tested patients on the CYCLE, a sentence-picture-matching test in which two-thirds of the subtests have both lexical and syntactic foils, and did not separate out lexical from syntactic errors in the patients’ data, thus making it unclear whether abnormal performances reflected failures to understand words or sentences.

Dronkers et al. (2004) used a new and potentially promising approach to relating lesions to deficits – voxel-based lesion-symptom mapping – but their application of this approach was problematic. The approach requires that each voxel in each patient’s scan be classified as lesioned or unlesioned, but these judgments were made subjectively and inter-observer reliability of these types of judgments is known to be low (Naeser & Hayward, 1978; Naeser, Hayward, Laughlin, & Zatz, 1979; Hayward, Naeser, & Zatz, 1977). Voxels with lesions in 7 or fewer cases were eliminated, which may have eliminated many voxels in areas frequently affected by middle cerebral artery strokes (Hillis et al., 2004). In some analyses, lesioned voxels were grouped into regions of interest (ROIs), but how the boundaries of these regions were determined was not stated. Analyses were reported contrasting performance of patients with lesions in “the bulk” of these (undefined) ROIs with those with smaller lesions in these ROIs, but it is not clear what constituted “the bulk” of an ROI or whether this measure was the same in different ROIs. Dronkers et al. reported areas in which lesions associated with an (ambiguous) pattern of performance overlapped, but the “overlap” analysis approach does not consider the possible effect of size of lesion or the possibility that multiple regions must be lesioned to produce a pattern of performance. Finally, Dronkers et al. considered a lesion in a voxel to have resulted in abnormal performance on a task if patients with lesions in the voxel performed significantly less well than those without such lesions; however, when compared to the normal control subjects, all the lesions were associated with abnormal performance on all tests.

Tramo, Baynes, and Volpe (1988) reported quantitative lesion analyses and presented reversible sentences in a sentence–picture matching task, but studied only one contrast (active and passive sentences) and only reported three cases.

Caplan et al. (1996) studied the widest range of sentence types (25 sentence types testing many aspects of syntactic processing were examined in an enactment task) and analyzed CT scans. However, only 18 patients were studied, lesions were identified subjectively, and scans were normalized along a single linear dimension in one horizontal plane only, likely leading to significant inaccuracies in the estimates of percents of regions of interest that were lesioned.

We studied 32 of the 42 patients described above using both MR and PET scanning, and examined the relation of lesion site and size to performance using regression analyses (Caplan et al., in press b). Two main results were obtained.

First, percent lesion volume and mean PET counts/voxel in large regions accounted for a significant amount of variance in the performance measures. For instance, in stepwise regressions, percent MR lesion in the left hemisphere and in either the left hemisphere cortex or subcortical region accounted for a significant amount of variance in first factor scores for all tasks combined and for each task individually as well as for scores reflecting the effect of syntactic complexity (“combined syntactic complexity scores”) in sentence–picture matching. This is consistent with distributed models of functional neuroanatomy (Mesulam, 1990, 1998). It is also consistent with models that

maintain that the area that supports these processes differs in different individuals (“variable localization:” Caplan et al., 1985, 1996). It is consistent with invariant localization only if the predictive value of lesions in these large areas on any performance measure is due to the predictive value of lesions in a single smaller part of that region on that measure.

The second major finding is that this last statement was not the case. Percent lesion volume and mean PET counts/voxel in *several* small regions accounted for a significant amount of variance in performance measures. For instance, again considering the stepwise regressions, percent MR lesion in the inferior parietal lobe, and in both the anterior inferior temporal lobe and superior parietal lobe, and PET counts/voxel in Broca’s area accounted for a significant amount of variance in first factor scores for all tasks combined and for object manipulation; percent MR lesion in Wernicke’s area, and in the anterior inferior temporal lobe, and PET counts/voxel in the inferior parietal lobe accounted for a significant amount of variance in first factor scores for sentence–picture matching; percent MR lesion in Wernicke’s area and the inferior parietal lobe, and in the anterior inferior temporal lobe, and PET counts/voxel in the inferior parietal lobe, accounted for a significant amount of variance in combined syntactic complexity scores for object manipulation; and PET counts/voxel in the insula accounted for a significant amount of variance in combined syntactic complexity scores for sentence–picture matching. Since the regions that accounted for variance in these performance measures are discontinuous, not connected by major fiber tracts, and microscopically (cytoarchitecturally) different, this finding indicates that these performances are linearly related to the size of lesions in a variety of areas, not a single area. This is consistent with either a distributed multifocal or a variable localizationist model of functional neuroanatomy, but not with invariant localization. We undertook two further analyses to discriminate between these possibilities.

First, we examined the range of lesion sizes in regions in which the functions that underlie performance on the performance measures might be distributed in patients with equivalent levels of performance on five measures (first factor scores on each task with whole sentence presentation and combined syntactic complexity scores in object manipulation and sentence–picture matching). We examined lesion size in four regions in which functions might be distributed: the entire left hemisphere, the left hemisphere cortex, the perisylvian association cortex, and the combination of the perisylvian association cortex, the inferior anterior temporal lobe and the superior parietal lobe. If the functions that underlie performance on any of these measures are distributed throughout any of these regions, the ratio of the range of lesion size in that area in the selected patients to the range of lesion size in that area in all patients should be about the same as that of ratio of the range of performance of the selected patients to the range of performance of all patients. We examined patients whose performance on these measures fell within a very small range near the average of performance ( $\pm 0.25$  standard deviations of the mean performance for all patients scanned). Lesion size in each of these regions in these patients covered a considerable range of total lesion size (in some cases, the entire range of lesion size).

Second, we did the converse analysis. We examined the range of performance in patients with equivalent-sized lesions in these regions (within  $\pm 0.25$  standard deviations of the mean lesion size in each region for all patients scanned). Again, performances in patients with a small range of lesion size in each of these regions covered a considerable range of total performance (in some cases almost the entire range, and in one case the entire range, of performance).

These analyses are inconsistent with distributed models, or, at the least, show that the law of mass action does not apply in a linear fashion in such models. They are consistent with the view that the functions that underlie syntactically based sentence comprehension are localized in different regions in different individuals.

## 8. SUMMARY

Syntactically based sentence comprehension is an odd function. It is probably not needed for most face-to-face oral communication, which can probably be sustained quite well about matters pertaining to everyday living by the lexico-pragmatic route to meaning, the use of heuristics, and gestures and other non-linguistic aspects of communication. However, it is of importance for comprehension of more abstract topics, as well as for a considerable amount of reading comprehension, and it is of great theoretical interest because of Chomsky's claims about the specificity of syntax to humans and to language. Disorders of this function have been studied for about 30 years. They were originally taken as showing patterns quite consistent with aspects of Chomsky's models of syntax, and as pointing to one small area of the brain as the location where these abstract operations took place. As things progressed, this esthetically gratifying picture began to show cracks, and what is presently on display is a much more complex view of these disorders and the neural basis for syntactic operations. Reductions in general abilities to process sentences in the service of specific tasks, rather than isolated disturbances of particular syntactic operations, appear to be the most common problems affecting patients, and the lesions that produce these disturbances arise in many brain regions. It is customary to conclude chapters of this sort with the statement that "more research is needed." If we do want to understand these disorders, this remark applies in spades.

## ACKNOWLEDGMENTS

This research was supported by a grant from NIDCD (DC 00942).

## REFERENCES

- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Balogh, J., & Grodzinsky, Y. (2000). Levels of linguistic representation in Broca's aphasia: Implicitness and referentiality of arguments. In: R. Bastiaanse, & Y. Grodzinsky, (Eds), *Grammatical disorders in aphasia: a neurolinguistic perspective*. London: Whurr Publishers.

- Berndt, R., Mitchum, C., & Haendiges, A. (1996). Comprehension of reversible sentences in "agrammatism": A meta-analysis. *Cognition*, 58, 289–308.
- Berndt, R. S., & Caramazza, A. (1980) A redefinition of the syndrome of Broca's aphasia: Implications for a neuropsychological model of language. *Applied Psycholinguistics*, 1(3), 225–278.
- Bever, T. G. (1970). The influence of speech performance on linguistic structure. In: W. Levelt, & G. Flores d'Arcais (Eds), *Advances in psycholinguistics* (pp. 21–50). North-Holland: Amsterdam.
- Blumstein, S., Byma, G. et al. (1998). On-line processing of filler-gap constructions in aphasia. *Brain and Language*, 61(2), 149–169.
- Blumstein, S. E., Milberg, W., & Shrier, R. (1982). Semantic processing in aphasia: Evidence from an auditory lexical decision task. *Brain and Language*, 17, 301–315.
- Caplan, D. (1985). Syntactic and semantic structures in agrammatism. In: M. L. Kean. (Ed.), *Agrammatism* (pp.125–152). New York: Academic Press.
- Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, 21, 117–175.
- Caplan, D., DeDe, G., & Michand, L. (in press), Task-independent and task-specific syntactic deficits in aphasic comprehension. *Aphasiology*.
- Caplan, D., & Hildebrandt, N. (1988). *Disorders of Syntactic Comprehension*. Cambridge, MA: MIT Press (Bradford Books).
- Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain*, 119, 933–949.
- Caplan, D., Waters, G. (1990). Short-term memory and language comprehension: A critical review of the neuropsychological literature. In: T. Shallice, & G. Vallar (Eds): *The neuropsychology of short-term memory* (pp. 337–389). Cambridge, UK: Cambridge University Press.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 114–126.
- Caplan, D., & Waters, G. S. (2002). Working memory and connectionist models of parsing: A response to MacDonald and Christiansen. *Psychological Review*, 109, 66–74.
- Caplan, D., & Waters, G. S. (2003). On-line syntactic processing in aphasia: Studies with auditory moving windows presentation. *Brain and Language*, 84(2), 222–249.
- Caplan, D., Waters, G., & Hildebrandt, N. (1997). Syntactic determinants of sentence comprehension in aphasic patients in sentence–picture matching and enactment tasks. *Journal of Speech and Hearing Research*, 40, 542–555.
- Caplan, D., Waters, G., Kennedy, D., Alpert, N. Makris, N., DeDe, G., Michaud, J., & Reddy, A. (in press b). A study of syntactic processing in aphasia II: Neurological aspects. *Brain and Language*.

- Caplan, D., Waters, G. S., DeDe, G., Michaud, J., & Reddy, A. (in press a). A study of syntactic processing in Aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*.
- Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language*, *76*, 158–184.
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain & Language*, *3*, 572–582.
- Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cupples, L., & Inglis, A. L. (1993). When task demands induce “asyntactic” comprehension: A study of sentence interpretation in aphasia. *Cognitive Neuropsychology*, *10*, 201–234.
- DeDe, G., Caplan, D., & Waters, G. (2006). Factor analysis of Syntactic deficits in a phasic comprehension. *Aphasiology*, *20*, 123–135.
- DeDe, G., Caplan, D., Waters, G., Michaud, J., & Reddy, A. (2005b). Factors underlying sentence comprehension in older adults and patients with aphasia. *Clinical Aphasiology Conference*, Sanibel, FL.
- Dick, F., Bates, E., Wulfeck, B., Utman, J., Dronkers, N., & Gernsbacher, M. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, *108*, 759–788.
- Dickey, M., & Thompson, C. (2004) The resolution and recovery of filler-gap dependencies in aphasia: Evidence from on-line anomaly detection, *Brain and Language*, *88*(1), 108–127.
- Dickey, M., Thompson, C., & Choy, J. (2005). The on-line comprehension of wh-movement structures in agrammatic Broca's aphasia: Evidence from eyetracking. *Clinical Aphasiology Conference*, Sanibel, FL.
- Dronkers, N., Wilkin, D., Van Valin, R., Redfern, B., & Jaeger, J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*, 145–177.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–224.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.
- Frazier, L., & Freiderici, A. (1991). On deriving the properties of agrammatic comprehension. *Brain and Language*, *40*, 51–66.
- Geschwind, N. (1965). Disconnection syndromes in animals and man. *Brain*, *88*, 237–294, 585–644.
- Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language*, *27*, 135–159.



- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23, 47–117.
- Haarmann, H. J., & Kolk, H. H. (1991). Syntactic priming in Broca's aphasics: Evidence for slow activation. *Aphasiology*, 5, 247–263.
- Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language*, 59, 76–120.
- Hayward, R., Naeser, M. A., Zatz, L. M. (1977). Cranial computer tomography in aphasia. *Radiology*, 123, 653–660.
- Hillis A., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain*, 127, 1479–1487.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149.
- Karbe, H., Herholz, K., Szefiels, B., Pawlik, G., Wienhard, K., & Heiss, W. (1989). Regional metabolic correlates of Token test results in cortical and subcortical left hemispheric infarction. *Neurology*, 39, 1083–1088.
- Kempler, D., Curtiss, S., Metter, E., Jackson, C., & Hanson, W. (1991). Grammatical comprehension, aphasic syndromes and neuroimaging. *Journal of Neurolinguistics*, 6, 301–318.
- Love, T., Swinney, D., & Zurif, E. (2001). Aphasia and the time-course of processing long distance dependencies. *Brain and Language*, 79, 169–170.
- Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13, 361–392.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A reply to Just & Carpenter and Waters & Caplan. *Psychological Review*, 109, 35–54.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Marcus, M. (1980). *A theory of syntactic recognition for natural language*, Cambridge, MA: MIT Press.
- Martin, R. C., & He, T. (2004). Semantic short-term memory deficit and language processing: A replication. *Brain and Language*, 89, 76–82.
- McKoon, G., & Ratcliff, R. (1994). Sentential context and on-line lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 1239–1243.
- McKoon, G. & Ratcliff, G. (2005). Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review*, 110, 490–525.

- McKoon, G., Ratcliff, G., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1219–1228.
- Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104, 3–65.
- Mesulam, M. -M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5), 597–613.
- Mesulam, M. -M. (1998). From sensation to cognition. *Brain*, 121, 1013–1052.
- Milberg, W., & Blumstein, S. E. (1981). Lexical decision and aphasia: Evidence for semantic processing. *Brain and Language*, 14, 371–385.
- Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Mass: MIT Press.
- Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: Pathologic and clinical. *Neurology*, 28, 311–324.
- Naeser, M. A., & Hayward, R. W. (1978). Lesion localization in aphasia with cranial computed tomography and the Boston diagnostic aphasia examination, *Neurology*, 28, 545–551.
- Naeser, M. A., Hayward, R. W., Laughlin, S., & Zatz, L. M. (1979). Quantitative CT scans studies in aphasia. I. Infarct size and CT numbers. *Brain & Language*, 12, 140–164.
- Nicol, J., Fodor, J., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1229–1238.
- Prather, P., Zurif, E. B., Stem, C., & Rosen, J. (1992). Slowed lexical access in nonfluent aphasia: A case study. *Brain and Language*, 43, 336–348.
- Prather, P., Zurif, E. B., Love, T. Y., & Brownell, H. (1997). Speed of lexical activation in nonfluent broca's aphasia and fluent wernicke's Aphasia. *Brain and Language*, 59, 391–411.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning internal representations by error propagation. In: D. E. Rumelhart, & J. L. McClelland, (eds), *Parallel distributed processing, Volume I: Foundations* (pp. 318–364). Cambridge, MA: MIT Press.
- Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, 2, 179–183.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568.
- Swinney, D., & Zurif, E. (1995). Syntactic processing in aphasia. *Brain and Language*, 50, 225–239.

Swinney, D., Zurif, E., Prather, P., & Love, T. (1996). Neurological distribution of processing resources underlying language comprehension. *Journal of Cognitive Neuroscience*, 8, 174–184.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical theories of sentence processing. *Cognitive Science*, 23, 491–515.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.

Tramo, M. J., Baynes, K., & Volpe, B. T. (1988). Impaired syntactic comprehension and production in Broca's aphasia: CT lesion localization and recovery patterns. *Neurology*, 38, 95–98.

Thompson, C., Dickey, M., & Choy, J. (2004). Complexity in the comprehension of wh-movement structures in agrammatic Broca's aphasia: Evidence from eyetracking. *Brain and Language*, 91, 124–125.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49A, (1), 51–79.

Wernicke, C. (1874). *Der Aphasische Symptomenkomplex*. Breslau: Cohn & Weigart. Reprinted in translation in *Boston Studies in Philosophy of Science*, 4, 34–97.

Zurif, E. B. (1982). The use of data from aphasia in constructing a performance model of language. In: M. A. Arbib, D. Caplan, & J. C. Marshall (Eds), *Neural models of language processes* (pp. 203–207). New York: Academic Press.

Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's Aphasia. *Brain and Language*, 45, 448–464.

## Chapter 25

### Language Processing in Bilingual Speakers

Ana I. Schwartz and Judith F. Kroll

We couldn't agree where we wanted to go on vacation, *y tuvimos una larga discusion*  
[and we had a long argument]

In the border community of El Paso, Texas, it is not at all uncommon to hear code-switched phrases like the one above. English and Spanish conversations are heard throughout all contexts of the community, whether it be on a university campus, a small café, bus, or in a formal business meeting. What an interesting feat it is then for the typical, bilingual citizen of El Paso to comprehend the mixture of language that she may hear or read. How is it that bilinguals comprehend input from their two languages? At what point in comprehension is the language of the utterance identified? At what point, if at all, does a bilingual select one language over the other?

These questions seem particularly salient when one considers that most of the world's population is proficient in more than one language (Bhatia & Ritchie, 2004). However, it is important to point out that research on bilingualism is essential toward developing theories of language processing and cognition that extend beyond interests in bilingualism. For example, take the word “*discusión*” in the phrase quoted above. Although this is a Spanish word, it clearly looks like the word “discussion” in English, and this similarity makes the language membership of the word quite ambiguous. Furthermore, despite the high degree of superficial, form similarity between these two words, their meanings are actually quite distinct. In Spanish, “*discusión*” refers to a disagreement; whereas the meaning of the English word “*discussion*” does not include the same combative overtones. How do bilinguals disambiguate such words and apply the appropriate intended meaning? The issue of lexical disambiguation has been a focus in the development of psycholinguistic theories of language processing and has spurred considerable debate regarding the role of context in lexical access and the degree to which processes of language comprehension are modular vs. interactive. The study of bilingual language comprehension allows researchers to address these issues in ways that cannot be achieved through monolingual paradigms.

In this chapter, we review the recent literature on bilingual language comprehension and production. We will focus in particular on issues related to cross-language interactions that take place between a native language and a non-native language. Previous research has examined both the interactions that occur as a new language is being acquired as well as the interactions that take place during online language comprehension and production processes. In terms of the interactions that take place during acquisition, research has demonstrated that linguistic characteristics of the native language (L1), such as how phonologically or syntactically similar it is to the second language (L2) has an effect on its acquisition (e.g., Bosch, Costa, & Sebastian-Galles, 2000; Fernández, 1998; MacWhinney, 1997). Similarly, there is evidence that the process of acquiring an L2 also has an impact on the L1. For example, studies of syntactic processing suggest that parsing preferences that develop in L2 acquisition can transfer and modify the parsing strategies used during L1 processing (Dussias, 2003). With respect to the interactions that take place during online processing, research has provided compelling evidence that bilinguals are unable to selectively activate one of their languages during either comprehension or production. Bilingual language comprehension appears to involve the parallel activation of lexical information during both visual word recognition (Dijkstra, De Bruijn, Schriefers, & Brinke, 2000; Dijkstra, Grainger, & Van Heuven, 1999; Jared & Kroll, 2001; Jared & Szucs, 2002; Van Heuven, Dijkstra, & Grainger, 1998; Von Studnitz & Green, 2002; and see Dijkstra, 2005 for a recent review) and auditory word recognition (e.g., Spivey & Marian, 1999). Likewise, the spoken production of words in even one of the bilingual's two languages appears to activate lexical candidates in both languages (e.g., Colomé, 2001; Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998).

Only very recently have researchers begun to examine how language non-selectivity is modified, if at all, by language or sentence context (e.g., De Bruijn, Dijkstra, Chwilla, & Schriefers, 2001; Schwartz & Kroll, 2006; Ju & Luce, 2004; Van Hell & De Groot, 1998). To the extent that the target language cannot be immediately selected, there are also critical consequences for the cognitive control that bilinguals must develop to negotiate the potential competition across the two languages (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Green, 1998). In this chapter, we will review these studies and discuss the implications for current models of the bilingual lexicon in comprehension and production.

In the review that follows, we take an inclusive approach to consider bilinguals to be anyone who actively uses two languages to some degree of proficiency. Bilinguals are rarely equally proficient or balanced in their use of the two languages, rendering one of the languages the more dominant language. Typically, the more dominant language will be the first or native language, but for bilinguals who have lived in their L2 environment for many years, the L2 may be functionally more dominant, at least for certain language-processing tasks. Among the most commonly studied bilinguals (e.g., the Dutch–English bilinguals in The Netherlands or the Catalan–Spanish bilinguals in Barcelona), many individuals are actually trilinguals. These contexts of language acquisition and use will clearly have implications for comprehension although some of the consequences, particularly in studies of lexical access, have only recently begun to be explored systematically.

We discuss them in the cases in which there are documented consequences for processing and where there is enough evidence to assess their contribution.

We begin the chapter with a review of models of the bilingual lexicon. Although the issues we consider in this chapter extend beyond the lexical level, early models of cross-language representation and processing focused extensively on the way in which words and concepts are interconnected in the bilingual's two languages. These models provide a starting point for understanding how a learner's cognitive system adapts to the presence of a second language. We then provide an overview of the recent research that has examined the cross-language interactions that take place during comprehension and production at the lexical and sub-lexical levels. These studies consider how single words and sounds from a bilingual's two languages are identified and spoken. We then review the evidence for cross-language interactions at the sentence/syntactic levels. Finally, we examine a set of issues common to both the lexical and sentential level of processing that include the cognitive factors that modulate L2 processing, the consequences of cross-language similarity, the directionality of cross-language influences, and the more general implications of research on bilingualism for theories of word recognition, lexical access, and sentence processing. Although our review focuses primarily on behavioral evidence from studies of the performance of second language learners and relatively proficient bilinguals, we also consider the emerging literature on the neural basis of bilingualism.

## 1. MODELS OF THE BILINGUAL LEXICON

Early accounts of the bilingual's cognitive system examined the issue of how an adult who already has a fully formed lexicon and conceptual system for their native language represents newly acquired information in a second language. Potter, So, Von Eckardt, and Feldman (1984) contrasted two alternative models for how a learner might integrate new L2 knowledge into their existing L1 language system. The models are shown in Figure 1. According to the Word Association model, new L2 words are represented by means of associations to their translation equivalents. Thus, to learn that the word *kikker* in Dutch, which means *frog* in English, the learner would form an association directly between *kikker* and *frog*. On this view, L2 words access meaning indirectly via the L1. In contrast, the Concept Mediation model assumes that L2 words have direct access to their respective meanings. Potter et al. tested the two models in a series of experiments in which bilinguals named pictures and translated words. They reasoned that if words in the L2 were lexically mediated via the L1, then translation into the L2 should be performed more quickly than picture naming because the translation would be directly available without conceptual processing. Picture naming would necessarily be slower than translation because to name a picture in L2 would require first accessing its meaning and name in L1, and only then could the L2 name be retrieved by its association with the L1 translation. However, if words in the L2 access their respective meanings directly, then all other things being equal, picture naming in L2 and translation into the L2 should take about the same amount of time. Potter et al. found support for the Concept Mediation predictions.

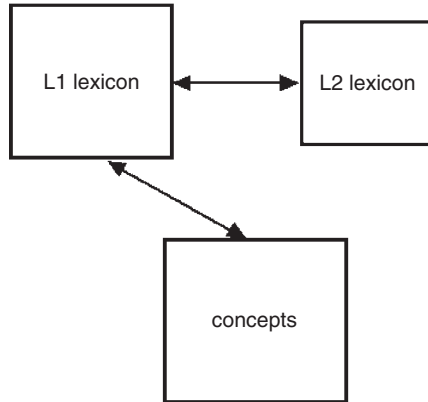


Figure 1. The Word Association and Concept Mediation models (adapted from Potter et al., 1984).

In two experiments, one with highly proficient Chinese–English bilinguals and the other with French–English second language learners, the time to name pictures and translate words was approximately the same. Contrary to the predictions of the Word Association model, the time to name pictures was, if anything, faster than the time to translate words. The results are counterintuitive because although the learners were slower in L2 than the proficient bilinguals, the pattern was the same, suggesting that even at early stages of L2 learning, individuals are able to access concepts directly.

The models shown in Figure 1 assume independent representations at the lexical level but a shared representation at the conceptual level. A question that we will return to is how the picture of cross-language activity suggested by the current research can be accommodated within this hierarchical framework. If even learners can access meaning directly for words in the L2 as the Potter et al. (1984) results suggest, then lexical-level interactions across languages might be expected to be quite limited. However, the immediate problem raised by the Potter et al. study was to understand why striking differences in L2 skill did not seem to influence the form of the connections between words and concepts in the bilingual’s mind. Two subsequent studies (Chen & Leung, 1989; Kroll & Curley, 1988) failed to replicate the Potter et al. results for the less-proficient L2 learners. In each of these studies, the performance of highly proficient bilinguals replicated the Concept Mediation model’s pattern, with similar latencies for naming pictures in the L2 and translating words into the L1. However, contrary to Potter et al., the performance of L2 learners followed the pattern predicted by the Word Association model, with translation into the L2 significantly faster than picture naming in the L2. The apparent discrepancy in the findings for L2 learners may be attributed to their level of proficiency. The learners in the Potter et al. study were high school students about to travel to France for a summer immersion experience. Although they were far from proficient in French, they were presumably highly motivated learners who were likely to have been beyond the very earliest stages of acquisition.

The results of Chen and Leung (1989) and Kroll and Curley (1988) suggested that learners initially associate new words in the L2 to their translation equivalents in the L1 and only later become able to access concepts directly. To account for the change in the connections between words and concepts as L2 skill develops, Kroll and Stewart (1994) proposed the revised hierarchical model (RHM). The model, shown in Figure 2, integrates the Word Association and Concept Mediation models but, unlike the earlier models, also makes assumptions about the strength of connections. At the lexical level, the RHM assumes that the connection from L2 to L1 is stronger than the connection from L1 to L2. Because learners presumably access L1 translations during early stages of acquisition for the purpose of retrieving the meaning of new L2 words, the requirement to transfer information from L1 to L2 will create a strong lexical-level connection from L2 to L1. With respect to the connections from words to concepts, the RHM proposes that the L1 is privileged with respect to accessing meaning and thus L1 connections to concepts will be stronger than those for L2 and only with increasing L2 proficiency will the L2 links to concepts begin to resemble those for L1.

The RHM can account for an asymmetry observed in the previous studies such that translation in the forward direction, from L1 to L2, is typically slower and more error-prone than translation in the backward direction, from L2 to L1. According to the model, L2 to L1 translation can be performed by retrieving lexical-level associations that provide a direct route to the translation, whereas translation from L1 to L2 is more likely to engage concept processing and subsequent lexicalization by virtue of the strong connections between L1 words and their respective meanings. The translation asymmetry is consistent with this account but can also be attributed to difficulties in accessing the phonology of the weaker L2. To provide a more critical test of the RHM, Kroll and Stewart (1994) examined the translation performance of relatively proficient Dutch–English bilinguals in contexts in which the words to be translated appeared in semantically categorized lists or in lists that were semantically mixed. If only translation from L1 to L2 engages meaning, then only translation in that direction should be influenced by the semantic organization of the list context. The results confirmed this prediction. Translation from L1 to L2 was slower in the semantically categorized lists

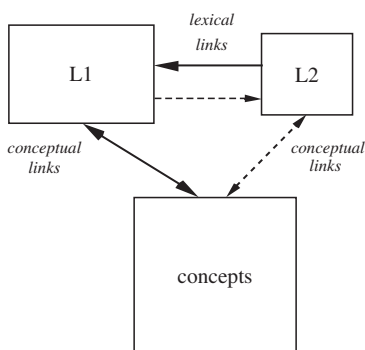


Figure 2. The Revised Hierarchical Model (adapted from Kroll & Stewart, 1994).



than in the mixed lists but translation from L2 to L1 was unaffected by this manipulation. The fact that semantically categorized lists induced interference was attributed to likely competition during L2 production (see Kroll, Bobb, & Wodniecka, 2006 for a recent discussion of this issue).

If conceptual processing in the L2 is a function of proficiency, then why did the Dutch–English bilinguals in the Kroll and Stewart (1994) study, who are relatively skilled speakers of English, show an asymmetry for the two directions of translation? One possibility is that the words to be translated included items that were relatively low frequency, making it more likely that the L2 would revert to reliance on the L1 translation. A number of recent studies have suggested that weighting of interlanguage connections is affected not only by the proficiency of the bilingual speaker but also by the nature of the words and concepts that are translated (e.g., Francis, Tokowicz, & Kroll, 2003). When a small set of items are well learned or highly practiced, even relative novices may appear to conceptually process words in the L2 (e.g., Altarriba & Mathis, 1997). However, under other circumstances, the development of L2 proficiency appears to shift from reliance on lexical cross-language connections to reliance on conceptual interconnections with increasing skill (e.g., Kroll, Michael, Tokowicz, & Dufour, 2002; Talamas, Kroll, & Dufour, 1999). Although there is controversy with respect to the issue of whether highly proficient bilinguals ever revert to lexical mediation during translation (e.g., Duyck & Brysbaert, 2004; La Heij, Hooglander, Kerling, & Van der Velden, 1996), the evidence on L2 learners suggests that there is an early stage of acquisition in which L1 translations are highly active (Sunderman & Kroll, in press).

The RHM and the earlier Concept Mediation model share an implicit assumption that once individuals achieve sufficient expertise in the L2, the L2 can function autonomously of the L1 (but see Kroll & De Groot, 1997, for a distributed model at both the lexical and conceptual levels). As we will see in the remainder of this chapter, it is this assumption that has been challenged in the recent literature in which cross-language activity has been observed for even highly proficient bilinguals. While no one would dispute the fact that bilinguals gain automaticity in processing the L2 with increasing skill (e.g., Segalowitz & Hulstijn, 2005), recent studies suggest that proficiency does not imply an ability to use each of the languages autonomously, as if the bilingual were functionally monolingual. However, a recent study (Sunderman & Kroll, in press) demonstrated that at the lexical level, it is the activity of the translation equivalent that changes with increasing proficiency, but not the activity of lexical form relatives. As we will review below in the section on lexical-level cross-language interaction, when bilinguals read or hear words in one language, there is activation of related information in the other language. However, what is active for proficient bilinguals is the orthographic and phonological information present in words in the two languages, not the translation equivalent itself. Thus the word *man* might be briefly activated in English when an English–Spanish bilingual reads the word *mano*, meaning hand in Spanish. These words are semantically unrelated but share similar lexical form. Sunderman and Kroll showed this type of cross-language activity occurred regardless of L2 proficiency. In contrast, activation of information related to the translation equivalent (e.g., the word *hambre*, meaning hunger in Spanish, resembles

*hombre*, the translation of man) occurred only for learners at relatively early stages of L2 acquisition. The account provided by the RHM with respect to lexical mediation thus appears to characterize the performance of learners who still depend on the L1 to access meaning and more proficient bilinguals who may revert to this strategy when the words to be translated are relatively low frequency in the L2, as seen in Kroll and Stewart (1994).

A comprehensive bilingual model will require that distinctions be made between those aspects of cross-language representation and processing that change dynamically with changes in proficiency and language use, and those that reflect the way in which the linguistic structure of the bilingual's two languages imposes constraints that are relatively stable (see Kroll & Tokowicz, 2005 for a discussion of these issues). Furthermore, it will require a principled account of the relation between sub-lexical, lexical, semantic, and syntactic processes. In the sections that follow we report the recent evidence at each of these levels of language processing and further consider the manner in which context and skill modulate the observed cross-language interactions.

## **2. CROSS-LANGUAGE INTERACTIONS AT THE LEXICAL AND SUB-LEXICAL LEVELS**

### **2.1. The Perception of Speech in Two Languages**

A focus of contemporary psycholinguistic research on both monolinguals and bilinguals concerns the mechanisms that allow individuals to perceive and uniquely identify the sounds and words of their languages. Indeed, one of the first tasks faced by all language learners is the identification of his/her native language(s) and all of its distinct sounds. In the earliest days of life, newborns are able to discriminate between unfamiliar languages, particularly if these languages belong to different rhythmic classes (e.g., Mehler et al., 1988). What about infants who are raised in multilingual environments? Recent findings suggest that bilingual infants might have enhanced language-discrimination abilities. Infants raised in bilingual environments show an early ability to discriminate languages; even those that are from the same rhythmic class (Bosch & Sebastián-Gallés, 2001).

In addition to making cross-language distinctions, young language learners must learn to identify and discriminate sounds within their native language(s). Research on monolingual language acquisition has demonstrated that within the first year of life monolingual infants are sensitive to the phonotactic patterns of their native language. For example, as early as nine months of age they are able to discriminate phonological sequences (e.g., CV structure) that conform to the regularities of the language from patterns that do not (e.g., Jusczyk, Goodman, & Baumann, 1999; Sebastián-Gallés & Bosch, 2002). How do the early sensitivities of infants exposed to multiple languages compare to those of infants raised in monolingual contexts? Recent evidence suggests that infants who are exposed simultaneously to two languages develop phonotactic sensitivities at the same rate as monolinguals. More interestingly, the degree of this sensitivity increases as

a function of the relative exposure to a given language; thus even at this early stage of life infants' perceptual abilities can reflect a perceptual dominance for the language to which they are most often exposed (Sebastián-Gallés & Bosch, 2002). This perceptual dominance of one language over another has been shown to persist into adulthood. It has been found that even highly proficient bilinguals apply one set of segmentation rules exclusively from the language to which they were most often exposed when parsing speech (e.g., Cutler, Mehler, Norris, & Seguí, 1989).

During the early months of life, infants are also developing cognitive representations of the different sounds or phonemes that make up their native language (Kuhl, Williams, Lacerda, & Stevens, 1992; Marean, Werner, & Kuhl, 1992; Werker & Lalonde, 1988). The perceptual system is quite flexible, allowing very young infants to make non-native phonemic contrasts. Remarkably, within just 6–12 months of age, there is a significant re-organization in which perceptual processes become specifically tuned to the native language and infants' ability to discriminate non-native phonemic contrasts diminishes (Werker & Lalonde, 1988). With increasing age and linguistic exposure, this language-specific tuning acts as a filter through which a non-native language is perceived. As a consequence, two phonemes that are contrastive within a second or less familiar language may be perceived as the same phoneme if the L1 does not distinguish between these sounds. For example, adult native Japanese speakers often have great difficulty in distinguishing between the /r/ and /l/ English phonemes, which are not contrastive in Japanese and both phonemes tend to be perceived as /l/.

This L1 filtering effect has been observed even for highly proficient bilinguals who have had long lifetime experiences and seemingly equal command of both languages (e.g., Pallier, Colomé, & Sebastián-Gallés, 2001; Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés, Echeverría, & Bosch, 2005). For example, highly proficient Spanish–Catalan speakers performing an auditory lexical decision in their L2 showed repetition priming for Catalan words that differed only by a single vowel contrast that does not exist in Spanish. These priming effects suggested that the bilinguals were perceiving the minimal pairs as the same entity. A separate group of Catalan-dominant bilinguals did not show these priming effects. These results were particularly remarkable since the two bilingual groups were both early bilinguals who were living in the same city and had received the same bilingual education. Indeed, the Spanish-dominant group was distinguished based only on the fact that they had been raised in Spanish monolingual homes during their earliest childhood prior to entering school (Pallier et al., 2001).

However, it is important to note that the perceptual system does not ever completely lose its ability to perceive new phonemes and distinguish new contrasts. Furthermore, non-native speakers vary greatly in their ability to discriminate L2 contrasts (e.g., Strange, 1995). The relative ease with which non-native phonemes are perceived and contrasted depends critically on several factors, including the phonetic similarity of the native and non-native languages and the age at which the second language was acquired (e.g., Flege, 1995).

## 2.2. Recognizing Words in Two Languages When They Are Read or Spoken

The studies reviewed thus far highlight the ways in which the acquisition of a native language has a fundamental and long-lasting effect on how the sub-lexical and lexical units of a new language are perceived and acquired. There is also evidence for cross-language interactions that occur during the online comprehension of language, and there is now a large body of literature demonstrating that when recognizing words bilinguals activate lexical and sub-lexical units from both of their languages in parallel (e.g., Dijkstra et al., 1999; Dijkstra et al., 2000a; Jared & Kroll, 2001; Ju & Luce, 2004; Marian & Spivey, 2003; Van Heuven et al., 1998).

Many of the studies that have looked at these online interactions have focused on visual word recognition tasks such as lexical decision and naming. The general strategy has consisted of presenting bilinguals with words in one language that share some lexical property or properties with words from their other language. For example, in a lexical decision task bilinguals may be presented with non-words (blart) and two types of words: Words that share lexical form and meaning with the non-target language (e.g., the cognate *piano* in English and Spanish) or control words that do not share any lexical property with the non-target language (e.g., *pencil*). One can then compare the bilinguals' word recognition performance, in terms of latency and/or accuracy, for these critical words relative to control words that do not share lexical properties. If bilinguals are able to selectively activate a single language during word recognition, then whether the word shares lexical properties with the non-target language should be of no consequence. If, on the other hand, there is parallel activation of both languages then the processing time and/or accuracy of performance should differ from control words.

The evidence has consistently demonstrated that bilingual word recognition involves the parallel, non-language-selective activation of both languages. For example, the processing of cognates is consistently facilitated relative to non-cognate control words across a wide variety of tasks including translation (Kroll & Stewart, 1994); word association (Van Hell & De Groot, 1998) and lexical decision (Dijkstra, Van Jaarsveld, & Ten Brinke, 1998). In translation, it is perhaps not surprising to find cognate facilitation because both languages are required to be active. However, in within-language lexical decision, it should be possible to function selectively, yet even under these circumstances there are robust effects of the non-target language. Cognate facilitation is obtained when bilinguals perform visual lexical decision in their native language with no expectation that the L2 will be used, suggesting that these effects are not simply a reflection of the stronger L1 influencing the weaker L2 (e.g., Van Hell & Dijkstra, 2002). Furthermore, evidence of non-selectivity persists irrespective of the surrounding language context, task instructions, or participant expectations to process one or multiple languages (Dijkstra et al., 2000b; Dijkstra & Van Hell, 2003).

It is important to note that the nature of the effects of non-selective activation depends on the relative match of the lexical codes (orthographic, phonological, and semantic)

across languages. While cognate facilitation has been observed across a variety of different tasks, the effects of inter-lingual findings regarding interlingual homographs, words that share form but not semantics (e.g., *fin* in Spanish means “end”) have not been nearly as consistent. Some studies have demonstrated inhibitory effects associated with homograph status (Dijkstra et al., 1998; Jared & Szucs, 2002; Von Studnitz & Green, 2002), while others have failed to find any effects at all (Dijkstra et al., 1998; Gerard & Scarborough, 1989). Furthermore, the specific nature of homograph effects, whether they are inhibitory or facilitative in nature has varied as a consequence of differences in task demands, the salience of the non-target language, and the relative frequency of the homographs’ lexical representations across languages (Dijkstra et al., 2000a; Dijkstra et al., 2000b; Dijkstra et al., 1998).

In a recent study, Schwartz, Kroll, and Diaz (in press) found further evidence that bilingual lexical processing of cognates is influenced by the relative match in orthographic and phonological codes across languages. In that study, English–Spanish bilinguals named cognates (e.g., *piano*) and non-cognate control words (e.g., *lapiz/pencil*) in their L1 and L2. The cognates were classified according to the relative orthographic and phonological similarity across English and Spanish. To illustrate, the English–Spanish cognate *base* maps on to very distinct pronunciations ([bas] vs. [báse]) and was classified as +O–P, whereas *piano* is pronounced much more similarly ([piæ’nou] vs. [pi’a’no]) and was classified as +O+P. The authors predicted that in the presence of highly similar orthography (e.g., *pianolpiano*; *baselbase*), the activation of the cross-language phonological representations would be particularly strong. When these representations were more distinct (e.g., [bas] vs. [báse]) the resulting competition would inhibit performance. The findings supported this prediction, the +O+P cognates (e.g., *piano*) were named faster and more accurately than the +O–P cognates (e.g., *base*) in both the L1 and L2 of the participants. This suggests that not only is lexical access non-selective across bilinguals’ two languages, but that the subtle interactions between the activated codes determine the manner in which cross-language competition is manifest.

The studies reviewed above all focused on cross-language activation that occurs during the recognition of visually presented words. In these studies, effects of cross-language activation may have been particularly robust since the visual input (i.e., a string of letters) can be completely language neutral (e.g., *fin* in English and Spanish), at least in the case in which languages share the same script. It should be noted that cross-language activation has also been observed for words that are not identical. For example, the identification of cross-language neighbors, words that share all but one letter across languages (e.g., *cinel/dine* in Spanish and English), has been shown to be slower and more error-prone relative to control words in lexical decision (Van Heuven et al., 1998) and naming (Jared & Kroll, 2001). There is also facilitation for recognizing cognates even when the cognates are non-identical (Van Hell & Dijkstra, 2002). Furthermore, there is evidence that even when the languages do not share the same script (e.g., Hebrew and English or Chinese and English), cross-language priming effects can be observed (e.g., Gollan, Forster, & Frost, 1997; Jiang, 1999).

Parallel activation of information in each of the bilingual’s languages during visual word recognition has been modeled in a variant of McClelland and Rumelhart’s (1981) Interactive Activation model for monolingual word recognition called the Bilingual Interactive Activation model, or BIA (Dijkstra et al., 1998). The model is shown in Figure 3. Like the monolingual model, BIA assumes that there is parallel activation of

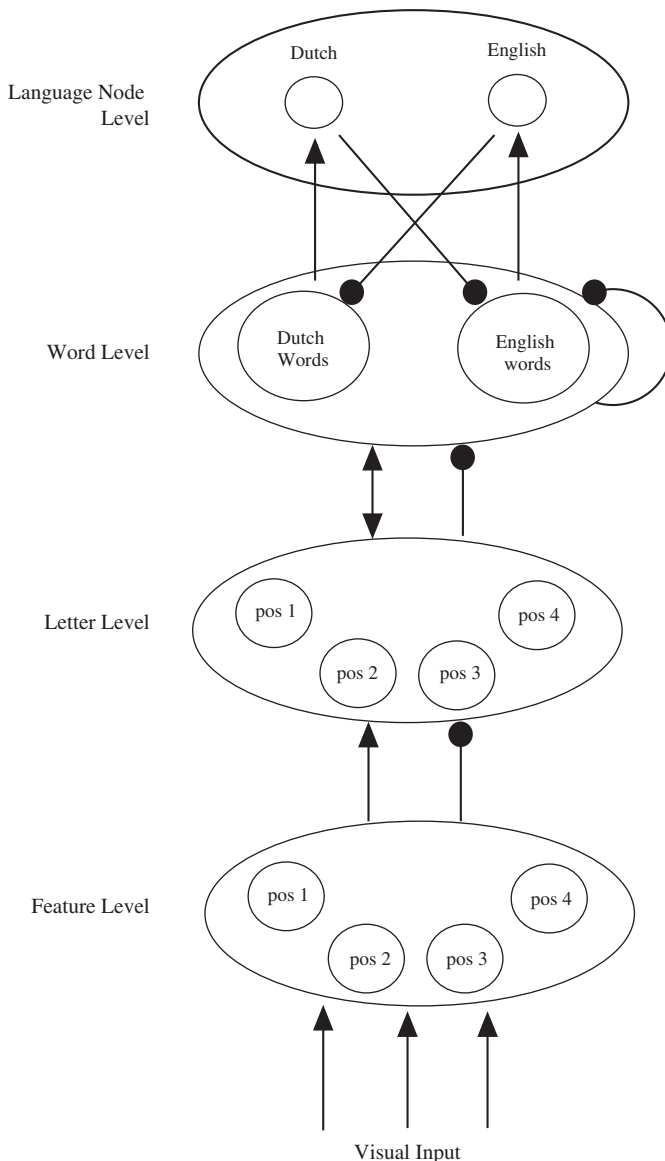


Figure 3. The Bilingual Interactive Activation Model (adapted from Dijkstra et al., 1998).

letter features, letters, and words, with information similar to the input string activated to some degree and producing competition across alternative candidates. However, unlike the monolingual model, BIA includes a layer of language nodes which serve to represent the top-down contextual biases and subsequently inhibit the bottom-up activation of the non-target language. In this scheme, the inhibitory effects occur relatively late in processing, once the initial components of word recognition are set in motion for all possible solutions in either of the bilingual's two languages. BIA has been implemented as a computer model and does an excellent job of simulating bilingual word recognition performance under conditions in which the words to be recognized differ in their within and across-language orthographic properties (Van Heuven et al., 1998). At the end of the chapter, we describe an updated version of the BIA model that includes phonology and semantics as well as orthography, and also makes some different assumptions about the way in which language selection is controlled.

The observation of parallel activity across the two languages during visual word recognition does not necessarily suggest that similar cross-language activation occurs during auditory processing of the speech signal. Since languages differ in their component sounds it can be argued that within the speech signal there are language-specific cues that are not as readily available within printed text. There have been a few studies that have examined cross-language interaction during speech processing. In a seminal study, Spivey and Marian (1999) asked Russian–English bilinguals to view an array of objects as they listened to instructions in either their L1 or L2 which indicated an object that they should select (e.g., “pick up the marker”). On critical trials the instructions indicated a target object whose phonological onset was the same as that of another object *in the non-target language* (e.g., “stamp” in Russian is “marka”). To test whether the non-target lexical representation of the object was activated, the authors monitored the bilinguals' eye-movements as they surveyed the array of objects and listened to the instructions. When the instructions indicated an object whose phonological onset was shared across languages, participants initially looked toward the object that shared this onset in the non-target language. This indicated that upon hearing the initial, shared phoneme, the bilinguals activated lexical candidates from both of their languages (and see Marian & Spivey, 2003).

Using a very different paradigm, Colomé (2001) found converging evidence that bilinguals activate phonemic representations from both languages in a non-selective manner. In that study highly proficient Spanish–Catalan bilinguals performed a phoneme monitoring task in their L2, in which they decided whether the name of a visually presented picture (e.g., a table) contained a target phoneme (e.g., /m/). On critical trials, the bilinguals had to reject phonemes that were not part of the Catalan name (e.g., /m/ is not present in the Catalan word *taula*) but *were* part of the contextually irrelevant Spanish translation of that object (e.g., *mesa*). The bilinguals took significantly longer to reject phonemes contained in the Spanish translation relative to phonemes that were not part of the picture's name in either language (e.g., /s/).

Subsequent studies have demonstrated that aspects of the linguistic input itself may make it possible to constrain the parallel activation of the non-target language when processing spoken language. Weber and Cutler (2004), tested Dutch–English bilinguals

with an eye-tracking paradigm very similar to Marian and Spivey (2003), and found significant cross-language effects from the L1 to the L2 (i.e., when bilinguals were processing the spoken targets in the non-native language) but not from the L2 to L1. Likewise, Ju and Luce (2004) replicated the basic pattern of cross-language phonological competitor effects, but then went on to demonstrate that cross-language competitor activation could be eliminated when the voice onset times (VOTs) of the initial phonemes were spoken like the L1. That is, L2 competitors were no longer activated when the target words were perceived to be native-like speech. These results contrast with results from experiments using written stimuli and illustrate the critical role that access codes play in the activation of lexical and sub-lexical representations in bilingual language comprehension.

### 2.3. Cross-Language Lexical Access in Context

Two types of studies have examined cross-language influences on lexical access in context. One adapts the standard semantic priming paradigm (e.g., Meyer & Schvaneveldt, 1971) to ask whether semantically related words prime target words when the primes and targets appear in different languages. The other asks whether the parallel activation of lexical alternatives in both languages is modulated by the presence of sentence context and whether lexical access within each language is open to the semantic and syntactic influences of sentence context in the other language. We consider briefly the evidence from each of these areas of research.

#### 2.3.1. *Semantic context*

Early research on the bilingual lexicon used the semantic priming paradigm as a means to test the independence of representations across the bilingual's two languages (e.g., McCormack, 1977; Snodgrass, 1984). In semantic priming, a prime word is typically presented briefly and then followed by a target word for lexical decision. When the prime word is semantically related to the target word, lexical decision time is facilitated relative to conditions in which the prime is unrelated to the target (Meyer & Schvaneveldt, 1971). A series of studies using variants of the semantic priming paradigm (e.g., Altarriba, 1990; Chen & Ng, 1989; De Groot & Nas, 1991; Kirsner, Smith, Lockhart, King, & Jain, 1984; Meyer & Ruddy, 1974; Schwanenflugel & Rey, 1986; Tzelgov & Eben-Ezra, 1992) demonstrated that it was possible to observe semantic priming across as well as within languages. The result of these studies suggested that bilinguals access semantic representations that are shared across the two languages (and see Dufour & Kroll, 1995, for related evidence from a categorization paradigm). Although some of these studies attempted to control the methodology of the priming paradigm to ensure that any observed priming could be attributed to automatic processes, many of the early studies can be criticized on methodological grounds because they included a high proportion of related trials and a long interstimulus interval between prime and target words that may have encouraged subjects to develop expectations for the upcoming targets (see Neely, 1991). The use of long prime-to-target intervals in the bilingual experiments is particularly problematic because in some of these experiments subjects may have been encouraged to translate the prime and/or target into the same language. If primes can be translated, even on a small



proportion of trials, then the observed cross-language effects may reflect only the presence of within-language priming in the cross-language conditions.

Subsequent cross-language semantic priming studies attempted to control for these factors with the result that cross-language priming was obtained but only under some conditions. For example, using a masked priming paradigm (see Figure 4 for an illustration), in which participants cannot consciously report the prime information and are therefore not even aware of the bilingual nature of the task, De Groot and Nas (1991) found evidence for cross-language associative priming only when translation equivalents in the two languages were cognates, sharing lexical form as well as meaning. Keatley, Spinks, and De Gelder (1994) reported that even when bilinguals are highly proficient in both languages there are asymmetries in the magnitude of semantic priming, with significant facilitation only with L1 primes and L2 targets (see also Altarriba, 1990, and Tzelgov & Eben-Ezra, 1992).

Recent studies of cross-language priming have reported mixed results with respect to the conditions under which priming is observed. On one hand, a series of masked priming studies (e.g., Gollan et al., 1997; Jiang, 1999) has shown that cross-language priming is observed in lexical decision for translation equivalents even when the two languages involved do not use the same script (Hebrew–English and Chinese–English), but is only consistently reliable with L1 primes and L2 targets (and see Grainger & Frenck-Mestre, 1998, for evidence that priming is observed in semantic categorization but not in lexical decision). Finding cross-language masked priming is particularly striking because participants are unaware of the prime words. When experiments are designed so that the language of target words is blocked, the experiment can be functionally performed in “monolingual mode” (Grosjean, 2001), thus reducing the likelihood that a translation

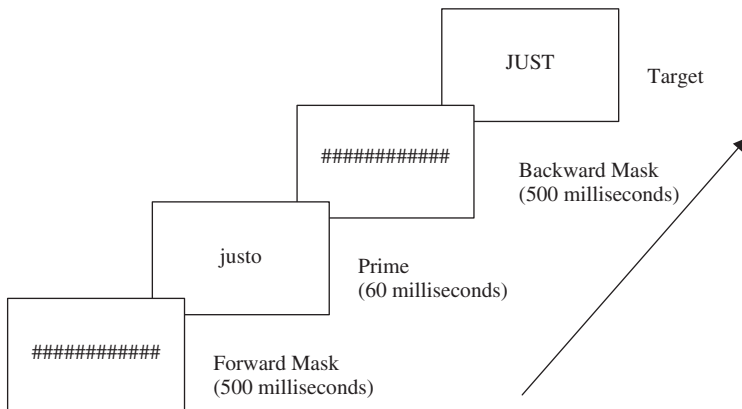


Figure 4. Schematic illustration of the masked priming procedure. In this example, a Spanish prime word (*justo* meaning “just”) is preceded and followed by a forward and backward visual mask. An English target word (*just*) is then presented. The combination of the brief presentation of the prime word in addition to the masking stimuli minimizes conscious processing of the prime word.

strategy can account for the observed priming. However, other studies suggest that there are limits to the degree of cross-language priming that reflect constraints in the nature of bilingual language representations attributable to the age at which individuals acquired the L2 (e.g., Kotz & Elston-Güttler, 2004; Silverberg & Samuel, 2004) and the amount and kind of semantic information that is accessed for each language (e.g., Finkbeiner, Forster, Nicol, & Nakamura, 2004; Kotz & Elston-Güttler, 2004). The constraints observed in cross-language priming also appear to reflect the degree of proficiency in the L2. For example, Kotz and Elston-Güttler found that late L2 learners were able to exploit associative but not categorical relatedness, whereas early bilinguals were able to use both types of information. Although the earlier research was largely compatible with a model of the lexicon in which the two languages access the same semantic information (see Francis, 2005, for a review), the recent studies support a mixed model in which some semantic information is shared and other semantic information is distinct (see De Groot, 1993, and Van Hell, 1998, for models of partly shared cross-language semantics). It remains to be seen to what extent the type of bilingualism determines the ability to acquire all of the subtle nuances of meaning in the L2 that are available in the highly skilled L1 (see Segalowitz & Hulstijn, 2005, for a discussion of issues of automaticity in the L2).

### 2.3.2. *Sentence context*

The findings from research on both auditory speech and text comprehension are compatible with a fairly open lexical system in which information activation flows across both languages. However, as mentioned previously, most of these findings stem from studies in which stimuli were presented in fairly de-contextualized conditions, such as single-word naming and lexical decision. Is the cross-language flow of activation at all constrained when there is a context such as a sentence that provides additional semantic information? To date there have been very few studies that have addressed this question directly (Elston-Güttler, 2000; Greenberg & Saint-Aubin, 2004; Schwartz & Kroll, 2006; Van Hell, 1998). Findings from these studies suggest that cross-language activation can be indeed constrained by the presence of a sentence context, allowing the system to operate in a more language-selective manner. For example, both Schwartz and Kroll (2006) and Van Hell (1998) found that although lexical access for cognates was facilitated in low-constraint sentences, this facilitation was eliminated when the same cognates were in high-constraint sentences.

What is the mechanism that allows processing to proceed in a more language-selective manner in high-constraint sentences? According to Schwanenflugel and colleagues (e.g., Schwanenflugel, 1991; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; but cf. Traxler & Foss, 2000), sentence constraint influences lexical activation through a set of feature restrictions that readers generate as they comprehend sentences. With increasing constraint, an increasing number of restrictions are generated. Lexical entries that conflict with these constrictions are inhibited. Thus, in the bilingual case increasing sentence constraint would lead to the generation of language-specific lexical feature restrictions which would inhibit lexical entries from the non-target language. Support for this account was observed in an eye-movement monitoring

study of Spanish–English bilinguals’ reading of code-switched sentences (Altarriba, Kroll, Sholl, & Rayner, 1996). The eye-movements of Spanish–English bilinguals were monitored as they read high-and low-constraint sentences in English. On half of the trials, one word in each sentence was a code-switched word from the non-target language, Spanish [e.g., He wanted to deposit all of his dinero (money) in the credit union]. Critically, this code-switched word was either a high-frequency word or a low-frequency word in Spanish. Analyses of the first fixation durations on the code-switched words revealed an interaction between sentence constraint and word frequency, such that fixation durations were elevated in high-constraint sentences when the code-switched word was of a high lexical frequency. This suggests that the participants generated both semantic and lexical-level feature restrictions when reading high constraint sentences. That is, when presented with the high frequency, code-switched word (e.g., *dinero*), processing was inhibited because the word met all of the semantic but not the lexical feature restrictions. However, it should be noted that effects of sentence constraint appear to operate at a later point, once processes of initial lexical access have been completed (Greenberg & Saint-Aubin, 2004). In other words, the studies to date cannot rule out an initial non-selective activation of lexical candidates from the non-target language followed by a subsequent inhibition.

An interesting feature of the results on the effects of sentence context, is that a similar pattern has been observed in studies that have examined the effects of non-linguistic factors on cross-language activation, such as instructions, expectations, and working memory resources (e.g., De Bruijn et al., 2001; Dijkstra et al., 2000a; Michael, Dijkstra, & Kroll, 2006; Van Hell & Dijkstra, 2002). Just as the results on sentence context suggest that knowing the language of the sentence has little effect on constraining lexical alternatives in the non-target language, the effects of single word context or expectations appear not to restrict the influence of the other language, at least for the recognition of visually presented words.

### 3. CROSS-LANGUAGE INTERACTIONS AT THE SENTENCE LEVEL

Far fewer studies have examined sentence processing in bilinguals relative to the substantial literature on word recognition and lexical processing. Much of the research on sentence processing in non-native speakers has focused on issues of acquisition, asking either to what extent access to the grammar of the L2 is mediated by transfer from the more highly skilled L1 (e.g., MacWhinney, 1997, 2005) or by the age of acquiring the L2 (e.g., Hahne & Friederici, 2001; Johnson & Newport, 1989; Weber-Fox & Neville, 1996). Only a small number of studies have addressed the issue of how language-specific constraints and biases in one of the bilingual’s languages affect processing in the other language (for recent reviews, see Frenck-Mestre, 2005; Kroll & Dussias, 2004).

Although there is a rich history in psycholinguistics of using cross-linguistic evidence to assess the universality of language processing mechanisms (e.g., Bates, Devescovi, & Wulfeck, 2001; Cuetos & Mitchell, 1988; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996)

only a small number of studies have asked directly how sentence processing is accommodated to the presence of two languages. Two types of experiments have been conducted to address these issues. In one, grammatical constructions that differ across language are the focus. The question is then how the bilingual resolves the potential conflict between the two alternatives in one and the same mind (e.g., Dussias, 2001, 2003; Fernández, 1998, 2003; Frenck-Mestre, 2005). In the other, a priming paradigm has been used to determine whether structural repetitions that typically facilitate performance within language, also facilitate performance across languages (e.g., Hartsuiker, Pickering, & Veltkamp, 2004; Loebell & Bock, 2003).

### 3.1. Cross-Language Parsing

If a bilingual's two languages follow different syntactic arrangements and if those differences give rise to distinct parsing preferences, then bilingualism potentially poses a problem for language processing if the two languages are not represented and accessed independently. A number of solutions to this problem are available in theory. One possibility is that the two languages are treated independently and the parsing preferences associated with each language are engaged appropriately as a function of the language context. As the literature on bilingual word recognition makes clear, there is very little evidence at the lexical level that the two languages function independently. Although it might seem that at the level of the grammar it might be easier to separate the two languages, our brief review of sentence context effects on lexical processing suggests that bilinguals do not use the language of a sentence itself as a strong cue to differentiate the two languages. As we will discuss later, bilinguals are also prepared to code switch with other similarly bilingual individuals and to understand code-switched utterances, suggesting the grammars of both languages are available and engaged.

A second solution to the problem is to bias parsing toward the native or more dominant language. This is a solution that has been discussed in detail in the literature on second language acquisition where there is a great deal of empirical support for transfer from the L1 to the L2 (e.g., see MacWhinney, 1997, 2005 for an illustration of this approach embodied within the Competition Model). However, even a strong transfer account such as the Competition Model includes a developmental component whereby successful L2 learners eventually acquire the cues for the weaker L2. The degree of transfer will depend on the relative proficiency of the bilinguals, with greater L1 influence for less than for more proficient bilinguals.

A third solution to the problem is to assume that the bilingual develops a set of parsing strategies that are uniquely bilingual, in that they represent a mix of the preferences used within each language by native speakers. This is an instance of Grosjean's (1989) well-known admonition that the bilingual is not two monolinguals in one. Recent studies of linguistic convergence (e.g., Bullock & Toribio, 2004; Malt & Sloman, 2003) provide support for the claim that language contact produces a pattern of language use that is distinctly bilingual at all levels of linguistic description.

Although research on bilingual parsing is at an early stage of investigation, recent findings suggest that although it is sometimes possible for advanced L2 learners and proficient bilinguals to acquire native-like parsing preferences for the L2 (e.g., Frenck-Mestre & Pynte, 1997), they may also be slower to process L2 than L1 sentences (e.g., Hoover & Dwivedi, 1998), and more likely to recruit additional working memory resources (e.g., Hasegawa, Carpenter, & Just, 2002; Miyake & Friedman, 1998). What is not yet clear within this emerging body of research is whether the demands on working memory are related to the presence of structural differences between the bilingual's two languages. That is, it may be more difficult to acquire a new form in L2 when it does not exist in the L1 or when the L1 requires distinctions that are not functional in the L2 (e.g., see Juffs, 1998 for an illustration of how otherwise proficient L2 speakers may be limited by these cross-language distinctions in the case of causative-inchoative constructions, and Jiang (2004), who shows that highly proficient Chinese-English bilinguals are restricted in their ability to comprehend subject-verb agreement in their L2 in an online task although they are able to recognize the correct English forms in an offline measure).

The experiments on sentence parsing that are perhaps most critical to a model of bilingual comprehension are those in which structural preferences in the two languages are in conflict. For example, Dussias (2001, 2003) has examined the resolution of such a cross-language conflict in the case of attachment preferences for temporally ambiguous sentences containing a relative clause. Dussias (2003) uses the sentence below to illustrate how native speakers of English and Spanish make distinct structural commitments.

Peter fell in love with the daughter of the psychologist who studied in California.

Native speakers of English prefer to attach the relative clause, *who studied in California*, to the immediately preceding noun, *psychologist*. Thus, in response to the question, "Who studied in California?" a native English speaker would respond, the "psychologist." This preference has been called low attachment (e.g., Frazier & Rayner, 1982). In contrast, native Spanish speakers prefer high attachment, so their answer to the same question would be, "the daughter" (but see also Gilboy, Sopena, Clifton, & Frazier, 1995). Carreiras, Salillas, and Barber (2004) recently provided evidence for the high-attachment preferences of native monolingual Spanish readers in an ERP study. They found a P600 effect, typically observed in event related potential (ERP) studies of sentence processing in response to a syntactic violation, when Spanish readers processed a sentence that was consistent with low attachment, contrary to their ordinary bias. What happens when both languages are available to highly proficient bilinguals? Dussias reports that Spanish-English bilinguals prefer the low-attachment strategy even when they are reading in Spanish, their native language. That is, the preference for high attachment appears to change once a high level of proficiency is achieved in the L2 (i.e., English). Like the results reviewed earlier on bilingual word recognition, these findings suggest that not only does the L1 affect the L2, but the L2 can come to influence the L1, even at the level of the grammar. Dussias considers the possibility that the dramatic shift to low attachment for the native Spanish speaking bilinguals may be due to nature of their exposure to English in a predominantly English-speaking environment in the US. A critical question, and one currently under

investigation, is whether the frequency of exposure to the L2 or proficiency in the L2 *per se* is the key factor determining this pattern of performance.

A summary of the results on bilingual sentence parsing at this early stage in the research program suggests that almost all of the possible outcomes can be obtained. Bilinguals can sometimes adopt L2 structures as if they were native speakers of the L2, but in other contexts they are influenced by transfer from the L1 or the L1 itself is modified by the use of the L2. A clear goal in the next stage of research will be to identify the range of factors that modulates these different patterns and particularly the linguistic constraints associated with each outcome. There are a number of existing results that are apparently conflicting and the resolution of these differences will be important to the development of a comprehensive model of bilingual sentence processing. For example, the studies that have examined the effects of age of acquisition on sensitivity to syntax in the L2 (e.g., Hahne & Friederici, 2002; Johnson & Newport, 1989; Weber-Fox & Neville, 1996), suggest that there is a limit in the extent that late second language learners can fully acquire the syntax of the L2, even those who are highly proficient in the L2 and have lived in the L2 environment for many years. However, the bilinguals in the studies we have reported typically acquired their L2 late and yet were able to adopt native-like processing preferences in the L2 or, even more dramatically, to have the L1 take on the properties of the L2. It remains to be seen whether the degree of constraint in L2 sentence processing is determined by the type of structures that have been examined, by the degree of proficiency of the bilinguals, a factor that is difficult to assess perfectly, or by other contextual factors that influence the degree to which the native language is maintained actively when bilinguals live in an L2 environment. However these issues are resolved, however, the fact that it is possible to demonstrate that the L2 can influence the L1 is again consistent with the view that the bilingual's two languages are open to interactions and that some of those interactions may have long-term consequences.

### **3.2. Cross-Language Syntactic Priming**

Although most syntactic priming studies examine production rather than comprehension, we describe the main results of these studies briefly because those results converge closely with the evidence reviewed above on parsing. Syntactic priming is the phenomenon whereby the production of a target sentence is influenced by the syntactic form of a previously produced prime sentence (Bock, 1986). For example, for a monolingual speaker of English, the probability of producing a sentence describing a picture in active vs. passive voice is a function of whether a spoken prime sentence is active or passive. In the bilingual research, the question has been whether a switch of language, and therefore syntax, from prime to target sentence, will disrupt priming relative to the monolingual case. The few studies that have examined cross-language syntactic priming have reported effects of priming that are very similar to those found within language, suggesting that there is a common basis for this effect across languages (e.g., Hartsuiker et al., 2004), although there are some suggestions as well that the range of priming may be more limited across than within languages (e.g., Loebell & Bock, 2003). Recent work has shown that structural priming can be observed in cases in which lexical priming alone cannot

account for the results and that lower frequency or less-dominant structures are more susceptible to priming (e.g., Scheepers, 2003). In the cross-language case, this means that syntactic priming will be more likely from L1 to L2, the more dominant to the less-dominant language (Flett, Branigan, Pickering, & Sorace, 2005). Hartsuiker et al. argue that the syntactic representations computed for each of the bilingual's two languages access the same abstract information, rendering the syntactic level integrated across languages and open to code switching, the topic to which we turn next.

### **3.3. Other Approaches to Bilingual Sentence Processing**

#### *3.3.1. Understanding code-switched sentences*

A surprising gap in the literature on bilingual sentence processing is that very few studies have investigated the comprehension of spoken code switches between languages. Code switching is a phenomenon common within bilingual communities. However, most of the research on this topic has been conducted from a linguistic rather than psycholinguistic perspective, with the goal to elucidate the syntactic constraints that govern allowable switches (e.g., Muysken, 2000; Myers-Scotton, 2002). The few studies that have examined the consequences of code switching for language processing have focused on processes occurring primarily at the lexical level rather than the syntactic level. A number of studies have examined lexical-level code switching while bilinguals read sentences (e.g., Altarriba et al., 1996; Moreno, Federmeier, & Kutas, 2002). Although language switching in comprehension tasks is an interesting phenomenon in and of itself for what it tells us about how effectively bilinguals can use expectations about the language they are processing to control subsequent language selection (e.g., Thomas & Allport, 2000), it can be argued that mixed language presentations during reading are rare, whereas code switches in spoken discourse are common. It would therefore seem critical to investigate this issue in spoken language contexts. Again, the few studies that have examined code switches in speech, have also focused at the lexical level. To illustrate, a number of experiments have asked how bilinguals comprehend a guest word spoken in the non-target language (i.e., not in the language of a preceding sentence context). The results of these studies support the conclusions of the word recognition research reviewed earlier in showing that information about the non-target language is available even during sentence processing, but demonstrate that the scope of activation of non-target alternatives is constrained to some degree by cues available to the listener (e.g., Li, 1996; Grosjean, 1988; Soares & Grosjean, 1984). An interesting observation is that code switches in written sentence contexts are in some respects less disruptive to reading than within-language lexical switches. Moreno et al. (2002) showed that the typical N400 effect observed in the ERP record when a lexical violation is encountered is greater when within-language synonyms are presented than when a cross-language translation appears. One interpretation of this result is that code switching is a relatively natural phenomenon, a conclusion that is compatible with the high degree of parallel activity observed across languages. It will be critical in the next phase of research to begin to examine syntactic-level constraints in the online comprehension of code switches, particularly under circumstances in which the bilingual's two languages differ syntactically.

### 3.3.2. *Reading for translation*

A final area of research in which sentence processing has been examined in bilinguals concerns the processes that are engaged when a translator is comprehending a sentence in one language for the purpose of producing it in the other language. The processes that support proficient translators in achieving real time simultaneous translation and interpretation are fascinating in and of themselves because they represent an extraordinary feat of cognition (see Christoffels & De Groot, 2005 for a recent review of the psycholinguistics of translation and interpretation). For present purposes, the question of interest with respect to translation, is how incoming material is comprehended as a function of how it will be used. Earlier research on sentence comprehension within the native language (e.g., Aaronson & Ferres, 1986) has shown that comprehension strategies are adapted to the goals of the task. Macizo and Bajo (2006) showed that when translators read for the purpose of translation, there is activation of information in the target language to be spoken during the comprehension process. When they read only for the subsequent task of repeating what they have read, there is little evidence of cross-language activation. These results suggest that task goals influence the degree to which both languages are active. Of particular interest is that translation makes greater demands on working memory resources than simple reading and those demands are reflected during the comprehension process, prior to actual production. One implication of the findings with translators is that the degree to which the other language is required in the larger discourse context appears to influence sensitivity to both languages during initial comprehension (see Grosjean, 2001 for a related argument about language mode). For translators, that context may be the likelihood of having to produce the currently processed sentence in the other language. For ordinary bilinguals, it may be related to the likelihood that they will interact with other bilinguals who are likely to switch into the other language completely or in part in code switched exchanges.

## 4. FACTORS THAT INFLUENCE BILINGUAL COMPREHENSION

In each area of research reviewed above, we focused on the performance of highly proficient bilinguals. However, comprehension is also open to the influence of a variety of factors that are likely to modulate performance. One factor includes the individual differences that affect the cognitive resources that are available to be recruited during language processing (see Perfetti, this volume for a review of this work within the native language). As we noted in the review of research on bilingual sentence processing, a great deal of evidence suggests that the L2 makes greater demands on memory and attention than the L1, even for relatively skilled bilinguals (e.g., Miyake & Friedman, 1998).

Another factor is the similarity between the two languages. Languages differ syntactically, morphologically, and phonologically and those differences are likely to affect the ease of cross-language comprehension in both listening and reading. In reading, languages also differ with respect to whether they are alphabetic and use the same or a



different script. A Dutch–English bilingual may easily mistake a written word in English for Dutch, but a Chinese–English bilingual will never make that mistake. Although differences across languages at each level modulate the pattern of processing (e.g., MacWhinney, 1997; Thomas & Allport, 2000; Vaid & Frenck-Mestre, 2002), there is very little evidence that suggests that once an individual becomes relatively proficient in the L2 they restrict the degree of cross-language interaction at either the lexical, sub-lexical, or sentence levels or serve as a cue to maintain greater separation across the bilingual's two languages.

If language status does not readily enable language selection, then how do proficient bilinguals effectively use the intended language without frequent intrusions from the unintended language? In this chapter, we have focused our review on comprehension but there is also a recent literature on bilingual production that demonstrates that language non-selectivity is not restricted to comprehension alone. Although spoken production in bilinguals is initiated by an idea to be expressed in words and sentences, a picture to be named or described, and words or sentences to be translated, the conceptually driven nature of production does not itself appear to restrict activation to the target language alone (e.g., Costa, 2005; Kroll et al., 2006). The presence of mutual activity across the bilingual's two languages in both comprehension and production suggests that another mechanism must be in place to allow attention to be appropriately directed so that the correct language choices are made and, at the same time, that systematic code-switching can be accommodated without incurring a significant processing cost.

One solution to this apparent problem is to hypothesize that cognitive mechanisms outside the linguistic representations themselves function to resolve the observed cross-language competition, either by modulating the relative activation of the unintended language or by actively inhibiting candidates from the non-target language (e.g., see Green, 1998 for an illustration of how such an inhibitory mechanism might operate in production). A model that incorporates an extra-linguistic mechanism for bilingual word recognition has been described by Dijkstra and Van Heuven (2002). The BIA+ model, shown in Figure 5, includes both a lexical identification system and a task schema system. The lexical identification system, an embellishment of the BIA model seen in Figure 3, represents lexical and sub-lexical information in each language and their interactions. The lexical system is hypothesized to be encapsulated in the sense that language-specific selection within the lexical system itself is possible in response to linguistic context but not affected directly by more cognitive, non-linguistic factors, such as expectations and instructions. On this view, bilingual word recognition is fundamentally a data-driven process that is uninfluenced by top-down factors until quite late in the process. The task schema system controls not only the output of the lexical identification system with respect to the mapping of language output to response processes, but also the manner in which language output is weighted with respect to decision criteria. As such, the model can account for many of the word-level phenomena we have reported. Bilingual word recognition is fundamentally language non-selective and even sentence context per se does not appear to

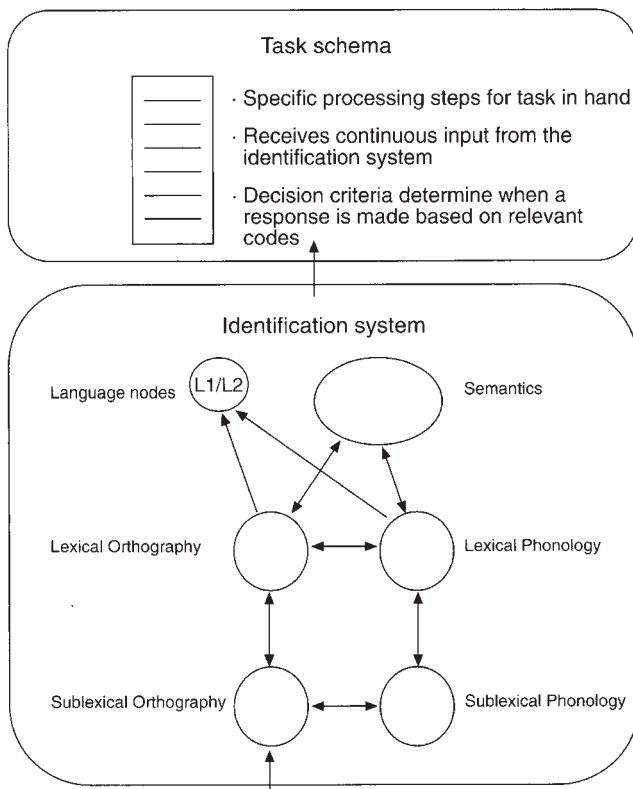


Figure 5. The BIA+ Model (adapted from Dijkstra & Van Heuven, 2002).

override the parallel activity across the two languages (see the earlier section on lexical access).

There remains a great deal of work to be done to determine to what extent the principles embodied within a model like BIA+ will serve as a general foundation on which a more comprehensive account of bilingual comprehension can be developed. However, it provides a useful basis on which to begin to identify those factors that constrain the basic architecture of the bilingual’s language system and those that reflect the manner in which bilingualism affects cognitive control. The recent work by Bialystok and colleagues (e.g., Bialystok, 2005; Bialystok et al., 2004) provides compelling evidence for the positive cognitive consequences that bilingualism appears to confer to young bilingual children and to elderly bilinguals in the realm of executive function. Bilinguals are not cognitively superior to monolinguals in general, but quite specifically in tasks that require inhibitory control in which irrelevant information or responses must be ignored. It is tempting to speculate that the control mechanisms of the sort included in Dijkstra and Van Heuven’s (2002) BIA+ model, required to

modulate proficient performance, may contribute to the development of enhanced executive function in bilinguals.

## 5. CONCLUSIONS

In this chapter, we reviewed some of the recent research on the processing of words and sentences by bilingual speakers. The picture presented by our review suggests a language system that is highly permeable across the bilingual's two languages, with information about words and grammatical structures activated concurrently even while a bilingual is reading text or listening to speech in one of his or her two languages alone. A very counterintuitive aspect of this body of research is that the activity of the unintended languages is not simply a matter of proficiency. Both languages appear to be active in even highly proficient bilinguals. Although much of this research is at a very early stage of development, it holds important implications for characterizing bilingual performance and for the way in which bilingualism provides a model more generally for investigating constraints and plasticity in language processing. At the heart of this review is the observation that bilinguals themselves are not special. To the contrary, more of the world's population is bilingual than not and most of the cross-language interactions we have reviewed are related to phenomena observed within language in the presence of ambiguity. However, the presence of two active and competing languages makes the bilingual an especially informative source for psycholinguists interested in how cognitive systems compete and in how the resulting competition is resolved. We are confident that the contribution of this approach will be increasingly valuable in mapping the relations between language and cognition and their neural underpinnings.

## ACKNOWLEDGEMENT

The writing of this chapter was supported in part by NSF Grant BCS-0418071 and NIH Grant R01MH62479 to Judith F. Kroll.

## REFERENCES

- Aaronson, D., & Ferres, S. (1986). Reading strategies for children and adults: A quantitative model. *Psychological Review*, *93*, 89–112.
- Altarriba, J. (1990). *Constraints on interlingual facilitation effects in priming in Spanish–English bilinguals*. Unpublished dissertation, Vanderbilt University, Nashville Tennessee.
- Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye-fixation and naming times. *Memory and Cognition*, *24*, 477–492.
- Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language*, *36*, 550–568.

- Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*, 52, 369–396.
- Best, C. T. (1995) A direct realist view of cross-language speech perception. In: W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–206). Baltimore: York Press.
- Bhatia, T. K., & Ritchie, W. C. (2004). Introduction. In: T. K. Bhatia, & W. C. Ritchie (Eds). *Handbook of Bilingualism* (pp. 1–2). Cambridge, MA: Blackwell Publishers.
- Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In: J. F. Kroll, & A. M. B. De Groot (Eds). *Lexical access in bilingual production. Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 417–432). New York: Oxford University Press.
- Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon Task. *Psychology and Aging*, 19, 290–303.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Bosch, L., Costa, A., & Sebastian-Galles, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12, 189.
- Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2, 29–49.
- Bullock, B. E., & Toribio, A. J. (2004). Introduction: Convergence as an emergent property in bilingual speech. *Bilingualism: Language and Cognition*, 7, 91–93.
- Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, 20, 98–105.
- Chen, H. -C., & Leung, Y. -S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 316–325.
- Chen, H. -C., & Ng, M. -L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition*, 17, 454–462.
- Christoffels, I. K., & De Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In: J. F. Kroll, & A. M. B. de Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 326–348). New York: Oxford University Press.
- Colomé, A. (2001). Lexical activation in bilinguals' speech production: language-specific or language-independent? *Journal of Memory and Language*, 45, 721–736.
- Costa, A. (2005). Lexical access in bilingual production. In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308–325). New York: Oxford University Press.
- Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1283–1296.

Costa, A., Miozzo, M., Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365–397.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*, 73–105.

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1989). Limits on bilingualism. *Nature*, *320*, 229–230.

De Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects on interlingual homograph recognition: Evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition*, *4*, 155–168.

De Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In: R. Schreuder, & B. Weltens (Eds), *The bilingual lexicon* (pp. 27–51). Amsterdam, Netherlands: John Benjamins Publishing Company.

De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*, 90–123.

Dijkstra, T. (2005). Bilingual word recognition and lexical access. In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). New York: Oxford University Press.

Dijkstra, T., De Bruijn, E., Schriefers, H., & Brinke, S. T. (2000a). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition*, *3*, 69–78.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 496–518.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000b). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, *42*, 445–464.

Dijkstra, T., & Van Hell, J. G. (2003). Testing the language mode hypothesis using trilinguals. *International Journal of Bilingual Education and Bilingualism*, *6*, 2–16.

Dijkstra, A. & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *23*, 175–197.

Dijkstra, A., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*, 51–66.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory and Cognition*, *23*, 166–180.

Dussias, P. E. (2001). Bilingual sentence parsing. In: J. L. Nicol (Ed.), *One Mind, two languages: Bilingual language processing* (pp. 159–176). Cambridge, MA: Blackwell Publishers.

- Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, 25, 529–557.
- Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 889–906.
- Elston-Güttler, K. (2000). *An enquiry into cross-language differences in lexical-conceptual relationships and their effect on L2 lexical processing*. Unpublished dissertation, University of Cambridge, Cambridge, UK.
- Fernández, E. M. (1998). Language dependency in parsing: Evidence from monolingual and bilingual processing. *Psychologica Belgica*, 38, 197–230.
- Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins Publisher.
- Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, 51, 1–22.
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In: W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–272). Baltimore: York Press.
- Flett, S., Branigan, H., Pickering, M., & Sorace, A. (2005). Syntactic processing in second language production. Paper presented at the 5th international symposium on bilingualism, Barcelona, Spain, March.
- Francis, W. S. (2005). Bilingual semantic and conceptual representation. In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 251–267). New York: Oxford University Press.
- Francis, W. S., Tokowicz, N., & Kroll, J. F. (2003). *Translation priming as a function of bilingual proficiency and item difficulty*. Poster presented at the 4th international symposium on bilingualism, Tempe, AZ, April.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.
- Frenck-Mestre, C. (2005). Ambiguities and anomalies: What can eye movements and event-related potentials reveal about second language sentence processing? In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 268–281). New York: Oxford University Press.
- Frenck-Mestre, C., & Pynte, J. (1995). Lexical influences on parsing strategies: Evidence from eye movements. In: J. M. Findlay, R. W. Kentridge, & R. Walker (Eds), *Eye movement research: Mechanisms, processes and applications* (pp. 433–444). Amsterdam: Elsevier, North Holland.
- Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology*, 50A, 119–148.

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 305–313.

Gilboy, E., Sopena, J. -M., Clifton, C, Jr., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, *54*, 131–167.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1122–1139.

Grainger, J., & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes*, *13*, 601–623.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81.

Greenberg, S. N., & Saint-Aubin, J. (2004). Letter detection for homographs with different meanings in different language texts. *Bilingualism, Language and Cognition*, *7*, 241–253.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, *3*, 233–274.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15.

Grosjean, F. (2001). The bilingual's language modes. In: J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Cambridge, MA: Blackwell.

Hahne, A., & Friederici, A.D. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*, 123–142.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? *Psychological Science*, *15*, 409–414.

Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage*, *15*, 647–660.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*, 213–229

Hoover, M. L., & Dwivedi, V. D. (1998). Syntactic processing by skilled bilinguals. *Language Learning*, *48*, 1–29.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, *44*, 2–31.

- Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition*, 5, 225–239.
- Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition*, 2, 59–75.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–634.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15, 314–318.
- Juffs, A. (1998). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Language Research*, 14(4), 406–424.
- Jusczyk, P. W., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, 40, 62–82.
- Keatley, C., Spinks, J., & de Gelder, B. (1994). Asymmetrical semantic facilitation between languages. *Memory and Cognition*, 22, 70–84.
- Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*, 23, 519–539.
- Kotz, S. A., & Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, 17, 215–235.
- Kroll, J. F., Bobb, S., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9, 119–135.
- Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In: M. Gruneberg, P. Morris, & R. Sykes (Eds), *Practical aspects of memory* (Vol. 2, pp. 389–395). London: Wiley.
- Kroll, J. F., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In: A. M. B. De Groot, & J. F. Kroll (Eds), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Mahwah, NJ: Lawrence Erlbaum Publishers.
- Kroll, J. F., & Dussias, P. (2004). The comprehension of words and sentences in two languages. In: T. K. Bhatia, & W. C. Ritchie (Eds), *Handbook of Bilingualism* (pp. 169–200). Cambridge, MA: Blackwell Publishers.



Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18, 137–171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). New York: Oxford University Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., & Stevens, K. N. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.

La Heij, W., Hooglander, A., Kerling, R., & Van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language*, 35, 648–665.

Li, P. (1996). Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language*, 35, 757–774.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, 41, 791–824.

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, 99, 1–34.

MacWhinney, B. (1997). Second language acquisition and the competition model. In: A. M. B. De Groot, & J. F. Kroll (Eds), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–142). Mahwah, NJ: Lawrence Erlbaum Publishers.

MacWhinney, B. (2005). A unified model of language acquisition. In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). New York: Oxford University Press.

Malt, B. C., & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, 6, 47–67.

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. *Developmental Psychology*, 28, 396–405.

Marian, V., & Spivey, M. J. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, 6, 97–115.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review*, 88, 375–405.

McCormack, P. D. (1977). Bilingual linguistic memory: The independence–interdependence issue revisited. In: P. A. Hornby (Ed.), *Bilingualism: Psychological, social, and educational implications*. New York: Academic Press.

- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143–178.
- Meyer, D. E., & Ruddy, M. G. (1974). Bilingual word recognition: Organization and retrieval of alternative lexical codes. Paper presented at the Eastern Psychological Association meeting, Philadelphia, PA, April.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.
- Michael, E., Dijkstra, T., & Kroll, J. F. (2006). *Individual differences in memory span do not modulate the degree of language nonselectivity in bilingual word recognition*. Manuscript in preparation.
- Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: The role of working memory. In: A. F. Healy, & L. E. Bourne, Jr. (Eds), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, 80, 188–207.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: D. Besner, & G. W. Humphreys (Eds), *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Erlbaum.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in vowel acquisition. *Cognition*, 64, B9–B17.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445.
- Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, 253, 437–438.
- Potter, M. C., So, K. -F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23–38.
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89, 179–205.
- Schwanenflugel, P. J. (1991). Contextual constraint and lexical processing. In: G. B. Simpson (Ed.), *Understanding word and sentence. Advances in psychology* (Vol. 77, pp. 23–45). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

- Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 344–354.
- Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual. *Journal of Memory and Language*, *25*, 605–618.
- Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232–252.
- Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*, 197–212.
- Schwartz, A. I., Kroll, J. F., & Diaz, M. (in press). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*.
- Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 974–989.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*, 240–255.
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In: J. F. Kroll, & A. M. B. De Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). New York: Oxford University Press.
- Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, *51*, 381–398.
- Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, *23*, 3–22.
- Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition*, *12*, 380–386.
- Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281–284.
- Strange, W. (Ed.). (1995). *Speech perception and linguistic experience*. Baltimore, MD: York Press.
- Sunderman, G., & Kroll, J. F. (in press). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*.
- Talamas, A., Kroll, J. F., & Dufour, R. (1999). Form related errors in second language learning: A preliminary stage in the acquisition of L2 vocabulary. *Bilingualism: Language and Cognition*, *2*, 45–58.

- Thomas, M. S. C., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language*, *43*, 44–66.
- Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1266–1282.
- Tzelgov, J., & Eben-Ezra, S. (1992). Components of the between-language semantic priming effect. *European Journal of Cognitive Psychology*, *4*, 253–272.
- Van Hell, J.G. (1998). *Cross-language processing and bilingual memory organization*. Unpublished dissertation, University of Amsterdam, Amsterdam, The Netherlands.
- Van Hell, J.G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*, 193–211.
- Van Hell, J.G., & Dijkstra, A. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review*, *9*, 780–789.
- Vaid, J., & Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French-English bilinguals. *Brain and Language*, *82*, 47–53.
- Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*, 458–483.
- Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, *11*, 407–442.
- Von Studnitz, R. E., & Green, D. (2002). Interlingual homograph interference in German–English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition*, *5*, 1–23.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.
- Weber-Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*, 231–256.
- Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*, 672–683.

This page intentionally left blank

## Chapter 26

### **Psycholinguistic and Neurolinguistic Perspectives on Sign Languages**

*David P. Corina and Heather P. Knapp*

Signed languages of the deaf are naturally evolving linguistic systems exhibiting the full range of linguistic complexity found in speech. While the field of sign language psycholinguistics is still in its adolescence, most phenomena that form the foundation of our understanding of speech perception and production, such as lexicality and frequency effects, semantic- and form-based priming, categorical perception, production slips, and tip-of-the-finger phenomena, are well attested in sign languages. In addition, neurolinguistic studies have helped identify brain regions critical for sign language and have documented the dissolution of sign language in cases of sign aphasia. New findings from neuroimaging have confirmed and extended our understanding of the intricacies of the neural system underlying sign language use. Taken together, these studies provide a privileged avenue for understanding the generality of the cognitive constraints evidenced in language processing and the biological basis for human language.

The field of psycholinguistics strives to understand the relationship between linguistic structure and cognitive processing. Linguistic, and especially psycholinguistic, research of signed languages is still in its infancy. It has only been approximately 30 years since American Sign Language (ASL) has been regarded as an autonomous and linguistically complex language (Klima & Bellugi, 1979). Additional descriptions and analyses of signed languages from around the world have emerged over the years, rapidly broadening the sign language linguistics landscape (Brentari, 2001).

There are at least two important yet oft-misunderstood facts regarding sign languages. The first concerns sign universality. Just as there are many different spoken language communities around the world (speakers of Quechua, Farsi, Hmong, Portuguese, etc.) there are many different sign language communities (e.g., Langue des Signes du Québec, Deutsche Gebärdensprache, Taiwan Ziran Shouyu, ASL). Although there has been far less documentation of the range of “foreign” sign languages, these languages are known to have evolved spontaneously in isolated communities where there has been a preponderance of deaf individuals. Such situations are not rare, as genetic influences on the transmission of deafness are well attested. In addition, just as spoken languages have

formal similarities due to historical connections (e.g., Dutch, German and English), so do signed languages (e.g., Langue des Signes Française, American Sign Language, Lenguaje de Signos Mexicano).

The second common confusion regarding sign languages is the degree to which they are related to spoken language. Signs are not invented hand symbols that simply represent the words of a spoken language, nor are they pantomime of everyday actions. The forms and structures of sign languages are autonomous from their surrounding spoken language communities. While many signed languages do have auxiliary systems for borrowing spoken language terms into a manual form (i.e., fingerspelling systems), these sub-systems are typically not considered “native” components of Deaf signed languages.

Detailed descriptions of the physical attributes of natural signed languages were first provided by William Stokoe in his seminal works (e.g., Stokoe, Casterline, & Croneberg, 1976). These notations classify ASL signs according to the shape taken by the hand during the articulation of a sign (hereafter *handshape*), the location of the shaped hand in relation to the body, and the movement in space produced during a sign's production. These separate dimensions of sign formation came to be known as “parameters.” Subsequent work by Battison (1978) argued for the inclusion of handshape orientation as an additional parameter. Over the last few decades, researchers have elaborated on these initial descriptions of sign structure, proposing models that encompass phonological, morphological, syntactic, semantic, and pragmatic properties of signed languages. Descriptions cast within these formal linguistic models help elucidate those properties of sign languages that are common to all human languages, as well as to isolate those properties that reflect the unique structure of manually articulated and visually perceived human languages.

## 1. SINGLE SIGN RECOGNITION

As in spoken language psycholinguistics, much of the literature on sign language psycholinguistics has sought to uncover the structure of the lexicon, and the processes governing the access of lexical items stored therein. Most contemporary models of spoken and written word recognition conceptualize lexical access as a matching process between a perceptual signal that accrues over time, and potential lexical candidates stored in memory (e.g., Cohort model, Neighborhood Activation model and TRACE; for a review, see Jusczyk & Luce, 2002). Competitive activation among lexical candidates ultimately yields a single word percept. Questions that are especially relevant to sign language psycholinguistics, therefore, include how the *physical* properties of the sign language signal affect the way in which the lexicon is initially accessed en route to word recognition, and what organizational principles dictate how signs are stored within the lexicon.

In spoken languages, word recognition is achieved through the mapping of the acoustic speech signal received by the ear onto word forms stored as mental representations in the listener's mind. The mapping process unfolds in time as the acoustic signal

of the word accrues, and is thought to be mediated by psychological processes which, in part, serve to rectify variances in the signal due to coarticulatory factors (e.g., allophonic variation) and cross-speaker differences. The details of the initial phonetic mapping, the structure of the representations, and the structure of the mental lexicon have been elucidated by psycholinguistic research. Well-accepted findings include lexicality and word-frequency effects, context effects (including semantic- and form-based priming), and neighborhood effects.

*Lexicality effects* refer to the finding that words are recognized faster than non-words. This basic finding suggests that the presence of a mental representation (i.e., a known word) provides a target for a search through the lexicon, whereas the lack of a representation engenders an exhaustive and unfilled search and thus incurs a processing disadvantage. *Word frequency effects* suggest that words that are highly frequent in the language (e.g., BABY) are recognized faster than low-frequency words (e.g., DRAGON). Psycholinguists have proposed a thresholding account of these data, with highly frequent words having a lower resting threshold and thus require less information for recognition. *Context effects* may have a temporal component, whereby a word that has been previously encountered may affect the subsequent processing of incoming words. For example, the recognition of the word CAT is faster if subjects previously heard the word PIG versus an unrelated word like PAIL. This robust effect is known as *semantic priming*. These data suggest that the lexicon is structured along a semantic dimension, such that semantic features may be shared between entries and thus may be co-activated, leading to speeded processing of related entries.

While only a few psycholinguistic studies of on-line processing in signed languages exist (see Emmorey, 2002 for a recent review of psycholinguistic studies of ASL), there is a growing body of literature confirming that factors affecting spoken word recognition also influence the recognition of signed language signs. For example, the importance of the lexical status of a sign form was demonstrated in a study in which participants were required to make a lexical decision about a target sign, after first being primed by a related sign or sign-form. Reaction times were significantly slower to formationally possible but non-existing ASL signs than to real ASL signs (Corina & Emmorey, 1993). Similarly, Dye and Shih (in press) reported data from native users of British Sign Language which showed lexicality effects (with slower responses to non-signs) in that language.

## 2. FROM SIGNAL TO REPRESENTATION

Although signs and spoken words are formationally quite different, we have reason to believe that words in each modality unfold in time in a lawful fashion, and that accessing lexical representations may vary as a function of sublexical properties. For example, in a gating task study by Emmorey and Corina (1990), signs were partitioned into 33 ms parts and presented to participants cumulatively. Participants identified the location of the sign first, followed quickly by the handshape and finally the movement. Signs located in neutral space were recognized before those located upon the face, presumably because the



target location of the sign was achieved earlier for neutral space signs. In addition, it was observed that signers could anticipate changes in signs that included a handshape change prior to full articulation of the handshape. This demonstrates that in sign recognition, as with spoken language, co-articulatory factors may assist word identification.

Finally, these studies confirmed that signs, though typically longer in duration than spoken words, were identified very quickly. In contrast to spoken English words in which approximately 83% of a word must be heard before identification occurs, sign identification occurred only after approximately 35% of the sign form had been seen. Two factors that may account for this finding include the relatively greater simultaneous packaging of phonological information within a sign and the fact that few signs share an initial phonological shape, leading to a reduced number of competing lexical items (i.e., reduced initial cohort size).

### 3. FORM-BASED STRUCTURE OF THE LEXICON

While the lexicon may be organized along a semantic dimension such that words sharing a meaning relation prime one another, there is also substantial support for structural properties of words influencing processing. For example, the processing of the word CAT can be affected by a previous encounter with a similar sounding word like MAT more than an unrelated word like LIP. This suggests a further dimension of lexical organization that honors a structural relationship between words (e.g., the phonological status of shared “rime”).

In sign, frequency and semantic context speed recognition. In contrast, structural relations (e.g., phonological composition of words and lexical neighborhoods) have been reported to typically produce inhibitory effects. Lexical decision studies have sought to establish processing effects based upon shared formational similarity (i.e., signs sharing one or more parameters), but have produced discrepant findings. For example, Corina and Emmory (1993) reported *inhibitory* effects for targets sharing an articulatory location with primes, no effects for shared handshapes and a facilitatory effect for movement. More recently, Corina and Hildebrandt (2002) investigated movement and location priming at 500 and 100 ms interstimulus interval (ISI) lags. They found no evidence of phonological priming for either Movement or Location at the 500 ISI lag, though inhibitory non-significant trends (Movement  $p < 0.064$ , Location  $p < 0.088$ ) were observed both for location and movement at 100 ISI lag. In contrast, Dye and Shih (in press) found some evidence of facilitatory phonological priming in native signers for signs that shared a common location and/or location and movement. Importantly, these effects were not observed when primes were non-signs, suggesting that the observed priming effect cannot be divorced from lexicality.

The inconsistent findings in structural priming studies of signed languages may reflect different sources of activation. The studies of Corina and Emmory (1993) and Dye and

Shih (in press), which report facilitation, document unusually long reaction times for sign target detection (>1100 ms), while the study that reported inhibition has reaction times in the 800 ms range. There is a growing consensus in the psycholinguistic literature that these longer reaction times may reflect post-lexically mediated or controlled processing strategies, while inhibition represents the effects of automatic priming. That is, these inhibitory effects reflect lexical rather than post-lexical access.

A related and well-known factor influencing lexical access across both written and spoken language modalities is the composition of the candidate set of lexical entries from which a single target form must emerge. In spoken languages these effects are conceptualized as owing to competition among formationally similar lexical entries—so-called “*neighborhood effects*.” The metric of lexical neighborhood similarity has been traditionally defined in terms of phonological properties in studies of spoken words, and orthographic properties in studies of written words. For example, Luce and Pisoni (1998) derive neighborhood similarity by considering the number of words that could be obtained by a single phoneme substitution, addition, or deletion. Numerous studies have demonstrated that these similarity properties influence word recognition, as shown in the perceptual identification accuracy, naming and lexical decision latencies, and priming effects, both in spoken and visual word recognition (for discussion, see Jusczyk & Luce, 2002). The frequency of individual lexical items may also interact with properties of similarity. Importantly, the effects of similarity and word frequency have been observed across many languages. A general characteristic of these effects is that lexical access tends to be inhibited as similarity increases, especially when lexical frequency is held constant.

A series of studies conducted with Lengua Signos Español (LSE) have examined the interaction between sign similarity and frequency (Carreiras, Gutiérrez-Sigut, Baquero, & Corina, submitted). As sign frequency counts are unavailable for most signed languages, this measure was formalized as a familiar rating based upon how often a particular sign was used. Lexical similarity was evaluated using the Hamburg Sign Language Notation System (Prillwitz, Leven, Zienert, Hanke, & Henning, 2004), whereby the degree of similarity for the parameters of location and handshape were calculated by counting how many signs were produced in a particular location or articulated with a specific handshape. These studies reveal that for neighborhoods defined by location, low-frequency signs are recognized slower in high density neighborhoods compared to sparse neighborhoods. A different pattern was observed for neighborhoods defined by handshape. Here, low-frequency signs were recognized faster in dense neighborhoods compared to sparse neighborhoods.

Several explanations for the processing differences for location and handshape are offered. In many theoretical models of sign structure, handshapes are represented as multi-featured, compositional and hierarchically ordered entities that serve as auto-segments within skeleton-based phonological treatments (Corina & Sandler, 1993). Featural representations of location tend to be far less compositional, often adopting a unitary representation such as [shoulder] for a sign articulated on or near the shoulder (Brentari, 1998). These variations in richness of representation may differentially impact lexical activation.

It is interesting to note that in spoken word recognition there is some evidence for word onset information producing inhibitory effects (e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1998). As noted above in the gating studies of sign recognition, location was the first parameter to be accurately and consistently identified during the unfolding of a sign, followed by handshape. This is not to suggest that location serves as a syllabic onset in sign languages, but rather that there may be a special status of first recognized word/sign elements across languages.

#### 4. EFFECTS OF LANGUAGE EXPERIENCE

A different approach to the study of the underlying structural properties of signs and the sign lexicon is through the study of *phonological awareness* in deaf signers. Phonological awareness reflects an understanding that words are not unsegmented wholes, but are made up of component parts. Some structural groupings in spoken languages have a special status that can often be observed in stylized language usage such as poetry, language games, and song. For example, words that rhyme (e.g., *juice-moose*) are generally appreciated as more similar-sounding than are words that share only onsets (e.g., *juice-june*).

Signed languages exhibit similar privileged groupings. Controlled studies reveal that similarity judgments of signs sharing two sign parameters (i.e., handshape and movement) produced much more consistent judgments than signs sharing only one parameter (Hildebrandt & Corina, 2002). Both deaf and sign-naïve subjects judge signs that share movement and location as the most similar, indicating that this combination of parameters enables a robust perceptual grouping. This fact accords well with the observation that languages commonly capitalize on perceptual distinctions as a basis for linguistic distinctions. For example, theories of the syllable structure of ASL have proposed that the combination of movement and location properties serves as the skeletal structure from which syllables are built, and that movement is the most sonorous element of the sign syllable (see, for example, Sandler, 1989). Interestingly, however, response patterns of the deaf native signers indicate that under some circumstances, tacit knowledge of ASL syllable structure may override purely perceptual factors.

The effects of language experience on the perception of signing are further illustrated in the studies investigating the phenomenon of *categorical perception*. In the spoken language domain, this phenomenon refers to the finding that speech stimuli are perceived categorically rather than continuously despite a continuous variation in form (Liberman, Cooper, Shankweiler, & Studdard-Kennedy, 1967). The principle of categorical perception helps explain how listeners may resolve the many-to-one mapping between continuous acoustic patterns and discrete phonological categories. Initially, the phenomenon of categorical perception was taken as evidence for a hard-wired, language-specific mechanism, but more recent work has placed the phenomenon within the framework of a general psychophysical mechanism that may be observed over a variety of non-language domains.

Studies of categorical perception for phonological parameters in ASL have been conducted for place of articulation, handshape, and facial expressions in signers and non-signers. Recently, categorical perception was found for the parameter of handshape, but not place of articulation (Emmorey, McCullough, & Brentari, 2003), an effect that was limited to deaf signers (see also Baker, Isardi, Golinkoff, & Petitto, 2005). This indicates that linguistic knowledge can influence perceptual judgments.

While it has been demonstrated that hearing non-signers exhibit categorical perception for emotional facial expressions (Calder, Young, Perret, Etcoff, & Rowland, 1996), the study of categorical perception for facial information is interesting in signed languages, as facial information in ASL not only conveys emotional information about the state of the speaker, but also serves a linguistic function. Several well-defined facial configurations serve to signal adverbial modulations and specify syntactic forms. Recent studies have found categorical perception of both emotional and linguistic facial expression (McCullough & Emmorey, 1999; Campbell, Woll, Benson, & Wallace, 1999). However, these effects were observed in both deaf signers and hearing non-signers, suggesting that categorical perception for facial expression is not mediated by linguistic knowledge *per se*.

## 5. SIGN LANGUAGE PRODUCTION

As with the study of spoken languages, psychological processes underlying the *production* of sign language have been underrepresented in the literature. However, examples of “slips of the hand” akin to “slips of the tongue,” have been reported in signed languages. In these data, semantic- and form-based errors are attested, but dual (semantic + phonological) errors are rare. For example, in a report of production errors in German Sign Language (Deutsche Gebaerdensprache [DGS]), 38/40 sign-substitution errors are semantically based; only one is semantically + form-based (Hohenberger, Happ, & Leuninger, 2002). Sign language form-based errors manifest primarily as anticipations and perseverations. Interestingly, there is a marked disparity in the frequency with which the individual sub-lexical form parameters are subject to error, with hand configuration being far more susceptible to error than place of articulation and movement (Hohenberger et al., 2002). These data are consistent with previous slip literature (Klima & Bellugi, 1979; Newkirk, Klima, Pedersen, & Bellugi, 1980) in which the majority of form-based errors are in hand configuration.

Production errors akin to the “tip of the tongue” phenomenon have been observed in sign languages. A recent report of the so-called “Tip of the Finger” (TOF) experiences in deaf signers indicate that, as is observed in studies of spoken language, proper names tend to invoke a large percentage of TOF states (Thompson, Emmorey, & Gollan, 2005). However, as proper names in ASL are nearly always fingerspelled (rather than signed using a lexical sign), it cannot be assessed whether the contributions of TOF state is based upon a failure to retrieve of the English spelling of a proper name or its corresponding fingerspelled form. TOF states do occur in lexical signs (though less frequently), and these

forms provide an unconfounded means by which to assess the properties partial retrieval of lexical forms in signing. These studies indicate that a deaf subject in a TOF state has some access to the phonological composition of an otherwise unaccessible sign. Specifically, it appears that some information regarding the initial configuration of the handshape, location, and orientation is more (and equivalently) reportable, relative to details of the sign movement. Thus similar to spoken language, word onsets appear to have a privileged status on word retrieval in signed languages. The apparent simultaneous accessibility of three of the four major phonological parameters of sign suggests that lexical retrieval may not be guided by a single phonological parameter.

Further evidence of this comes from a sign language production experiment (Corina & Knapp, *in press*) in which the time-course of semantic and phonological form retrieval in sign was assessed. In this paradigm, signers were required to articulate sign names of an object while observing an overlay of a distractor sign. This Stroop-like task found that native signers of ASL exhibit longer naming latencies in the presence of semantically related sign distractors than in the presence of unrelated distractors at early, but not late, stimulus onset asynchronies (SOAs). In contrast, phonologically related stimuli produce naming facilitation both early and late in the naming process, but effects vary by degree and type of phonological relatedness. Interestingly, we observed different amounts of interference between distractors that shared one, two, and three parameters with the target, with greatest effects observed when the sign target to be articulated and the interfering stimulus shared both movement and location parameters. Note that the combination of location and movement was an important factor in the lexical decision experiment of Dye and Shih (*in press*). As discussed previously, location and movement components of signs are linguistic categories that may comprise the skeletal structure of a sign.

## 6. MORPHOLOGY

Morphological theories are formal statements about how word forms are built. Traditional approaches provide accounts of how new words are adopted into the existing lexicon (*i.e.*, coinage), the relationships between word forms that change grammatical category and thus extend meaning usage (*i.e.*, derivational morphology), and how word forms adapt in the face of syntactic phenomena (inflectional morphology).

Psycholinguistic studies of sign language morphology have been largely driven by the fact that morphological processes do not easily fit into a traditional segment-based approach. In contrast to spoken languages in which morphological forms are commonly created by prefixing or suffixing onto a stem, many of the morphological form changes observed in signed languages are non-concatenative. Under morphological inflection, the entire base form undergoes complex movement modulations.

Consider, for example, the morphological changes that affect a verb like TELL. In the citation form of this sign, the extended index finger of the otherwise closed handshape touches the chin with the palm facing the signer. It moves forward in a straight path via

the extension of the elbow to approximately a 120 degree angle. In one aspectual inflection that signals intensity or forcefulness, we find a reduction in elbow extension and an increase in movement velocity. A temporal inflection, giving rise to a meaning that this action occurred again and again over time, adopts an elliptical path movement that passes by the chin in repeated cycles. Inflection for grammatical number agreement may take on a sequence of repeated straight movements directed towards two or three laterally and evenly displaced points in space. Finally, grammatical verb agreement inflections modulate the beginning and end of the path movements which in turn signal the grammatical subject and object of the sentences, respectively.

Many psycholinguistic studies of spoken language morphology have examined whether morphologically complex forms are deconstructed into constituent parts or treated in a more holistic fashion. Current indications are that the degree of decomposability may vary with the formal devices used to modulate word meaning in a specific language. The question of how complex sign forms are parsed has been dealt with in only a handful of studies. Emmorey (1991) investigated the organization of the lexicon for morphologically complex forms using a repetition priming experiment. Here, a morphologically inflected sign was used as a prime, and a lexical decision was made about its uninflected form. These studies found that whereas sign primes inflected for aspect (e.g., habitual, continual) did produce facilitation in the recognition of the associated uninflected stem, signs inflected for grammatical agreement (e.g., dual, multiple, reciprocal) did not prime their citation forms. Emmorey (2002) has suggested that differing degrees of productivity may affect the lexical representation of morphologically complex forms in ASL. The interplay between the number of verb types that participate in a particular inflectional process may influence the association between related forms in the lexicon. For example, more verb forms can be marked with the habitual inflection than the reciprocal inflection.

## 7. SYNTAX

Syntax refers to a level of language form that specifies the interrelationship between words in a sentence. Formal studies of syntax reveal universal and language-specific patterns and constraints that reliably contribute to interpretation of the sentence meaning, especially those specifying thematic roles. Psycholinguistic studies of syntactic processing strive to understand how these structured patterns are exploited during on-line comprehension of sentences, and how this information is combined with word meaning to give rise to a conceptualization of the intended message.

Early spoken language studies examined the psychological reality of clause boundaries and syntactic complexity (Garrett, Bever, & Fodor, 1966). More recent studies have attempted to document the precedence of sentence interpretation as it unfolds in time, with some suggesting a leading role of syntactic over semantic properties. Other studies have examined how structural dependencies in a sentence are instantiated – for example, the integrative processes by which a noun and its associated pronoun are coupled in sentences

like “John went to the store and he bought some bread.” The few studies of syntactic processing in signed languages have focused on the phenomenon of co-reference, exploiting the unusual means by which pronouns and their noun antecedents are associated. Specifically in signed languages like ASL, a signer may associate a noun with a specific location in articulatory space (typically on a horizontal plane in front of the signer) with a later indexical point to this space, signaling a co-referential relationship. Thus in the sentence above, the proper noun *John* may be articulated in a right-sided location in neural space, a subsequent point to that same location would signal the antecedent “John.” In many theories of sign language grammar this pointing sign is considered a pronoun.

Spoken language studies have found that at the time one encounters a pronoun in a sentence one can find evidence that its noun antecedent has been “reactivated.” Using a probe recognition technique, Emmorey and colleagues found evidence for antecedent re-activation during sign comprehension that was similar to that observed for spoken languages. A semantically related probe sign that followed the pronoun was recognized faster than a semantically inconsistent sign. The authors also explored several sign language-specific aspects of co-reference. For example, the re-activation was noted even when the grammatical co-reference was signaled by verb agreement rather than by an overt pronoun, thus providing a psycholinguistic validation of the theoretically motivated analysis of null-pronoun phenomena in ASL (Emmorey & Lillo-Martin, 1995). In addition, Emmorey, Corina, and Bellugi (1995) examined whether the spatial location of the probe interacted with the entailed semantic relationships. They found that the consistency of the spatial location of the probe item did not influence response times. That is, while a semantically (and co-referentially) appropriate probe sign was recognized faster than a semantically inconsistent sign, these recognition times were not modulated by absolute spatial location of the probe (which could appear in the same location as the pronoun [and antecedent] or a different location). Interestingly, this lack of spatial effect was observed only in sentences in which the spatial locations signaled grammatical relations. In contrast, when the spatialization of the nouns and pronouns made reference to actual topographic relationships (i.e., real-world space, such as “chair located to the left of a table”) the consistency between the probes and the spatially established referents did positively influence reaction times. This finding is consistent with a theoretically motivated distinction between grammatical space and topographic use of space in sign. In sum, these studies have found that the same processing mechanisms are required to interpret co-reference in signed and spoken languages, but for signed languages, the type of information represented by the spatial location can influence how co-reference relations are processed (Emmorey, 2002).

## 8. EFFECTS OF SIGN LANGUAGE STRUCTURE ON MEMORY

Classic theories of verbal working memory (Baddeley, 1986) propose that verbal information is maintained in memory via the existence and partial interaction of two components: a phonological buffer in which phonological representations of words are transiently stored, and an articulatory rehearsal mechanism through which those

representations are continuously updated. This proposed architecture of verbal working memory accounts for four well-attested phenomena associated with spoken language immediate serial recall (discussed below) – the phonological similarity effect, the irrelevant stimulus effect, the word length effect, and the articulatory suppression effect (reviewed in Baddeley, 1990; Neath, Surprenant, & LeCompte, 1998; Wilson & Emmorey, 1997a).

However, it is not *a priori* obvious whether this architecture can account for working memory for sign language signs. While signs clearly have phonological structure that is amenable to verbal storage and rehearsal mechanisms, their visual-manual nature likely requires processing routines that draw heavily from those subserving non-linguistic visuo-spatial working memory. In recent years, a great deal of knowledge about visuo-spatial working memory for sign language signs has come from the studies of Margaret Wilson and colleagues (e.g., Wilson & Emmorey, 1997a; Wilson, 2001a). These studies demonstrate that the similarities between spoken and sign working memory functions are profound, encompassing each of the phenomena on which classic working models are predicated.

One of the four effects that form the foundation of classic working memory models is that serial recall is poorer for words situated in the context of phonologically similar words (e.g., dog, log, lot, dot) than with words with different phonological representations. This phonological similarity effect (for a review see Gathercole & Baddeley, 1996) has been taken as evidence that the code by which words are stored in working memory is phonological in form. An analogous effect has been found in sign language serial recall, in which performance is worsened for signs that share phonological parameters, such as hand configuration, with other signs in a signed list (Hanson, 1982; Klima & Bellugi, 1979). That word similarity effects exist in signed languages is evidence that signs are stored in working memory by a phonological or kinesthetic code, just as spoken words. That this code is visual–manual rather than auditory–oral, however, is evidence that the entries in verbal working memory, no matter the architecture of the system, are themselves modality-specific representations.

Parallel patterns of serial recall disruption in speech and sign are also evidenced by performance decrements in the presence of irrelevant, modality-specific stimuli. Specifically, hearing subjects' spoken word recall performance is known to be impaired when they are presented with irrelevant auditory stimuli, such as speech or non-linguistic tones, immediately following the presentation of a list of words to be remembered. This irrelevant sound effect (for a review see Neath, 2000) occurs when the words to be remembered are either auditory or printed. However, in contrast to findings from studies of the phonological similarity effect, effective interfering stimuli for speech recall are seldom or never visual (Wilson & Emmorey, 2003), presumably because the interfering visual stimuli used in irrelevant speech tasks are not coded into an auditory form capable of causing interference. In contrast, Wilson and Emmorey (2003) have found that irrelevant visual stimuli do reduce *sign* recall performance. Interestingly, both pseudo-signs and non-linguistic moving visual objects were markedly effective in doing so. This finding is consistent with a view of sign language working memory in which signs are stored as visual–manual phonological representations, and is further evidence of a common



principle (irrelevant stimulus effects) being instantiated in a modality-specific way (auditory for speech, and visual for sign).

Both the phonological similarity effect and the irrelevant stimulus effect are associated with the buffer component of the classic verbal working memory model. In contrast, the rehearsal component is associated with two different effects, the word length effect and articulatory suppression. The word length effect is the finding that short words are remembered better than longer words on immediate serial recall tasks. Generally this is believed to be a result of the greater amount of time required to articulate (and thus rehearse) longer words, although it is controversial as to whether the critical factor is actually the number of syllables or phonemes in a given word, rather than the length of the word *per se* (Baddeley, Thomson, & Buchanan, as cited in Wilson, 2001a). By varying the physical distances that signs must traverse during articulation, Wilson and Emmorey (1998) were able to demonstrate that greater articulatory time, rather than a greater number of phonological units, is indeed responsible for the word length effect in sign languages. This study suggests that although spoken and signed languages call upon different articulators and motor programs in the service of word production, the processes associated with retaining words and signs in working memory are constrained by common properties.

Effects of articulation on memory span are not confined to word length. Span is also reduced when subjects engage in non-speech articulatory behaviors, while holding a list of words in memory. This finding, known as the articulatory suppression effect (Murray, 1968; Smith, Reisburg, & Wilson, 1992), has been shown by Wilson and Emmorey (1997b) to be robust for sign as well as speech. Specifically, signers' immediate serial recall performance was reduced when they opened and closed their hands (from an S hand to a 5 hand and back) repeatedly during stimulus presentation. The articulatory suppression effect is a testament to the importance of modality-specific articulatory rehearsal in verbal working memory. By engaging in an unrelated motoric task, planning and/or implementation of the motor program normally used to rehearse and produce a spoken or signed word is suppressed, and articulatory rehearsal is no longer possible. As a consequence, memory performance suffers.

Broadly, the functional architecture of working memory for spoken and signed language thus appear remarkably similar. Each has a storage buffer in which lexical items are represented by phonological codes – codes that are susceptible to interference both by the concurrent presence of highly similar phonological codes, and by the presence of irrelevant stimuli that share the modality – auditory or visual – of the words to be remembered. The lexical items in the phonological buffer, spoken or signed, decay unless sub-vocally or sub-manually rehearsed, a process that is itself sensitive to the length of the words to be remembered. Perhaps most interestingly, blocking articulatory rehearsal will not only result in worsened memory performance, it will suppress other effects that it gates, such as the phonological similarity effect (Baddeley, Lewis, & Vallar, 1984; Wilson & Emmorey, 1997a, 1997b). Articulation thus serves as a mechanism for getting words into the phonological buffer, and for keeping them active while there.

## 9. NEURAL REPRESENTATION OF SIGNED LANGUAGES

Our understanding of the neural representation of human language has been greatly enriched by the consideration of signed language of the deaf. Outwardly, this language form poses an interesting challenge for theories of cognitive and linguistic neural specialization, which classically has regarded the left hemisphere as being specialized for linguistic processing, while the right hemisphere is specialized for visual-spatial abilities. Given the importance of putatively visual-spatial properties of sign forms (e.g. movements trajectories and paths through 3-dimension space, memory for abstract spatial locations, assessments of location and orientation of the hands to the body, etc.), one might expect a greater reliance of right-hemisphere resources during sign language processing. However, as discussed above, despite major differences in the modalities of expression, once we acknowledge the structural homologues of spoken and signed language forms, striking parallels in the psycholinguistic and cognitive processing of these languages emerge. Thus, these commonalities suggest a possible uniformity in neural systems, and the cognitive processes they mediate, underlying both signed and spoken processing.

Case studies of deaf signing individuals with acquired brain damage and neuroimaging studies of healthy deaf subjects have provided confirming evidence for the importance of left hemisphere system in the mediation of signed language. Deaf signers, like hearing speakers, exhibit language disturbances when left-hemisphere cortical regions are damaged (e.g., Hickok, Love-Geffen, & Klima 2002; Marshall, Atkinson, Smulovitch, Thacker, & Woll, 2004; Poizner, Klima, & Bellugi, 1987; for a review see Corina, 1998a, 1998b). In addition, there is good evidence that *within* the left hemisphere, cerebral organization in deaf signers follows the familiar anterior/posterior dichotomy for language production and comprehension, respectively, that is familiar from speech. In addition, neuroimaging studies have raised new questions regarding the unique role of the right hemisphere in sign language comprehension, as some evidence suggests that posterior-parietal regions may play a special role in the mediation of signed languages (Newman, Bavelier, Corina, Jezzard, & Weville, 2002).

## 10. SIGN LANGUAGE APHASIA

### 10.1. Sign Language Production

In spoken language aphasia, chronic language production impairments are typically associated with left-hemisphere frontal anterior lesions that involve the cortical zone encompassing the lower posterior portion of the left frontal lobe, e.g., Broca's area. These lesions often extend in depth to the periventricular white matter (e.g., Mohr et al., 1978; Goodglass, 1993). The anterior insula has also been implicated in chronic speech production problems (Dronkers, Redfren, & Knight, 2000). In the context of understanding prefrontal contributions to sign language, a pertinent example of left-hemisphere language mediation is that of patient G.D., reported in Poizner et al. (1987). G.D., a deaf signer with a large lesion in a left, anterior frontal region encompassing BA 44/45, presented with non-fluent, aphasic signing with intact sign comprehension. Specifically,

G.D.'s signing was effortful and dysfluent, with output often reduced to single-sign utterances. The signs she was able to produce were agrammatic, devoid of the movement modulations that signal morpho-syntactic contrasts in fluent signing. As with hearing Broca's aphasics, this signer's comprehension of others' language productions was undisturbed by her lesion. Both at the single sign and sentence level, her comprehension was on par with control subjects. That this deficit is not simply motoric in nature is indicated by the fact that the deficits were exhibited on both her motorically and non-motorically (i.e., ipsilesional) limb.

## 10.2. Sign Paraphasia

Sign language breakdown following left hemisphere damage is not haphazard, but affects independently motivated linguistic categories. This observation provides support for viewing aphasia as a unique and specific cognitive deficit rather than as a subtype of a more general motor or symbolic deficit. A fascinating example of the systematicity in sign and spoken language breakdown is illustrated through consideration of paraphasia errors (Corina, 2000). The substitution of an unexpected word for an intended target is known as verbal paraphasia. Most verbal paraphasias have a clear semantic relationship to the desired word and represent the same part of speech, hence, they are referred to as "semantic paraphasias" (Goodglass, 1993). In contrast, phonemic or "literal" paraphasia refers to the production of unintended sounds or syllables in the utterance of a partially recognizable word (Blumstein, 1973; Goodglass, 1993). Theoretically, sound distortions arising from phonetic impairment are not considered to be instances of paraphasia; however, in practice, it is quite difficult to distinguish true paraphasic errors from phonetic-based sound distortions. Phonemic sound substitution may result in another real word, related in sound but not in meaning (e.g., *telephone* becomes *television*). Also attested are cases in which the erroneous word shares both sound characteristics and meaning with the target (*broom* becomes *brush*; Goodglass, 1993).

Several reports of signing paraphasia can be found in the sign aphasia literature. In an early report of "neologistic" signing, Leischner (1943) describes a deaf subject with left hemisphere damage who produced "fluent but nonsensical signing." Unfortunately, little description of these errors was provided. Several well-documented examples of semantic paraphasias have been reported (Poizner et al., 1987; Brentari, Poizner, & Kegl, 1995; Corina et al., 1992). For example, subject P.D. (Poizner et al., 1987) produced clear lexical substitutions: BED for CHAIR, DAUGHTER for SON, QUIT for DEPART, etc. In general, the semantic errors of P.D. overlap in meaning and lexical class with the intended targets; this pattern has been routinely observed in spoken language semantic paraphasia. Subject W.L. (Corina et al., 1992) evidenced interesting semantic blends in signing, errors conditioned, in part, by perseverations from earlier cued items. For example, in the context of a picture-naming task, when shown a picture of a tree, W.L. signed TREE with the G handshape. Previously, W.L. had been asked to name the color green. The lexical signs GREEN and TREE share a motion (twisting of the wrist) and evidence similar articulatory postures. These ASL semantic paraphasias suggest that the lexicon is structured according to semantic principles, whereby similar semantic

items share representational proximity. In this view, co-activation of closely related representations and/or an absence of appropriate inhibition from competing entries may lead to substitutions and blends.

One of the most striking characteristics of aphasic signing is formational paraphasia. As with spoken languages, ASL formational errors encompass both phonological and phonetic levels of impairment (see Corina, 2000, for some discussion). A priori, we may expect to find true phonological errors affecting the four major formational parameters of ASL phonology: handshape, movement, location, and orientation. However, the distribution of paraphasic errors among the four parameters of sign formation appears to be unequal; handshape configuration errors are the most widely reported, while paraphasias affecting movement, location, and orientation are infrequent (see, e.g., Poizner et al., 1987). Not only have handshape errors been observed across many different aphasic signers, but the frequency of occurrence in individuals who exhibit this disruption is quite high. The globally aphasic signer W.L. (Corina et al., 1992) produced numerous phonemic errors, nearly all of which were errors involving handshape specification. For example, W.L. produced the sign TOOTHBRUSH with the Y handshape rather than the required G handshape, and produced the sign SCREWDRIVER with an A handshape rather than the required H handshape. The higher incidence of handshape errors is interesting, as recent linguistic analyses of ASL have suggested that handshape specifications (and perhaps static articulatory locations) may be more consonantal in nature, while movement components of ASL may be analogous to vowels (see Corina & Sandler, 1993, for some discussion). In spoken language phonemic paraphasias, a homologous asymmetry exists; the vast majority of phonemic paraphasias involve consonant distortions. Another similarity between spoken and sign paraphasic error is that in each case, errors do not compromise the syllabic integrity of a sign or word (Brentari et al., 1995; Corina, 2000).

### 10.3. Sign Language Comprehension

Fluent spoken language aphasias are associated with lesions to left-hemisphere posterior temporal regions. Wernicke's aphasia, for example, is often associated with damage to the posterior regions of the left superior temporal gyrus. More recent work has suggested the contribution of posterior middle temporal gyrus in cases of chronic Wernicke's aphasia (Dronkers, Redfern, & Ludy, 1995; Dronkers et al., 2000). Two prominent features of Wernicke's aphasia are impaired comprehension and fluent, but often paraphasic (semantic and phonemic) output. Additionally, persistent neologistic output sometimes occurring with severe Wernicke's aphasia is associated with lesions extending to the supramarginal gyrus (Kertesz, 1993).

Signers with left-hemisphere posterior lesions also evidence fluent sign aphasia. Two cases are reported in Chiarello, Knight, and Mandell (1982), Poizner et al. (1987), and Corina et al. (1992). These patients presented with severe comprehension difficulties in the face of relatively fluent but paraphasic output. Interestingly, while damage to left temporal cortex has been demonstrated to impair sign comprehension in some patients (Hickok et al., 2002), the lesions in the two case study patients above did not occur in

cortical Wernicke's area proper, but rather involved more frontal and inferior parietal areas. In both cases, lesions extended posteriorly to the supramarginal gyrus. This is interesting, as lesions associated with the supramarginal gyrus alone in users of spoken language do not typically result in severe speech comprehension deficits. These two cases suggest that sign language comprehension may be more dependent than speech on left-hemisphere inferior parietal areas, a difference that may reflect within-hemisphere reorganization for cortical areas involved in sign comprehension (Leischner, 1943; Chiarello et al., 1982; Poizner et al., 1987).

#### **10.4. Cortical Stimulation Mapping**

Additional insights into the neural control of paraphasic errors have been reported by Corina et al. (1999), who investigated sign language production in a deaf individual undergoing an awake CSM procedure for the surgical treatment of epilepsy. During the language mapping portion of the procedure, a subject is required to name pictures or read written words. Disruption of the ability to perform the task during stimulation is taken as evidence of cortical regions integral to the language task (Stemmer & Whitaker, 1998).

In this deaf patient, all testing was conducted using ASL. The patient was to sign the names of line drawings of pictures. All signs were one-handed, and the subject signed each with his left hand. Because this subject was undergoing left-hemisphere surgery, language disruption as a result of cortical stimulation cannot be attributed to the suspension of primary motor functioning.

Stimulation to two anatomical sites led to consistent naming disruption. One of these sites, an isolated frontal opercular site, corresponds to the posterior aspect of Broca's area, BA 44. A second site, located in the parietal opercular region, also resulted in robust object-naming errors. This parietal area corresponds to the supramarginal gyrus (SMG, BA 40). Importantly, the nature of these errors was qualitatively different. Stimulation of Broca's area resulted in errors involving the motor execution of signs. These errors are characterized by a laxed articulation of the intended sign, with non-specific movements (repeated tapping or rubbing) and a reduction in handshape configurations to a laxed-closed fist handshape. Interestingly, there was no effort on the part of S.T. to self-correct these imperfect forms. Our results are consistent with the characterization of the posterior portion of Broca's area as participating in the motoric execution of complex articulatory forms, especially those underlying the phonetic level of language structure.

The sign errors observed with stimulation of the SMG are qualitatively different. With stimulation to this site, S.T. produced both formational and semantic errors. Formational errors are characterized by repeated attempts to distinctly articulate the intended targets, commonly with successive formational approximations of the correct sign. For example, the sign PEANUT is normally signed with a closed fist and outstretched thumb with a movement composed of an outward wrist rotation (the thumb flicking off the front

of the teeth). Under stimulation, this sign began as an incorrect but clearly articulated, “X” handshape (closed fist with a protruding bent index finger) produced at the correct location, but with an incorrect inward rotation movement. In two successive attempts to correct this error, the subject first corrected the handshape, and then went on to correct the movement as well. Notably, we do not find the laxed and reduced articulations characteristic of signing under conditions of stimulation to Broca’s area. Instead, as these examples illustrate, under stimulation to the SMG, the subject’s signing exhibits problems involving the selection of the individual components of sign forms (i.e., handshape, movement, and, to a lesser extent, location).

Semantic errors were also observed under stimulation of the SMG, and the characteristics of these errors are particularly noteworthy. Specifically, all of these errors involve semantic substitutions that are formationally quite similar to the intended targets. For example, the stimulus picture “pig” elicited the sign FARM; the stimulus picture “bed” was signed as SLEEP; and the stimulus picture “horse” was signed as COW. In ASL, these semantic errors contain considerable formational overlap with their intended targets. For example, the signs PIG and FARM differ in movement, but share an identical articulatory location (the chin). Each is made with a similar handshape; the signs BED and SLEEP share handshape and are both articulated about the face; finally, the signs COW and HORSE differ only in handshape. In English, these mistakes might be similar to uttering “lobster” when one intended to say “oyster,” or “plane” when one intended to say “train”. That is, these errors share both semantic and formational properties.

In summary, the analysis of these data suggests that stimulation to Broca’s area has a global effect on the motor output of signing, whereas stimulation to parietal opercular site (the SMG) disrupts the correct selection of the linguistic components (including both phonological and semantic elements) required in the service of naming.

## 11. NEUROIMAGING STUDIES

Neuroimaging techniques like PET and fMRI also make unique contributions to our current understanding of the neurological processing of signs. In particular, these studies reaffirm the importance of left-hemisphere anterior and posterior brain regions for sign language use and emphasize that some neural areas appear to participate in language perception and production, regardless of the modality of the language.

Sign language production tasks are especially likely to recruit the left hemisphere. For example, when signers name objects (Emmorey et al., 2003), generate verbs to accompany nouns (e.g., CHAIR → SIT) (McGuire et al., 1997; Petitto, Zatorre, Gauna, Nikelski, Dostie, & Evans, 2000; Corina, San Jose-Robertson, Guillemin, High, & Braun, 2003), or sign whole sentences (Braun, Guillemin, Hosey, & Varga, 2001), their left hemispheres show significant increases in blood flow, relative to control tasks. It has been suggested that this heightened blood flow reflects, in part, the activation of motor systems needed for the production of complex linguistic actions.

Sign language comprehension also recruits the left hemisphere in some studies, for both word- and sentence-level tasks. For example, classic Broca's area has been found to be involved in sign comprehension when subjects observe single signs (Levanen, Uutela, Salenius, & Hari, 2001; Petitto et al., 2000) and sentences (Neville et al., 1998; MacSweeney et al., 2002). This activation is not limited to anterior regions. When signers of BSL view their language, posterior left-hemisphere regions are activated, including the posterior superior temporal gyrus and sulcus, and the supramarginal gyrus (MacSweeney et al., 2004). This heightened activation is relative to complex non-linguistic gestures, and does not occur for non-signers.

Interestingly, there is growing evidence that right-hemisphere regions may also be recruited for aspects of sign-language processing in ways that are not required in the processing of spoken languages. At least one sign-language production task is known to recruit right-hemisphere brain regions. When deaf signers were asked to use classifier constructions to describe the relative positions of two objects depicted in a line drawing, both left- and right-hemisphere regions were found to be active (Emmorey et al., 2002). When ASL prepositions were used instead of classifiers, only the right hemisphere was recruited.

Other evidence suggests that right-hemisphere posterior parietal regions may contribute to the processing of some aspects of sign comprehension (Bavelier et al., 1998; Capek et al., 2004; Corina, 1998b; Newman et al., 2002). For instance, both left- and right-hemisphere cortical regions were recruited when hearing native signers of ASL passively watched ASL sentences (Newman et al., 2002). Some right-hemisphere structures appear to be specialized for processing spatial information, including biological motion. It may be that ASL phonological distinctions that make use of space are the trigger for right-hemisphere recruitment in sign perception.

Moreover, right-hemisphere involvement may be related to the age at which the signer first acquired a sign language. One particular structure, the right angular gyrus, was found to be active only when hearing native users of ASL performed the task. When hearing signers who learned to sign after puberty performed the same task, the right angular gyrus failed to be recruited. Thus, the activation of this neural structure during sign language perception may be a neural "signature" of sign competence developing during the critical period for language (Newman et al., 2002).

In sum, while left-hemisphere regions are undisputedly recruited in a similar fashion for both sign and speech, it has been argued that the right-hemisphere activation seen during sign language comprehension is more robust than that observed in studies of spoken language processing. Continued research using cognitive neuroscience tools such as PET and fMRI will provide more opportunities to investigate these findings.

In conclusion, preliminary findings from psycholinguistic and cognitive studies of signed languages indicate a great deal of commonality between the recognition, access, and memory structures of the representations of spoken and signed words in the mental lexicon. These similarities extend to those processes underlying the parsing and

interpretation of morphological and syntactic structure. Moreover, burgeoning cognitive neuroscience research has also begun to specify the commonalities and differences between neural systems underlying spoken and signed language forms. While some unique and subtle differences have been found between their respective processing, overwhelming evidence suggest a great deal of homology between speech and sign neural representations. These data suggests that core properties underlie the neural capacity for human language, regardless of the surface form taken by the linguistic communicative system. Future studies will doubtless continue to further specify both the common and unique aspects of these forms of language while striving to better understand the processing domains under which the modality of language expression does and does not affect the final form of the mental structure of language. In this way, we may come to understand both the biological and environmental contributions that shape human linguistic communication.

### ACKNOWLEDGEMENTS

This work was supported by NIH-NIDCD grant R01DC003099-06 to David P. Corina and by pre-doctoral National Research Service Award (NRSA) 5F31DC006796-02 to Heather P. Knapp.

### REFERENCES

- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36A, 233–252.
- Baker, S. A., Isardi, W. J., Golinkoff, R. M., & Petitto, L. (2005). The perception of handshapes in American Sign Language. *Memory and cognition*, 33, 884–904.
- Battison, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.
- Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., Rauschecker, J. P., Braun, A., Turner, R., & Neville, H. J. (1998). Hemispheric specialization for English and ASL: Left invariance-right variability. *Neuroreport*, 9 (7), 1537–1542.
- Blumstein, S. E. (1973). *A phonological investigation of aphasic speech*. The Hague: Mouton.
- Braun, A. R., Guillemin, A., Hosey, L., & Varga, M. (2001). The neural organization of discourse an H<sub>2</sub>O-Pet study of narrative production in english and American Sign Language. *Brain*, 124 (10), 2028.
- Brentari, D. A. (1998). *Prosodic model of sign language phonology*. Cambridge, MA: MIT Press.
- Brentari, D. (2001). *Foreign vocabulary in sign languages: A cross-linguistic investigation of word formation*. Mahwah, NJ: Lawrence Erlbaum Associates.



- Brentari, D., Poizner, H., & Kegl, J. (1995). Aphasic and parkinsonian signing: Differences in phonological disruption. *Brain and Language*, 48(1), 69–105.
- Calder, A. J., Young, A. W., Perret, D. I., Ectoff, N. L., & Rowland, D. (1996). Early experience is associated with the development of categorical perception. *Visual Cognition*, 3, S1–S117.
- Campbell, R., Woll, B., Benson, P. J., & Wallace, S. B. (1999). Categorical perception of face actions: Their role in sign language and in communicative facial displays. *Quarterly Journal of Experimental Psychology*, 52A(1), 67–95.
- Capek, C. M., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P., & Neville, H. J., (2004). The cortical organization of audio–visual sentence comprehension: An fMRI study at 4 tesla. *Brain Research Cognitive Brain Research*, 20(2), 111–119.
- Carreiras, M., Gutierrez-Sigut, E., Baquero, S., & Corina, D. P. (submitted) Phonological similarity effects in Spanish sign language (LSE).
- Chiarello, C., Knight, R., & Mandell, M. (1982). Aphasia in a prelingually deaf woman. *Brain*, 105, 29–51.
- Corina, D. P. (1998a). The processing of sign language: Evidence from aphasia. In: H. Whitaker, & B. Stemmer (Eds), *Handbook of neurology* (pp. 313–329). San Diego, CA: Academic Press.
- Corina, D. (1998b). Studies of neural processing in deaf signers: Toward a neurocognitive model of language processing in the deaf. *Journal of Deaf Studies and Deaf Education*, 3(1), 35–48.
- Corina, D. P. (2000). Some observations regarding paraphasia in American sign language. In: K. Emmorey, & H. Lane (Eds), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 493–507). Mahwah, NJ: Lawrence Erlbaum Associates.
- Corina, D. P., & Emmorey, K. (1993). Lexical priming in American sign language. Paper presented at the Linguistic Society of American Conference, Philadelphia, PA.
- Corina, D. P., & Hildebrandt, U. C. (2002). Psycholinguistic investigations of phonological structure in ASL. [References]. In: M. P. Richard, & K. Cormier et al. (Eds) *Modality and structure in signed and spoken languages*, (pp. 88–111). New York: Cambridge University Press.
- Corina, D. P., & Knapp, H. P. (2006). Lexical retrieval in American sign language production. In: L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds), *Papers in laboratory phonology 8: Varieties of phonological competence*. Berlin: Mouton de Gruyter.
- Corina, D. P., McBurney, S. L., Dodrill, C., Hinshaw, K., Brinkley, J., & Ojemann, G. (1999). Functional roles of Broca's area and smg: Evidence from cortical stimulation mapping in a deaf signer. *Neuroimage*, 10(5), 570–581.
- Corina, D. P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., & O'Grady-Batch, L. (1992). Dissociation between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Lang*, 43(3), 414–447.

- Corina, D., & Sandler, W. (1993). On the nature of phonological structure in sign language. *Phonology*, *10*, 165–207.
- Corina, D. P., San Jose-Robertson, L., Guillemin, A., High, J., & Braun, A. R. (2003). Language lateralization in a bimanual language. *Journal of Cognitive Neuroscience*, *15* (5), 718–730.
- Dronkers, N. F., Redfern, B. B., & Knight, R. T. (2000). The neural architecture of language disorders. In: M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*, Cambridge: The MIT Press, (2nd ed., pp. 949–960).
- Dronkers, N. F., Redfern, B. B., & Ludy, C. A. (1995). Lesion localization in chronic Wernickes aphasia. *Brain and Language*, *51*, 62–65.
- Dye, M., & Shih, K. (2006). Mock attitudes and old chairs: Phonological priming in British sign language. In: L. Goldstein, D. W. Whalen, & C. T. Best (Eds), *Papers in laboratory phonology 8: Varieties of phonological competence*. Berlin: Mouton de Gruyter.
- Emmorey, K. (1991). Repetition priming with aspect and agreement morphology in American Sign Language. *Journal of Psycholinguistic Research*, *20*(5), 365–388.
- Emmorey, K. (2002). *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Lawrence Erlbaum and Associates.
- Emmorey, K., & Corina, D. (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills*, *71*, 1227–1252.
- Emmorey, K., Corina, D., & Bellugi, U. (1995). Differential processing of topographic and referential functions of space. In: K. Emmorey, & J. Reilly (Eds). *Language, gesture, and space* (pp. 43–62). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Emmorey, K., Damasio, H., McCullough, S., Grabowski, T., Ponto, L. L., Hichwa, R. D., & Bellugi, U. (2002). Neural systems underlying spatial language in American sign language. *Neuroimage*, *17*(2), 812–824.
- Emmorey, K., Grabowski, T., McCullough, S., Damasio, H., Ponto, L. L., Hichwa, R. D., & Bellugi, U. (2003). Neural systems underlying lexical retrieval for sign language. *Neuropsychologia*, *41*(1), 85–95.
- Emmorey, K., & Lillo-Martin, D. (1995). Processing spatial anaphora: Referent reactivation with overt and null pronouns in American sign language. *Language and Cognitive Processes*, *10*(6), 631–664.
- Emmorey, K., McCullough, S., Brentari, D. (2003). Categorical perception in American sign language. *Language & Cognitive Processes*, *18*(1), 21–45.
- Garrett, M., Bever, T., Fodor, J. (1966). The active use of grammar in speech perception. *Perception & Psychophysics*, *1*(1), 30–32.
- Gathercole S. E., & Baddeley, A. (1996). *Working memory and language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Goodglass, H. (1993). *Understanding aphasia*. San Diego: Academic Press.
- Hanson, V. L. (1982). Short-term recall by deaf signers of American sign language: Implications of encoding strategy for order recall. *Journal of Experimental Psychology: Learning, memory, and cognition*, 8(6), 572–583.
- Hickok, G., Love-Geffen, T., & Klima, E. S. (2002). Role of the left hemisphere in sign language comprehension. *Brain and Language*, 82 (2), 167–178.
- Hildebrandt, U. C., & Corina, D. P. (2002). Phonological similarity in American sign language. *Language and Cognitive Processes*, 6, 593–612.
- Hohenberger, A., Happ, D., & Leuninger, H. (2002). Modality-dependent aspects of sign language production: Evidence from slips of the hands and their repairs in German sign language. In: R. P. Meier, K. Cormier, & D. Quinto-Pozos (Eds). *Modality and structure in signed and spoken language*. Cambridge: Cambridge University Press.
- Jusczyk, P. W., & Luce, P. A. (2002). Speech perception. In: Pashler, H., & Yantis, S. (Eds), *Steven's handbook of experimental psychology: Vol. 1: Sensation and perception*. (3rd ed., pp. 493–536). New York: Wiley.
- Kertesz, A. (1993). Clinical forms of Aphasia. *Acta Neurochirurgica Supplementum*, 56, 52–58.
- Klima, E., & Bellugi U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.
- Leischner, A., (1943). Die “Aphasie” der Taubstummen. *Archiv fuer Psychiatrie und Nervenkrankheiten*, 115, 469–548.
- Levanen, S., Uutela, K., Salenius, S., & Hari, R. (2001). Cortical representation of sign language: Comparison of deaf signers and hearing non-signers. *Cerebral Cortex*, 11, 506–512.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., Calvert, G., & Brammer, M. J. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *Neuroimage*, 22(4), 1605–1618.
- MacSweeney, M., Woll, B., Campbell, R., Calvert, G. A., McGuire, P. K., David, A. S., Simmons, A., & Brammer, M. J. (2002). Neural correlates of British sign language comprehension: Spatial processing demands of topographic language. *Journal of cognitive neuroscience*, 14, 1064–1075.
- Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., & Woll, B. (2004). Aphasia in a user of British sign language: Dissociation between sign and gesture. *Cognitive Neuropsychology*, 21(5), 537–554.

- McCullough, S., & Emmorey, K., (1999). Perception of emotional and linguistic facial expressions: A categorical perception study with deaf and hearing subjects. Poster presented at the Psychonomics Society Meeting, November, Los Angeles, California.
- McGuire, P. K., Robertson, D., Thacker, A., David, A. S., Kitson, N., Frackowiak, R. S. J., & Frith, C. D. (1997). Neural correlates of thinking in sign language. *NeuroReport*, 8(3), 695.
- Mohr, J. P., Pessin, M. S., Finkelstein, H. H., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: Pathologic and clinical. *Neurology*, 28(4), 311–324.
- Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78(4), 679–684.
- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, 7(3), 403–423.
- Neath, I., Surprenant, A. M., & LeCompte, D. C. (1998). Irrelevant speech eliminates the word length effect. *Memory & Cognition*, 26(2), 343–354.
- Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., Braun, A., Clark, V., Jezzard, P., & Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proceedings of the National Academy of Sciences*, 95, 922–929.
- Newkirk, D., Klima, E. S., Pedersen, C. C., & Bellugi, U. (1980). Linguistic evidence from slips of the hand. In: V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 165–198). New York: Academic Press.
- Newman, A. J., Bavelier, D., Corina, D., Jezzard, P., & Neville, H. J. (2002). A critical period for right hemisphere recruitment in American Sign Language processing. *Nature Neuroscience*, 5(1), 76–80.
- Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., & Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: Implications for the neural basis of human language. *Proceedings of the National Academy of Sciences of the USA*, 97(25), 13961–13966.
- Poizner, H., Klima, E. S., & Bellugi, U. (1987). *What the hands reveal about the brain. MIT press series on issues in the biology of language and cognition*. Cambridge, MA: The MIT Press.
- Prillwitz, S., Leven, R., Zienert, H., Hanke, T., & Henning, J. HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide. (International Studies on Sign Language and Communication of the Deaf, 5) Hamburg: Signum. Retrieved December, 2004 from <http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html>.
- Sandler, W. (1989). *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language. Publications in language sciences*, 32. Providence, RI: Foris Publications.

- Smith, J. D., Reisberg, D., & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In: D. Reisberg (Ed.), *Auditory imagery* (pp. 95–119). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stemmer, B., & Whitaker, H. A. (1998). *Handbook of neurolinguistics*. San Diego, CA: Academic Press.
- Stokoe, W., Casterline, D. C., & Croneberg, C. G. (1976). *A dictionary of American Sign Language on linguistic principles*. Washington, DC: Gallaudet University Press.
- Thompson, R., Emmorey, K., & Gollan, T. H. (2005). Tip of the fingers: Experiences by deaf signers. *Psychological Science*, *16*(11), 856–860.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, *9*, 325–329.
- Wilson, M. (2001a). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review*, *8*(1), 44–57.
- Wilson, M. (2001b). The impact of sign language expertise on perceived path of apparent motion. In: M. D. Clark, & M. Marschark (Eds), *Context, cognition and deafness* (pp. 28–48). Washington, DC: Gallaudet University Press.
- Wilson, M., & Emmorey, K. (1997a). Working memory for sign language: A window into the architecture of the working memory system. *Journal of Deaf Studies and Deaf Education*, *2*, 121–130.
- Wilson, M., & Emmorey, K. (1997b). A visuospatial ‘phonological loop’ in working memory: Evidence from American Sign Language. *Memory & Cognition*, *25*, 313–320.
- Wilson, M., & Emmorey, K. (1998). A ‘word length effect’ for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition*, *26*, 584–590.
- Wilson, M., & Emmorey, K. (2003). The effect of irrelevant visual input on working memory for sign language. *Journal of Deaf Studies and Deaf Education*, *8*, 97–103.

## **SECTION 3**

### **LANGUAGE DEVELOPMENT**

This page intentionally left blank

## Chapter 27

### Language Learning in Infancy

*Anne Fernald and Virginia A. Marchman*

#### 1. INTRODUCTION

Learning language is one of the most impressive and intriguing human accomplishments. Think about the vast differences between the healthy 12-month-old child who says “*Ah! Ah!*” with hands held up in the air, eager to be lifted from the highchair, and the same child six months later using recognizable two-word combinations coordinated with gestures (e.g., “*Mommy out!*”). Within the next year that child will start using an impressive complement of morphosyntactic skills to produce utterances that reflect considerable linguistic sophistication (e.g., “*Mom! I wanna get outta this chair now!*”). The child’s desire may be equally intense in each of these situations, yet clearly the typical 2-year-old has advanced significantly beyond the 1-year-old in her ability to effectively use her native language to make that particular desire known to those around her. The goal of developmental psycholinguistics is to map the endogenous and exogenous forces that converge to shape and guide this set of developmental achievements.

Over the past five decades, the field of language development research has been at the center of the debate between nativist and constructivist approaches to understanding human cognition. In the early 1960s, Chomsky’s proposal that language acquisition was innately guided by a Language Acquisition Device offered a powerful solution to the logical problem of how children learn language, a view still ardently embraced by many in the field. Since that time, however, an alternative view has been gaining momentum, gathering logical and empirical support for the idea that a child’s linguistic knowledge is constructed rather than triggered, emerging as a consequence of the child’s experiences with the linguistic and non-linguistic world (e.g., Bates & MacWhinney, 1979; Braine, 1976; Slobin, 1973). Proposals on how exactly the child accomplishes this task have taken several different forms over the years, and with each new decade, theoretical and methodological advances have strengthened the case for this alternative to nativism. The goal of this chapter is to outline some key features of current proposals on how the child constructs a language. We first briefly review the standard nativist approach, and then discuss some recent developments in theory and research from diverse disciplines



that have contributed to a shift in emphasis in research on language development in productive new directions. Then we review three lively areas of current research on language learning in infancy: early speech perception, lexical development, and listening for meaning. This review will of necessity be quite eclectic, focusing on a few studies within each of these areas that exemplify new perspectives that are now coming to the forefront in this field.

## 2. NATIVIST VIEWS OF LANGUAGE DEVELOPMENT

How and why does the child's linguistic behavior change so dramatically over such a short period of time? Much of the research on this topic has assumed that this process is driven by innate and highly specialized mental structures (a "mental organ," Chomsky, 1981). That is, learning language involves the operation of a specifically linguistic maturational bioprogram (i.e., Universal Grammar) as it processes specifically linguistic input (e.g., Chomsky, 1981; Lenneberg, 1967). According to this perspective, this innately specified system is what makes it possible for the child to determine which of all the possible linguistic rule systems characterizes their particular native language. Indeed, the goal of much research in modern linguistics has been to map the diverse set of principles and features that describe the rule systems of any and all of the world's languages. Of course, in setting out on this daunting task, one is soon struck by the vast richness and complexity of human grammars.

If the central question is how grammars come to be mastered, such complexity is particularly disheartening, especially in light of the prevailing assumption among nativist theorists that the environment falls far short of providing what children need in order to learn rich systems of linguistic representations on their own. According to Chomsky (1981), speech by adults is so full of hesitations, false-starts, mispronunciations, and ungrammaticalities that it could not possibly be an adequate model from which to abstract complex and subtle linguistic regularities. Even if one acknowledged that child-directed speech is typically more coherent in structure than adult conversation (e.g., Snow & Ferguson, 1977), ethnographic research reveals substantial differences in the extent and nature of linguistic interactions with infants across cultures (e.g., Schieffelin & Ochs, 1986). Thus, it was apparently impossible to identify a universal set of features of child-directed input necessary for acquisition to take place (e.g., Lieven, 1994). Moreover, the wisdom of relying on a simplified child-directed register as the basis for grammatical development was called into question. Since speech to children is typically simpler than speech to adults, learning a grammar may be hindered by the fact that the input is limited in the scope and extent of the detailed syntactic information it can provide (Gleitman et al, 1988). Most significantly, several studies have demonstrated that caregivers do not provide enough explicit information to prevent the child from building overly general grammatical systems. This mistake, it was assumed, can only be overcome by linguistic input that provides "negative evidence", i.e., information about what sentences are not permitted by the target language (Marcus, 1993, cf. Sokolov & Snow, 1994).

These limitations of the input must be interpreted in the context of assumptions regarding what types of mechanisms are used by the language-learning child. In a famous examination of what would make languages “learnable,” Gold (1967) proposed that children are general learners who test hypotheses about grammatical rules against example sentences that they hear in the target language. In this demonstration, Gold provided what was considered to be compelling evidence that a general learner cannot induce the grammars of certain types of formal languages (which are derived from the class of natural languages) if it receives only “positive evidence” (i.e., only sentences that are grammatical in the language). The only conditions under which learning could be successful are: (1) if the learner is provided with negative evidence (i.e., cues to what sentences were ungrammatical), or (2) if the learner possessed strong initial biases about the types of hypotheses to consider in the first place. We should note that Gold’s learner incorporated an all-purpose learning mechanism that is clearly unlike any that has been proposed for young children. However, since several studies showed that explicit negative evidence is rare and is not universal in child-directed speech (e.g., Brown & Hanlon, 1970; Marcus, 1993), the only conclusion deemed reasonable at the time was that children must come pre-wired with a universal set of representational constraints on the types of grammars that are possible in human languages (e.g., Pinker, 1979). In other words, the complexity of the end-product and the indeterminacy of input to children (i.e., the “poverty of the stimulus”) appear to comprise compelling evidence that grammars cannot be learned. As Tomasello (2003) has recently put it, this view assumes that adult grammars go far beyond what children are capable of building given the resources available to them, i.e., “you can’t get there from here” (p. 2). Since most children do become relatively proficient at grammar within the first few years of life, it was logical to assume that it was only via a rich system of innately specified rules and representations, i.e., Universal Grammar, that all children could possibly zero-in on the particular set of rules that characterize their native language (e.g., Chomsky, 1975; Pinker, 1999, 2003).

Finally, this view also makes strong claims regarding the relations between linguistic and non-linguistic cognition. That is, the language faculty involves special processing mechanisms that are specifically dedicated to mediating the acquisition and processing of language. Moreover, sub-systems are themselves “modularized” in terms of components of the language faculty as traditionally defined by linguists, i.e., phonology, semantics, grammar, pragmatics (e.g., Fodor, 1983; Pinker, 1991; Levelt, 1989). Because these sub-components of language are assumed to be distinct in terms of the representations they employ, they are viewed as structurally autonomous and informationally encapsulated, not only from each other but also from the rest of non-linguistic cognition (e.g., Pinker, 1991). In sum, this nativist view of language development focuses on the specificity of the young child’s complex grammatical knowledge, the biological origins of its nature, and the universal course of its acquisition. As framed by Pinker (1994):

Language is a complex, specialized skill, which develops in the child spontaneously, without conscious effort or formal instruction, is deployed without awareness of its underlying logic, is qualitatively the same in every individual, and is distinct from more general abilities to process information or behave intelligently.

For these reasons, some cognitive scientists have described language as a psychological faculty, a mental organ, a neural system, and a computational module. But I prefer the admittedly quaint term ‘instinct’. It conveys the idea that people know how to talk in more or less the sense that spiders know how to spin webs...spiders spin spider webs because they have spider brains, which give them the urge to spin and the competence to succeed (p. 18).

### **3. A PARADIGM SHIFT: NEW PERSPECTIVES ON LANGUAGE LEARNING**

Although nativist views of language acquisition are forceful and still widely endorsed, there has been ongoing controversy about the adequacy of such theories as an account of how children develop competence in language. Some critiques directly challenge the logic of arguments made by Chomsky, Pinker, and like-minded theorists, questioning such core assumptions as the universality of generative grammar, the autonomy of syntax in language processing, and the fundamental unlearnability of language (e.g., Bates & Goodman, 1999; Braine, 1994; Pullum & Scholz, 2002; Tomasello, 1995). Other critiques focus on empirical evidence inconsistent with particular nativist assertions. For example, the claim that negative evidence is not available when children make grammatical errors, an assumption central to the “poverty of the stimulus” argument at the heart of Chomsky’s theory, is not supported by a recent analysis of parents’ reformulations in speech to children (Chouinard & Clark, 2003). These diverse challenges, both philosophical and data-driven, have fueled debate over four decades about the explanatory adequacy of nativist theories of language learning.

However, in recent years this debate has begun to change in focus and tenor, not only in response to explicit critiques within linguistics and developmental psychology, but also in response to research findings and theoretical insights from farther afield. An alternative perspective on language learning has been gathering force, amplified by new developments in research areas that formerly made little contact with theoretical debates on the nature of language development (see Kuhl, 2004; Seidenberg & MacDonald, 1999; Tomasello, 2003). We focus on four such developments that have begun to change the direction of research on early language acquisition: first, the emergence of more “user-friendly” theories of language and language use; second, the contribution of computational approaches to modeling language processing and learning; third, provocative findings from experimental research on learning and cognitive processing by infants; and fourth, insights from studies with children and non-human primates on the role of social cognition in communication. In different ways, these diverse areas all motivate and support an emerging alternative view of language learning.

#### **3.1. New Ways of Understanding Language and Language Use**

While generative theories have favored a view of linguistic competence defined exclusively in terms of grammatical knowledge, recent developments in both linguistics

and psycholinguistics have shifted the focus to a more inclusive view of competence, one that incorporates performance factors that guide language use. Within the newly emerging area of cognitive-functional linguistics, “usage-based” theorists emphasize the essential connection between the structure of language and how language is used to communicate (e.g., Croft, 2001; Goldberg, 1995). According to this view, linguistic competence cannot be reduced to knowledge of a core grammar as Chomsky claimed, but rather draws on a wide range of cognitive and social capabilities and on knowledge from diverse domains. Rejecting fundamental nativist assumptions about the nature of language, usage-based theories demand a very different view of what is involved in language learning (Tomasello, 2003).

In psycholinguistics as well there has been a dramatic shift away from models of speech processing that embody nativist assumptions, in favor of models that emphasize statistical and probabilistic aspects of language (Seidenberg, 1997). Until recently, the dominant processing theories have been those that presuppose the modularity of language, focusing on syntactic parsing strategies presumed to be automatic (see Frazier, 1987). For example, adults reading potentially ambiguous sentences such as “Put the apple on the towel into the box” are confused in predictable ways, presumed to result from of an irresistible initial tendency to interpret the prepositional phrase (PP) “on the towel” as modifying the verb and thus specifying the destination of the action. In fact, the first PP modifies “apple” rather than “put,” a reduced relative clause that attaches to the noun phrase rather than the verb phrase, catching the reader by surprise. Such classic “garden-path” effects were replicated in hundreds of experiments based on nativist assumptions, providing support for the idea that default syntactic processing strategies are automatic and impervious to influence from other sources of information. But this picture is changing, with the emergence of new experimental paradigms that use more revealing techniques for monitoring on-line comprehension. For example, when adults are able to look at a relevant visual scene while hearing “Put the apple on the towel into the box,” different results emerge (Tanenhaus, Spirey-Knowlton, Eberhard & Sediry, 1995). If a single apple on a towel is present in the scene along with a second towel and a box, listeners look briefly at the lone towel when they hear the first PP, before looking at the box. This is the behavioral equivalent of the garden-path effect observed previously in reading studies. However, if two apples are present in the scene, one on a towel and the other on its own, the same sentence is no longer perceived as ambiguous and there is no evidence of misinterpretation. To the contrary, the presence of the second apple provides immediate non-linguistic contextual support for interpreting “on the towel” as modifying the noun, crucial information that enables the listener to identify the correct referent. This experiment by Tanenhaus et al. was one of many recent studies using eye-tracking techniques to show that listeners integrate probabilistic information from multiple sources in interpreting spoken language, rather than defaulting to inflexible syntactic processing strategies (see Trueswell & Tanenhaus, 2005).

These new theoretical perspectives on language use emerging within linguistics and psycholinguistics are not only “user-friendly” in their emphasis on the flexibility and resourcefulness of mature language processing, but also “child-friendly” in their

developmental implications. Their influence on current research on early language learning is evident at several levels. Developmental theorists using observational techniques are testing predictions from usage-based theories, namely that children develop linguistic competence gradually, learning to produce new constructions item by item, rather than advancing by triggering innately specified grammatical rules that function in an all-or-none fashion (e.g., Lieven, Pine, & Baldwin, 1997; Tomasello, 2003). Researchers are also using new experimental techniques for assessing early speech-processing abilities to explore how infants in the first year track distributional information in spoken language (Saffran, Werker, & Werner, *in press*), and how by the second year they are able to interpret speech incrementally based on probabilistic information, similar to adults (Fernald, McRoberts, & Swingley, 2001). These are just a few examples to illustrate the more general point, that the application of probabilistic constraints is not only central to understanding adult competence in language processing but is now being extended to theories of language acquisition as well. Seidenberg and MacDonald (1999) make the case that the processes of constraint satisfaction that are critical to mature language production and interpretation are the same processes used by infants as they begin to make sense of speech and break into language.

### **3.2. Computational Approaches to Language Use and Language Learning**

The idea that attending to distributional information in speech might be critical in language learning was proposed many years ago (Maratsos & Chalkley, 1980), at a time when nativist views of language dominated the field. Another reason why this early interest in distributional learning was initially eclipsed was that the computational resources necessary for exploring such questions empirically were not yet widely available. In recent years, however, computational approaches of different kinds have become increasingly influential in research on language development, ranging from statistical analyses of language patterns to connectionist models. Because large corpora of spontaneous speech by parents and children are now accessible through the Child Language Data Exchange System (CHILDES) data bank (MacWhinney & Snow, 1990), researchers are able to undertake detailed analyses of the kinds of distributional information available in the language directed to the child. For example, statistical models have been used to reveal cues in child-directed speech that could potentially aid the young language learner in identifying word boundaries (e.g., Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998; Swingley, 2005) and in classifying new words in the appropriate grammatical form class (e.g., Mintz, 2003).

Statistical models such as these provide evidence that information about the distribution of linguistic units at various levels is available in the speech stream that could, in principle, facilitate learning by the child. Connectionist models are well suited to tackle the next question, asking what kinds of outcomes are possible at different phases in development given a particular input and a general-purpose learning mechanism (Elman et al., 1996). These models typically represent information in a distributed fashion across a set of connections between input and output units, although representational features and network architectures have been varied in many interesting ways (e.g., Shultz, 2003;

Munakata & McClelland, 2003). Over training, the networks extract and represent patterns of regularities in the input, abstracting information from multiple sources simultaneously and at multiple levels of granularity. Guided by the non-linear learning mechanism, the networks allow solutions to be represented as the coordinated activity of the network as a whole (i.e., the patterns of connections across the weights) that can be evaluated at different points in the training.

For example, in a series of models of the acquisition of inflectional morphology (Plunkett & Marchman, 1991), networks were trained to map words from their stem (e.g., walk) to past tense (e.g., walked) forms using artificial languages with different proportions of regular and irregular verbs. These factors were evaluated in a parametric fashion, revealing information about the conditions under which the networks would make use of phonological regularities across stem–past tense pairs as well as the role of both token frequency (i.e., how many times a particular stem–past tense mapping was seen) and type frequency (i.e., how many different stem–past tense pairs shared the same overall pattern, e.g., *ring-rang*, *sing-sang*) in determining learning. These features predicted the learning patterns in the networks and have been examined in several studies of natural languages (e.g., Bybee, 1995). Plunkett and Marchman (1993) next examined learning of stem–past tense mappings in the context of a lexicon that gradually increased in size over the course of the training, i.e., incremental learning. That is, the ability of the networks to memorize particular stem–past tense pairs or to generalize to novel forms was sensitive to developmental changes in the overall size and composition of the training set. Again, the predictions of these models have led to examinations of the role of vocabulary size in children’s learning of morphosyntax and the impact of individual differences in vocabulary size on later grammatical outcomes (e.g., Marchman & Bates, 1994; Bates & Goodman, 1999).

Another example is Elman’s (1993) simple recurrent network (SRN) model which abstracted syntactic regularities across sequential occurrences of lexical items. The model was presented strings of words in “sentences” generated by a human-like artificial grammar. The task for the SRN was to predict the upcoming word in the sequence, a task that is inherently probabilistic given the many possible words that could come next, especially across sentence boundaries. However, the task of “listening ahead” in essence forced the network to track distributional relations across the words, encoding the syntax of the sentences in terms of the varying conditional probabilities that were inherent in the example sentences. Interestingly, the network was successful in this task only when limitations were placed on the size of the “working memory” early in learning and then memory size was gradually increased across the course of training. The network’s limitation, as it turned out, was an advantage in learning the syntax of this artificial language, illustrating the importance of “starting small” (see also Newport, Bavelier, & Neville, 2001; Hertwig & Todd, 2003). While “starting small” may not always be necessary for successful learning (Rohde & Plant, 1999) and there are clearly limitations to what these models can tell us about human learning, such endeavors have served as natural tools for testing the conditions under which knowledge can emerge with exposure to different kinds of learning environments (e.g., Kersten & Earles, 2001). In addition, they have fueled new interest in exploring the ways in which knowledge is used and represented,

what kinds of information are available to the language-learning child, and what types of learning mechanisms can account for both the general patterns and individual variation seen across development (Elman et al., 1996; Elman, 2004, 2005).

### 3.3. Learning Strategies in Infancy

A third area in which recent developments have challenged traditional assumptions about language acquisition is research on infant learning strategies. When Skinner (1957) proposed that language was a behavior like any other animal behavior that could only be learned through gradual shaping and external reinforcement, Chomsky (1959) emphatically rejected this idea. The common observation that children begin to talk correctly in the absence of explicit positive or negative feedback was clearly at odds with the idea that language learning is achieved through a process of conditioning. Chomsky's claim that principles of learning could not possibly explain how children master language was based on the behaviorist learning theory of the day, which was much too simplistic to account for the complexities of children's linguistic ability. In recent years, however, researchers investigating early perceptual and cognitive development have made stunning discoveries about the learning capacities of young infants, and a new view of the potential role of learning in language development has emerged. Thus, while computational models reveal how much information about language structure is potentially available to the young language learner, these new behavioral studies confirm that infants are in fact able to learn from this information.

For example, Saffran, Newport, and Aslin, (1996) showed that eight-month-olds can segment a stream of meaningless syllables containing no acoustic or prosodic cues to word boundaries after only a few minutes of listening experience. The information infants are using to identify word-like units in this case is distributional evidence, the regularities in the relative position and order of particular syllables over the whole sequence. For example, one string of syllables consisted of *pa bi ku go la tu da ro pi ti pu do da ro pi go la tu...* After familiarization with this sequence, infants were tested with "words" that had occurred in the string, i.e., sequences of syllables that always occurred in the same order, such as *pabiku* and *golatu*. They were also tested with "non-words," combinations of familiar syllables that spanned two different words, such as *kugola*. Although there was no acoustic information specifically marking word boundaries, the transitional probabilities were much higher between syllables within words than between words. Thus there was statistical information that could enable infants to identify the familiar word-like units in the stream of speech. The finding that they are capable of performing such computations reveals the sophisticated talents young learners bring to the task of segmenting speech, months before they are able to understand meanings in the words they hear.

### 3.4. Social Cognition in Infants and Non-Human Primates

Another domain of research that is yielding surprising findings relevant to language learning focuses on the abilities of human infants and animals of other species to

appreciate the mental states of others. Ethologists have long been interested in the communicative function of animal signals, attempting to establish criteria for determining whether primate vocalizations constitute intentional and semantically meaningful signals to conspecifics. Vervet monkeys, for example, give at least three acoustically distinctive calls in response to particular predators (Cheney & Seyfarth, 1997). When one of these call types is played back in field experiments in the absence of an actual predator, adult vervets react appropriately, taking a different escape route in response to a “snake call” than to an “eagle call.” These and other recent ethological observations have dispelled the prevailing assumption that primate communication is entirely reflexive and emotional in nature and thus must be very different from human communication. These findings are of interest to many outside the field of biology, posing questions of concern to philosophers, linguists, and psychologists. What sorts of mental representations underlie the production and perception of these primate calls? When a vervet gives a snake alarm call, is this analogous to a child crying out “Snake! Watch out!” on seeing slithery movement in the grass?

Cheyney and Seyfarth (1997) conclude that vervet and human communication differ fundamentally, in that monkeys call and look at each other in order to influence each other’s behavior, whereas children do so in order to influence their attention or knowledge. According to these researchers, the lack of a theory of mind in vervets is one fundamental reason they are incapable of language. However, these questions are complex, and studies with other species show that animals as diverse as sea lions, parrots, and bonobos are capable of learning to use symbols and to distinguish between objects, actions, and modifiers. There is energetic debate about how these abilities should be interpreted and how they differ among species, in particular whether chimpanzees and other higher primates show more sophisticated competence in reading the goals and intentions of others than do monkeys (Tomasello & Carpenter, 2005).

This new wave of research on animal communication resonates in interesting ways with two flourishing areas of research with human children. The first area explores the growth of theory of mind in preschoolers (e.g., Wellman, 2002), while the second focuses on children’s early sensitivity to communicative behaviors such as emotional expressions, gaze direction, and pointing (Baldwin & Moses, 1996). Experimental studies using looking-time measures show not only that young infants can use such vocal and visual cues to guide their attention (e.g., Mumme & Fernald, 2003), but also that they use non-verbal referential cues in combination to make inferences about the goals and intentions of others (see Rochat, 1999). It may not be obvious what these studies of non-verbal communication in animals and infants have in common with the paradigms described in the three previous sections, which all seem more directly relevant to the new focus in language research on how children and adults evaluate multiple sources of probabilistic information in interpreting and using language. But the relevance of research on intentionality to this new focus will become clear when we review current research on lexical development. Through the work of Tomasello (2003) and others, it is increasingly apparent that the ability to learn what a word means and to use words in communication depend crucially on fundamental skills of joint attention and intention-reading. Because



language use involves mind-reading, children must learn to interpret others' mental states and to integrate probabilistic information on this level with information at other linguistic and non-linguistic levels in order to interpret language and to communicate effectively.

In summary, for half a century research on language development has been dominated by theoretical claims and assumptions emerging from a powerful linguistic theory that defines competence in terms of innately specified grammatical knowledge. Recent developments in diverse fields outside the mainstream of linguistics are forcing re-examination of these assumptions from different angles, all suggesting alternative ways of thinking about how human language functions and what it means to learn a language. A dominant theme emerging from the new paradigms and findings in the four areas reviewed above is that human communication relies on the integration of many different sources of information, rather than on innately specified knowledge in an encapsulated, modular system isolated from other cognitive and social capacities. Seidenberg and MacDonald (1999) refer to this emerging perspective as the "probabilistic constraints" approach, which has as its central idea that language learning by children and language processing by adults both involve the use of "multiple, simultaneous, probabilistic constraints defined over different types of linguistic and nonlinguistic information" (p. 570).

This theme is reflected in some of the most interesting new research on infant speech perception, early word learning, and the emergence of efficiency in spoken language understanding, to be reviewed in the following sections. The purpose of this brief review is to point to a few of the recent contributions to research on language acquisition in infancy that exemplify this sort of probabilistic perspective. We do not attempt to provide a comprehensive review of this area and certainly do not assume that this approach has all the answers. Instead, our goal is to describe new research on how children start out by attending to patterns in speech sounds in the first year and learn to listen for meaning in speech in the second year. By focusing on how infants, from the beginning, attend to multiple sources of information in making sense of spoken language, this approach emphasizes the continuity between current psycholinguistic theories about how language is used and emerging developmental perspectives on how language is learned.

#### **4. LEARNING ABOUT THE SOUNDS OF SPEECH IN THE FIRST YEAR**

To begin making sense of speech, infants must discern regularities in the sequences of sounds used by speakers of the particular language they are hearing. Hundreds of experiments on speech perception in the first year of life have shown that months before understanding or speaking a single word, infants become attuned to characteristic sound patterns in the ambient language (see Jusczyk, 1997). While early research in this area focused on discrimination and categorization of consonants and vowels, more recent studies are exploring infants' implicit learning of complex distributional patterns in spoken language and how these learning strategies enable infants to find the words, using multiple sources of information available in the sound patterns of continuous speech.

#### 4.1. Early Attention to Speech Sounds

Given the abundance of findings in what is now regarded as one of the most exciting areas of research in cognitive development, it is hard to believe that infants' sophisticated speech-processing abilities are a relatively recent discovery. To give some historical perspective, in 1970 Bernard Friedlander published a research overview entitled *Receptive Language Development in Infancy: Issues and Problems*, motivated by the following concerns:

Judging by the theoretical and speculative literature as it stands today, receptive language development in infancy is a minor topic of marginal significance. Issues related to infant listening and receptive processes are virtually ignored in ... the new wave of language studies that assumed torrential proportions in the early 1960's. Though there is a general acknowledgment... that language input is a necessary prerequisite for the organization of speech, the topic is seldom accorded more than a few sentences... and some of these discussions are highly patronizing in tone. They seem to suggest that auditory perception in general and language perception in particular are topics on which thoughtful observers would hardly need to spend much time. There is little in this literature to suggest that the problem of how babies come to recognize the phonological, lexical, semantic, and grammatical systems in the language they hear represents a psychological, linguistic, and developmental problem of the greatest magnitude (1970, p. 7).

Friedlander's (1970) paper was more a lament than a review, because there was almost no research available at the time: "Hardly enough is known at a factual level about early listening processes and their role in language growth even to organize the phenomena in reasonably durable categories" (p. 8). Then just one year later the situation changed, when the first reasonably durable categories were discovered through the pioneering experiments of Eimas, Siqueland, Jusczyk, and Vigorito (1971). Using an innovative operant technique called the high-amplitude-sucking procedure, these researchers were able to show for the first time that young infants can discriminate and categorize speech sounds.

To fluent English speakers the speech sounds /b/ and /p/ sound clearly distinct, so it is difficult to appreciate how formidable this task could be to a linguistically inexperienced listener. In fact, these two consonants are acoustically very similar; moreover, /b/ and /p/ vary acoustically when combined with different vowels or when they occur in different positions in a word. Eimas et al. showed that even very young infants could discriminate /b/ from /p/, and just like adults, they failed to distinguish different tokens of /b/ that were acoustically distinct yet were members of the same phonetic category. Infants can also appreciate the fact that vowel tokens that are acoustically dissimilar may be equivalent in terms of their phonetic identity. Using an operant head-turn procedure, Kuhl (1979) showed that five-month-old infants readily discriminated /a/ from /i/ when spoken with the same intonation by the same female speaker. However, they grouped together several different tokens of /a/ that were acoustically variable, produced by male and female

speakers using both rising and falling pitch contours. These classic studies on early speech processing abilities showed that infants can attend to the acoustic variability relevant to the phonetic identity of speech sounds, while ignoring acoustic variability that is linguistically irrelevant.

The first experimental studies on early speech perception focused on infants' ability to distinguish isolated syllables. The questions initially of interest derived from controversial issues in research on adult speech perception, with the infant representing the "initial state," or the listener innocent of experience. Studies with infants were seen as test cases relevant to current debates about which acoustic features are most critical for human speech perception (e.g., Eimas & Corbit, 1973), whether speech and non-speech sounds are processed in fundamentally different ways (e.g., Jusczyk, Rosner, Cutting, Foard, & Smith, 1977), and whether speech sounds are represented in terms of phonetic features or syllables (Bertoncini, Bijelijac-Sabic, Jusczyk, Kennedy, & Mehler, 1988). While providing valuable information about early perceptual abilities crucial for speech processing, most of these early studies were "developmental" only in the sense that they showed these capabilities were already present at birth. A more dynamic picture has emerged in recent years as researchers have begun to focus on developmental change in speech processing strategies, first by exploring how experience with a particular language shapes perception, and second by asking how infants learn to recognize patterns in speech that may help them identify linguistic units.

#### 4.2. Becoming a Native Listener

Although infants are clearly born with perceptual abilities and biases that equip them for organizing speech sounds into linguistically relevant categories, these perceptual grouping strategies are neither unique to humans nor unique to speech sounds. Other primates also organize human speech sounds categorically, and some other kinds of acoustic stimuli are perceived in a similar fashion (see Kuhl, 2004). What *is* presumably unique to humans is the perceptual learning that occurs over the first few months of life as a result of hearing a particular language. Adults often find it difficult or even impossible to distinguish certain speech sounds in an unfamiliar language. For example, native speakers of Hindi can easily discriminate the consonants /Ta/ and /ta/, but to monolingual English-speaking adults they sound like indistinguishable tokens from the English category /t/. However, six-month-old infants growing up in English-speaking families can effortlessly discriminate the Hindi contrast /Ta/ - /ta/ (Werker & Tees, 1984). Studies of adult perception of native and non-native speech sounds show that adults have become specialists, attentive to phonetic distinctions relevant in the languages they have learned but less discerning in making other distinctions. Yet infants must start out with the potential to make a wide range of distinctions. When does this process of perceptual specialization begin? Werker and Tees tested English-learning infants at three ages between 6 and 12 months, to investigate whether they retained their ability to discriminate non-native speech contrasts across the first year. Infants at each age listened either to the Hindi consonants /Ta/-/ta/ or to consonants from the Nthlakampz language, /k'i/-/q'i/,

which are also very difficult for English-speaking adults to discriminate. Almost all of the infants at 6–8 months could discriminate both non-English contrasts, although very few of the infants at 10–12 months were able to distinguish either pair.

Further evidence for the influence of the ambient language on infants' emerging phonetic categories comes from research by Kuhl, Williams, Lacerda, Sterens, & Lindblom (1992), who showed that six-month-old infants hearing only Swedish or English already grouped vowels perceptually in categories appropriate to the language they were learning. Recent studies measuring brain activity are generally consistent with the behavioral findings showing increasing specialization for familiar speech sounds over the first year. At six months of age, infants show an electrophysiological response to changes in both native and non-native speech contrasts, but by 12 months the response is elicited only by changes in speech sounds native to the language the child has been hearing (e.g., Cheour-Luhtanen et al., 1995). These results indicate that auditory experience over the first-year results in neural commitment to a particular perceptual organization of speech sounds appropriate to the ambient language. Through early experience with the speech around them, infants adapt their perceptual strategies for efficiency in processing the language they are learning.

### 4.3. Finding the Words in Fluent Speech

Other studies of developmental change in speech perception have focused on the discovery procedures infants use to identify higher-order elements in spoken language. An influential article by Lila Gleitman and colleagues stimulated this new research direction (Gleitman et al., 1988). They proposed that infants might be able to use certain prosodic features in continuous speech, such as pauses and the vowel lengthening typically preceding pauses, as cues to the boundaries of phrases and clauses, a perceptual discovery strategy that could be useful to the child beginning to learn syntax. This “prosodic bootstrapping hypothesis” generated considerable interest, leading to experiments showing that 10-month-old infants seemed to recognize violations of common prosodic rhythms in the ambient language (e.g., Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). Although there were also counterarguments against the view that prosodic cues are sufficiently regular as to provide reliable cues to syntactic units in speech (e.g., Fernald & McRoberts, 1995), the prosodic bootstrapping hypothesis stimulated the first wave of research exploring how infants might use lower-level acoustic cues in speech to gain access to linguistic structure at higher levels (see Morgan & Demuth, 1996).

The strong claim that young infants must first rely on prosodic information in order to organize segmental information in continuous speech has receded in light of new findings showing that infants are much more adept at identifying word-like units in fluent speech than anyone had imagined. Jusczyk and Aslin (1995) investigated the ability of seven-month-old infants to detect repeated words embedded in fluent speech. When infants were first familiarized with multiple repetitions of a word such as *bike* or *feet* and then tested in an auditory preference procedure with passages that either did or did not

contain the familiarized word, they preferred to listen to passages containing the familiar word. This finding indicated that infants were able to segment speech into words without benefit of exaggerated prosodic cues. However, prosody at the level of word stress does play a role in facilitating such segmentation. English-learning infants are more successful in segmenting words such as *bor'der* that have a strong–weak accent pattern than words such as *guitar'* that have the opposite pattern, because they have already learned that the strong–weak pattern is dominant in the language they are hearing (Jusczyk, 1998). In contrast, French-learning infants appropriately show the opposite bias, based on their experience hearing words with weak–strong accent patterns.

Many studies have now demonstrated infants' sensitivity to particular cues in the ambient language such as phonotactic regularities (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993) and lexical stress (Jusczyk, Cutler, & Redanz, 1993), and their ability to take advantage of these cues in identifying word boundaries (e.g. Johnson & Jusczyk, 2001). In an influential study mentioned earlier, Saffran et al. (1996) showed that infants were also able to use sequential statistics to discover word-like units in continuous speech, in the absence of any other acoustic cues to word boundaries. Unlike the experiments on segmentation by Jusczyk and Aslin (1995) in which infants were first familiarized with samples of natural language, Saffran et al. exposed infants briefly to strings of meaningless syllables, i.e., stimuli that were language-like but entirely novel in their organization. After only a few minutes of passive exposure to these sounds, infants picked up on the regularities and attended longer during testing to sequences that deviated from these regularities. Using a similar training procedure, Chambers, Onishi, and Fisher (2003) showed that infants can also learn new phonotactic regularities after minimal exposure. Moreover, they can also quickly take advantage of such newly learned phonotactic patterns, using them as cues to identify the boundaries of novel words (Saffran & Thiessen, 2003). Thus although infants in the second half of the first year may already show a strong commitment to the particular sound patterns they have absorbed from hearing their native language, early speech processing remains a highly dynamic process. Infants remain open to new experience as they build on prior learning, drawing on multiple sources of information to find order in novel sounds.

Experimental studies of early speech processing proceed parametrically, typically investigating one isolated variable at a time. However, infants listening to natural speech are in fact confronted with multiple sources of information at any moment, some redundant and others in conflict. Building on a strong foundation of research on infants' use of individual cues to word boundaries, several recent studies have begun to explore how young language learners tackle this challenging problem. Just as research on adult speech processing now focuses on how listeners integrate probabilistic information from numerous sources (Seidenberg, 1997), developmental researchers are beginning to ask parallel questions of very young infants (e.g., Curtin, Mintz, & Christiansen, 2005; Mattys, White, & Melhorn, 2005; Thiessen & Saffran, 2003, 2004). These and many other new findings confirm the wisdom of Friedlander's (1970) intuition 35 years ago, "that the problem of how babies come to recognize the phonological, lexical, semantic,

and grammatical systems in the language they hear represents a psychological, linguistic, and developmental problem of the greatest magnitude” (p. 7).

## 5. LEXICAL DEVELOPMENT

These studies of early speech perception show that infants in the first year of life are becoming skilled listeners, capable of making detailed distributional analyses of acoustic–phonetic features of spoken language. Although such accomplishments are often cited as evidence for early “word recognition,” they are perhaps more appropriately viewed as evidence of pattern detection abilities pre-requisite for recognizing words in continuous speech. Identifying particular sound sequences as coherent acoustic patterns is obviously an essential step in word recognition, but this can occur without any association between sound and meaning. Laboratory studies show that around 5–6 months of age infants respond selectively to their own name (Mandel, Jusczyk, & Pisoni, 1995), and by 10 months appear to have some kind of acoustic–phonetic representation for a number of frequently heard sound patterns (e.g., Halle & de Boysson-Bardies, 1994). Because this selective response to familiar words in the early months of life can occur with no evidence of comprehension, it may constitute word recognition in only a limited sense. However, most infants do begin to respond to and utter sounds in meaningful ways by their first birthday, and one year later are able to speak dozens of words quite convincingly. In this section, we review research on children’s first speech productions and the course of early vocabulary growth, as well as the factors that influence word learning in infancy.

### 5.1. First Words

According to parents’ reports of their children’s spontaneous responses to speech, infants typically begin to associate sound sequences with meanings toward the end of the first year. By eight months, on average, many children respond appropriately to about 10 familiar phrases, such as “Where’s Daddy?” (e.g., by turning and crawling toward the door), or “It’s time for bath!” (e.g., by plopping down and attempting to remove their shoes) (Fenson et al., 1994). While it could be tempting to assume that the child is actually interpreting each of the words in these phrases, it is more likely that children are using a variety of cues, both linguistic and contextual, to make sense of these frequent expressions (e.g., Daddy just left the room, Mom is holding a towel standing next to a running faucet). At the same time, the fact that children do respond in these ways indicates that they are paying attention to the speech around them and beginning to understand it by associating certain sound patterns with particular contexts. Only a short time later, children begin to demonstrate an ability to understand individual words with less and less contextual support, an ability that will continue to improve over the next several months. Based on reports from more than 1000 parents, Fenson et al. (1994) cite that the median-level 10-month-old understands approximately 40 words, while the median-level 18-month-old understands more than 250 words, a more than six-fold increase.

The production of recognizable words begins, on average, just before a child's first birthday. These early words cross-cut a variety of linguistic categories, but are typically names for caregivers (e.g., *mama*), common objects (e.g., *bottle*, *shoe*), social expressions (e.g., *bye-bye*), with some modifiers (e.g., *hot*) and actions or routines (e.g., *peek-aboo*, *throw*) (Nelson, 1973). New words tend to enter children's expressive vocabularies over the next several months at a relatively slow but steady pace, reaching an average of 300 words by 24 months and more than 60,000 by the time they graduate from high school (Fenson et al., 1993). Thus, after the slow start, many children appear to undergo a "vocabulary burst," a sudden and marked increase in how many words children use (e.g., Goldfield & Reznick, 1990; Mervis & Bertrand, 1995). Putting aside the difficulties of defining how much of an increase constitutes a "spurt" in rate of learning, the characterization of lexical development in terms of a sudden increase in learning rate has been interpreted to indicate two distinct phases of lexical acquisition (i.e., pre-spurt vs. post-spurt). Some researchers have associated this "burst" with the achievement of linguistic milestones, for example, children's new understanding about what words are for (the "naming insight") (e.g., Dromi, 1987; Bloom, 1973), improved word segmentation abilities (Plunkett, 1992), or enhanced word retrieval skills (Dapretto & Bjork, 2000). Other researchers have associated increases in lexical growth with cognitive advances related to the nature or organization of object concepts (e.g., Gopnik & Meltzoff, 1987).

However, other researchers have suggested that these increases in the rate of learning are relatively constant across the period, questioning the idea of a "spurt" at all, and hence, its role as a marker for other cognitive or linguistic events (Bates & Goodman, 1999; Bloom, 2000; Ganger & Brent, 2004). Moreover, work within a connectionist framework suggests that it may not be necessary to assume that any shift in the trajectory of vocabulary growth would be associated with the emergence of a new insight or learning mechanism at all. Instead, both the slow and steady pace of learning early on and the accelerated learning later can be accounted for within a single explanatory framework. In Plunkett et al. (1992) a connectionist network was trained to associate labels (i.e., words) to random-dot "images" (i.e., pictures). In the "language" and "world" of this network, several images had the same "name" and there was no information in the images regarding what the label should be. Just like in natural languages then, the mappings between labels and images were arbitrary and many-to-one. Importantly, however, in the network, a vocabulary "spurt" occurred without any shift in the underlying mechanism guiding the learning. That is, even though the identical network was solving an identical task throughout development, shifts in learning rate were observed in the behavior of the network. As summarized in Elman et al. (1996), "there is no need to build in additional architectural constraints or to invoke changes in the input to explain the vocabulary spurt. It is simply an emergent function of the processing" (p. 128). The shifts were a natural consequence of processing limitations that arise over the course of a gradual and continuous learning process that involves complex and multiply determined mappings.

The fact that learning words involves gradually building a system of mappings using a variety of linguistic and non-linguistic constraints is also evident when examining changes in the types of words that children produce. Even though children's first words

come from a range of lexical categories (e.g., Bloom, 1973), first words are typically open class or content words (nouns, verbs, adjectives), and only later grammatical function words such as prepositions, determiners, and pronouns. Within the open class, children's first words tend to be referential (i.e., concrete nouns), and only later are children producing predicative terms (e.g., verbs and adjectives) (Bates, Bretherton, & Snyder, 1988; Benedict, 1979; Brown, 1973; Nelson, 1973). This dominance of concrete nouns in early vocabularies (sometimes referred to as a "noun bias") is most evident in the first 200 words or so, after which there tends to be an increase in the proportion of vocabularies devoted to predicative terms, for example, verbs (e.g., *go*), adjectives (e.g., *hot*), and closed-class functors (e.g., *and*, *of*) (Dromi, 1987; Nelson, 1973; Bates et al., 1994).

Several explanations have been proposed for these developments. First, it is possible that open-class words are learned early because they are longer in duration, generally stressed, and phonologically less reduced than closed-class words (Morgan, Shi, & Allopenna, 1996). Even six-month-old infants show a preference for listening to open class rather than closed-class words (Shi & Werker, 2001), and newborns are sensitive to the acoustic differences between these word types (Shi, Werker, & Morgan, 1999). Second, it might also be the case that the ability to learn predicates is dependent on amassing a particular body of referential terms, and that using words that do grammatical work (e.g., functors) is dependent on the acquisition of a set of content words on which they can operate (i.e., "from reference to predication to grammar," Bates et al., 1994, p. 98). Third, concrete nouns are conceptually simpler than verbs and other predicates, and both of these types of open-class words are more conceptually transparent than grammatical function words (Gentner, 1982; Gasser & Smith, 1998). More specifically, it has been proposed that the early priority of nouns in children's vocabularies reflects the fact that nouns are more "cognitively dominant" than verbs and closed-class items (Gentner, 1982; Gentner & Boroditsky, 2001). That is, nominals serve primarily to denote concrete objects that are more bounded and more perceptually individuated than, for example, verbal terms denoting states and processes which must rely on their arguments (i.e., nouns) to make sense. Relational words (e.g., closed class terms), in contrast, are "linguistically dominant," in that they derive their meaning from other parts of the linguistic context and sometimes reflect relatively opaque grammatical constructs (e.g., *gender*) (Gentner & Boroditsky, 2001).

The general pattern of early nominals appearing before other more relational terms has been observed in the vocabularies of children from several language-learning communities, even though their languages have different typological features that potentially make them more or less "noun friendly" (Gentner, 1982; Caselli, Casadio & Bates, 1999; Jackson-Maldonado et al., 2003; Bornstein, et al., 2004). For example, Caselli, Casadio, and Bates, (1999) used parent report to contrast early vocabulary composition in young English- and Italian-learners. In spite of the fact that Italian, but not English, is a pro-drop language (i.e., the subject noun phrase can be omitted leaving verbs in the very salient sentence-initial position), a similar level of noun bias was observed in the two languages. Likewise, in a recent study of middle-class children learning English, Italian, Spanish, Dutch, French, Hebrew, and Korean, Bornstein et al. (2004) report few crosslinguistic



differences in vocabulary composition, with mothers of children in all language communities reporting more nouns than words in other language classes. Interestingly, Caselli et al. also report that Italian-learning children were likely to have more social terms and names for people in their vocabularies than English-learners, suggesting that cross-linguistic differences in vocabulary composition may be more attributable to social or cultural factors (e.g., a tendency to live near extended family) than specific features of the language. Similarly, in a study of children learning English, Italian, and Spanish in urban and rural communities (Bornstein & Cote, 2005), the observed variation in vocabulary size and composition was generally attributable to cultural factors, favoring urban over rural settings, rather than to the particular language being learned.

However, the claim that a “noun bias” may be a universal feature of early conceptual and linguistic development has been challenged by other studies of children learning Korean (Choi & Gopnik, 1995), Mandarin Chinese (Tardif, Gelman, & Xu, 1999), and Japanese (Fernald & Morikawa, 1993). All of these languages have features that make them more verb-friendly than English and Italian. For example, while both Italian and Mandarin are “pro-drop” languages, Mandarin verbs are considered to be more morphologically transparent than Italian verbs with their rich morphology. Indeed, naturalistic observational data indicated that young Mandarin-learners produced a higher proportion of verbs than nouns, compared to their Italian- and English-speaking peers (Tardif et al., 1999). This is in contrast to a consistent pattern of noun dominance using parent report, suggesting that parent report may tend to over estimate nouns (and underestimate verbs) in children’s vocabularies. Interestingly, Tardif et al. (1999) found that Mandarin mothers produced more verbs than nouns in their spontaneous speech to their children, and Fernald and Morikawa (1993) found that Japanese mothers labeled objects less frequently and less consistently than English-speaking mothers. In several studies, English-speaking mothers were more likely to use and elicit more nouns than verbs from their children, and place them in salient positions in the sentence when engaged in activities such as object-naming or book-reading (e.g., Hoff, 2003; Tardif, Shatz, & Naigles, 1997). Again, it appears unlikely that a single factor can account for the “dominance” of nouns in children’s early vocabularies, but rather children’s early vocabulary compositions are determined by a multitude of factors that happen to vary across languages and language-learning situations.

While many studies have examined the early stages of vocabulary development, it is still difficult to ascertain what a child really knows when they understand or produce a word. When talking about early lexical development, it is tempting to credit a child with “knowing” a word, as if word knowledge is something that the child either has or does not have, i.e., as if words are acquired in an all-or-none fashion. However, early studies noted that children’s early words (e.g., *bottle*) do not necessarily have the same meanings (e.g., *white plastic 6 oz cup with the bright red screw-on lid*) as they would for the adult (e.g., *receptacles of all shapes and sizes from which one generally pours liquids*). Early words are frequently used in very context-specific ways (i.e., under-extensions), only with reference to specific objects in specific situations (Bloom, 1973; Barrett, 1986; Harris, Barrett, Jones, & Brookes, 1988). At the same time, children’s early word uses

might also be considerably more broad (i.e., overextensions) than one would expect based on adult-like meaning categories (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979). For example, a child using the word “dog” to refer to all four-legged animals (dogs, but also cats, cows, and horses) could be an example of a child misrepresenting the adult-like meaning for words. Yet, research has also shown that over- or under-extensions are actually relatively rare (Rescorla, 1980; Harris et al., 1988; Clark, 2003b), and may be less of a reflection of how children represent the word’s meaning in some sort of mental lexicon and more related to the child’s ability to put their lexical knowledge to work in real time (e.g., Huttenlocher & Smiley, 1987). That is, a child who uses the word “dog” when the family cat runs across his path may not actually think that dogs and cats are the same thing, but simply cannot generate the appropriate word in the heat of the moment. Interestingly, children’s over- and under-extensions are considerably more frequent in production than comprehension (e.g., Clark, 2003a), and an experimental study using a looking-preference procedure has shown little concordance between comprehension and children’s over- and under-extensions in production (Naigles & Gelman, 1995). Taken together with research on children’s processing of speech in real time, described in more detail below (e.g., Fernald, Perfors, & Marchman, 2006), these studies suggest that early lexical development is quite gradual. It involves not only building-up “adult-like” meaning representations, but also learning to use words in more and more contextually flexible ways and in more and more challenging contexts (Bates et al., 1979; Barrett, 1976).

## 5.2. Individual Differences in Vocabulary Development

So far, we have been talking about the general features of lexical development in “the modal child” (Fenson et al., 1994, p. 1). However, there is considerable variation in both *when and how* children build their receptive and expressive vocabularies (Bates et al., 1988, 1994; Bloom, Lightbown & Hood, 1975; Fenson et al., 1994; Goldfield & Snow, 1985; Nelson, 1973, Peters, 1977, 1983). For example, while many children show signs of word comprehension at 8 or 10 months of age, other children do not respond systematically to the speech around them until several months later. Similarly, some children produce their first words well before their first birthday, while others do not do so until 14 or 15 months of age. The “modal” 18-month-old has already built up a 50–75 word expressive vocabulary, yet other children do not amass this many recognizable words until 22 months or later. Some of these “late talkers” will catch up in vocabulary a few months down the road, while others will remain late and continue to be at risk for language or learning disorders (Bates, Dale, & Thal, 1999). Studies of variation in vocabulary composition have noted that some English-speaking children tend to adhere to a strong “noun bias” tendency, a so-called “referential style” of early word learning (e.g., Nelson, 1973; Pine & Lieven, 1990). In contrast, other children with a more “expressive” style tend to have a smaller proportion of concrete nouns, preferring more “canned phrases” (e.g., *I wanna do that!*) and social expressions (e.g., *no way Jose!*) (Nelson, 1973).

Such individual differences have been well-documented since the mid-1970s, based primarily on diary studies (e.g., Nelson, 1973). However, important progress in understanding

the extent of individual variation has been facilitated by large-scale studies which rely on reports from parents, e.g., MacArthur-Bates Communicative Development Inventories (CDI) (Fenson et al., 1993) and the Language Development Survey (Rescorla, 1989). While there are clearly limitations to this methodology (e.g., Mervis & Tomasello, 1994), this technique has enabled the examination of variation in lexical milestones in several languages, for example, English (Bates et al., 1994), Italian (Caselli et al., 1999), Mexican Spanish (Jackson-Maldonado et al., 2003), Hebrew (Maitel, Dromi, Sagi & Bornstein, 2000), as well as in children learning two languages simultaneously (Pearson, Fernández & Oller, 1993; Marchman & Martínez-Sussmann, 2002) and in children from urban and rural settings (Bornstein & Cote, 2005). Although there is clearly variation in early acquisition across languages (e.g., Caselli et al., 1999; Choi & Bowerman, 2001; Tardif et al., 1999), research has consistently demonstrated remarkable similarities across languages in the overall size of children's vocabularies and the *extent* of the variation that is observed. Commenting on their comparative data across Spanish, English, and Italian, Bornstein and Cote (2005) noted a strikingly similar range of vocabulary knowledge in all three languages, leading these authors to conclude that individual variability is probably a universal feature of early language acquisition.

What are the sources of individual differences so early in development? Some studies have looked to child-factors, such as gender or birth order, to explain variation in lexical development. Studies have documented somewhat larger vocabularies and faster rates of growth in girls compared to boys (Fenson et al., 1994; Huttenlocher, Haight, Bryk, & Seltzer, 1991) and first-borns compared to later-borns (e.g., Hoff-Ginsberg, 1998). Recently, Bornstein and Cote (2005) note a consistent advantage for girls over boys in reported vocabulary in English, Italian, Spanish, across both urban and rural communities. While it is striking to see consistent gender effects in communities that likely vary in gender-based social expectations, these effects are generally small relative to overall developmental effects (i.e., the impact of gender is considerably smaller in magnitude than age effects). Like gender, the impact of birth order is relatively minor compared to other factors, but points to the suggestion that children, even those living in the same family, can differ in the frequency and character of interactions in which they engage on a regular basis.

Indeed, it is well-known that there are considerable individual differences in the quantity and quality of the talk that children hear (e.g., Huttenlocher et al., 1991; Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Hart & Risley, 1995), and that several of these features of maternal talk are directly linked to children's vocabulary outcomes (Hoff, 2003). In a recent large-scale study of low-income families, Pan, Rowe, Singer, and Snow (2005) found that variation in growth in children's vocabulary from 14 to 36 months was significantly related to diversity of maternal talk, in particular, the number of different words produced during mother-child interaction. Thus, children who hear a rich vocabulary that includes a higher proportion of low-frequency or complex words are likely to develop their own vocabularies at a faster rate (see also Weizman & Snow, 2001; Hoff & Naigles, 2002). However, Pan et al. (2005) also found that features

of maternal knowledge (e.g., scores on standardized tests of language and literacy) and maternal mental state (e.g., depression index) also contributed to child outcomes.

Other researchers have characterized individual differences in terms of cognitive or processing abilities. Typically based on naturalistic data, several early studies proposed that children may vary in the tendency to select analyzed vs. unanalyzed units (e.g., Peters, 1983), the use of strategies for segmentation which favor “word-sized” vs. “phrase-sized” units (Plunkett, 1992; Bates et al., 1988), a predilection for “imitativeness” (Bloom et al., 1975), or the ability to use contextual or linguistic cues to retrieve words (Bloom, 1973). Other studies have used more processing-based measures to assess various skills that could underlie vocabulary development. Using a phonetic discrimination task, Werker, Fennell, Corcoran, and Stager (2002) found that 14-month-olds who were successful at learning phonetically similar words had relatively larger vocabularies. Interestingly, no relation was observed in 18-month-olds, suggesting that this type of skill may only be helpful at the beginning phases of building a vocabulary. Similarly, Swingley, and Aslin (2000, 2002) also found little relation between vocabulary size and the ability of 14- and 18-month-olds to identify words pronounced correctly and incorrectly. However, Fernald and colleagues (Fernald et al., 2001; Fernald, 2002) found that 18- and 21-month-olds who had larger production vocabularies were faster than their lower-vocabulary peers to recognize words based on partial phonetic information and were more efficient at using verb semantics to predict what was coming up in the sentence. Using a similar procedure, Zangl, Klarman, Thal, Fernald, and Bates (2005) found that infants with larger vocabularies were more efficient at processing words that were perceptually degraded. Finally, in a longitudinal sample, Fernald et al. (2006) have recently shown that efficiency of spoken language understanding was related to trajectories of growth in vocabulary from 12 to 25 months, as well as several indices of early grammar.

Thus, individual differences in children’s burgeoning vocabulary knowledge appear to be linked to a variety of skills that come into play during the processing of both linguistic and non-linguistic information during real-time language comprehension. While we are still a long way from knowing exactly how those factors operate over the course of development, it is likely that individual differences in lexical development are linked to a host of factors, both child-related and experience-related, that all contribute to the variation that is so pervasive in vocabulary development.

### 5.3. Early Word Learning

Observational studies and research using parental report measures can provide normative data on the rate and composition of vocabulary growth by children at different ages, as well as correlational data showing how environmental factors such as the amount and quality of parental speech relate to lexical development (e.g., Huttenlocher et al., 1991). However, an experimental approach is required to examine how young children make use of particular sources of information in the process of figuring out the meanings of novel

words. For example, imagine a scene where a two-year-old visiting her relatives is served an unfamiliar fruit pastry after dinner, and her mother exclaims “What a surprise! Rhubarb!” How is the child to make sense of this remark? Studies exploring versions of this question now number in the hundreds (see Bloom, 2000; Woodward & Markman, 1997). Here we provide a very brief overview of research on some of the factors influencing young children’s interpretation of novel words.

Discussions of this research question often start by framing the problem in terms of Quine’s (1960) vignette of a linguist visiting an unfamiliar culture, accompanied by a native guide who does not speak his language. As a rabbit hops across the scene, the guide exclaims *Gavagai!* How, Quine asks, can the stranger possibly interpret this utterance, given that the speaker could be referring to the rabbit, to a part of the rabbit, to the animal’s action, or to an indeterminate number of other aspects of the scene? Quine does not provide an answer to this conundrum, emphasizing instead the fundamental impossibility of knowing the intended meaning based on the evidence at hand. In the language development literature, however, it is assumed that a young child in this situation would most likely rapidly and automatically interpret *gavagai* as *rabbit*, without considering the myriad other possibilities. Thus the inherent indeterminacy of meaning that was the focus of Quine’s argument is circumvented by the young language learner thanks to interpretative biases that guide early word learning. The idea that word learning gets started with help from some sort of “object-category bias” is supported by experiments in which children are asked to consider novel words in relation to novel objects. When an unfamiliar object such as a toy animal is labeled with an unfamiliar name (e.g., *ferret*), children typically assume that the new word refers to the animal as a whole, rather than its tail, its color, or the stuff it is made of (e.g., Markman & Hutchinson, 1984).

Although a bias for naming the whole object is predictably observed in experiments of this sort, there is considerable debate as to how this phenomenon should be interpreted. What factors lead the child to guess that the new word *ferret* is a name for the animal as a whole and other animals of the same kind, rather than its parts or properties? Three different approaches to this question have been discussed extensively: The first approach emphasizes the importance of *preverbal perception and cognition* in guiding word learning. Because objects are perceived as bounded and coherent and thus are salient even to infants (e.g., Spelke, 1998), we are predisposed from infancy on to see the world as containing cohesive objects. This could explain why children as well as adults are biased to identify (and to name) an object as a whole before attending to its parts and other attributes (Gentner, 1982). Moreover, nouns naming concrete objects are conceptually simpler than relational words like verbs and adjectives, which can vary substantially in the perceptual features they refer to depending on the nouns they are associated with (e.g., *a good cookie* vs. *a good dog*). According to Gentner and Boroditsky (2001), such non-linguistic aspects of human perception and cognition in relation to language structure could account for the tendency of infants to learn names for objects before they learn names for actions and attributes. A second approach to studying factors that guide early word learning emphasizes the critical role of *social cognition*, a perspective that has roots in Vygotsky’s (1962) theory of social support for learning and in Bruner’s (1975) views

on how reference first emerges in preverbal communication. Current research on the social origins of linguistic knowledge focuses on children's emerging awareness of the referential intentions of others in figuring out what unfamiliar words might refer to. For example, several studies have shown that when children hear a novel word, they will connect it with an appropriate object only if they somehow appreciate that the adult intended to name the object (e.g., Baldwin et al., 1996; Tomasello & Barton, 1994).

A third approach to understanding early word learning proposes *linguistic constraints* that account for children's biases in interpreting novel words. According to Markman's (1989) influential formulation of this position, such constraints are default assumptions that serve to limit hypotheses as to possible meanings for a new word. In particular, the *whole object assumption* guides the child to associate a new word with the entire object, while the *taxonomic assumption* guides the child to extend that new word to other objects belonging to the same class rather than to thematically related objects. Another proposed word learning constraint is the *mutual exclusivity assumption*, a kind of exclusionary learning strategy that has recently been demonstrated in studies with infants as young as 15 months of age (Halberda, 2003; Markman, Wasow, & Hansen, 2003). When a young child is presented with a familiar object with a known name (e.g., a ball) along with an unfamiliar object (e.g., a whisk) and is then asked to find the *dax*, the typical response is to choose the whisk. One interpretation of this effect is that the child automatically maps the novel word onto the novel object rather than assigning a second name to the ball, guided by the default assumption that an object can only have a single name (e.g., Markman & Wachtel, 1988). Other researchers disagree that this effect is specific to word learning, arguing instead that it is grounded in pragmatic knowledge (see Bloom, 2000; Clark, 2003b). For example, Clark (1997) argues that children's bias against lexical overlap is best explained in terms of a "principle of contrast" which leads them to assume that differences in form should correspond to differences in meaning. It is interesting to note that even dogs show a related form of exclusionary learning, mapping novel spoken words onto objects for which no name has previously been learned (Kaminski, Call, & Fischer, 2004). However, although this finding suggests that a learning principle based on mutual exclusivity might not be specific to human language, it is also clear that animal learning of word-object associations differs in important ways from lexical learning by children (Bloom, 2004; Markman & Abelev, 2004).

The robust learning biases demonstrated in these experiments are certainly consistent with the notion that children rely on strategies specifically adapted for lexical learning. However, the view that such learning biases are automatic and language-specific has come under criticism from many directions. Children's early vocabularies do not consist only of nouns and include types of words (*Hi! Up! More!*) quite different from the object names supposedly favored by lexical constraints (Nelson, 1973). And for some first words (e.g., *bath*) that are technically nouns, it is not at all clear whether they are understood by the child as an object or an action, or as a routine involving both. Although word learning constraints are proposed as a solution to the problem characterized in Quine's (1960) dilemma, many scholars of early language learning (e.g., Bloom, 1993, 2000; Clark, 2003a, 2003b; Nelson, 1988; Tomasello, 2003) point out that the young child is

not really comparable to the linguist who speaks a first language different from that of the native guide, and that the guide is not at all like the parent of a small child. In the ecology of early parent–child interaction, the adult takes the perspective of the linguistic novice in many ways, providing both simplified language input (e.g., Lieven, 1994) and many pragmatic cues to reference using gaze and gesture that young children are able to use in figuring out what new words mean (Tomasello, 2003).

Another way in which Quine’s example of linguistic indeterminacy has been oversimplified as a model of early word learning is that novel objects do not always hop across the scene as the single most salient focus of attention. Imagine a variation of this scenario: the native guide fires his gun as he yells *Gavagai!* and the rabbit falls down dead. Under these circumstances, the linguist as well as the child would presumably not default automatically to a static whole-object interpretation of this string of sounds, but might assume *Gavagai!* meant something like *Watch out!* or *Got it!* or *Dinner!* The everyday situations in which word learning occurs are often much more dynamic and ambiguous than the experimental setups in which word learning is studied scientifically. In the example above in which the mother exclaims “Rhubarb!” as the dessert is placed on the table in front of her 2-year-old, the child will only gradually figure out what this word refers to, learning based first on the taste and texture of the cooked fruit and only later on other properties of rhubarb as a plant. Does this mean that a mapping error will occur, as the child automatically attempts to apply the new word to the pastry itself as a whole object? It seems more likely that the child will pick up on social cues indicating that the mother’s remark is addressed to others at the table and is not intended as a label, one of several reasons a word may simply be ignored on first exposure. After all, infants hear thousands of words in a week yet learn to use only one or two new words a day.

Although critics of the linguistic constraints position may object that the point of Quine’s (1960) example has been distorted in the developmental literature, they agree that young children face a daunting inductive problem in assigning meaning to an unfamiliar sequence of speech sounds. But it is possible to agree that children need to limit the potentially large number of hypotheses for word meanings without assuming either that this kind of inductive problem is unique to word learning, or that constraints in the form of default assumptions are the only way to solve the problem. Bloom (2000) points out that children face comparably complex inductive problems in other domains of experience all the time. When a child grabs the handle of an iron skillet sitting on a hot stove burner, she has to figure out what to avoid in the future in order not to get burned again. Is it the skillet or just the handle? Or could it be anything shiny and white like the stove? In this case, avoidance of the skillet itself might reflect a non-linguistic whole-object bias, a reasonable first guess until the child developed a deeper understanding of the causal processes involved. In another example of a proposed word-learning constraint that may be much more general in its scope, Markson and Bloom (1997) show that the phenomenon of “fast mapping” a novel word to a novel object is not limited to lexical learning. When children hear a novel word described as a *koba*, they remember which object the new word referred to; however, when they hear a novel object described as “the one my uncle gave me,” they are equally good at remembering which object the *fact*

referred to. Bloom and Markson (2001) argue that many of the findings about how novel words are extended are best explained in terms of general cognitive systems such as those involved in concept formation and intentional inference, rather than through proposed linguistic constraints that are dedicated uniquely to word learning.

In an early critique of the theory that linguistic constraints are essential for learning new words, Nelson (1988) pointed out that the first formulations of this position traced their intellectual roots to Chomsky's (1975) claims for innate mechanisms for learning grammar, extending this framework to lexical learning. This perspective is consistent with an emphasis on default learning strategies that privilege some sorts of information and are impervious to others. Just as some nativist theories of adult language comprehension posit autonomous parsing strategies favoring syntactic structure over all other kinds of linguistic information (e.g., Frazier & Rayner, 1987), the initial emphasis in constraints theory was on how children are restricted by strong, possibly innate, default assumptions as to what a new word might mean (e.g., Markman & Hutchinson, 1984). In more recent accounts of the linguistic constraints position, these word learning strategies are framed as more flexible heuristics, i.e., as somewhat "softer" constraints (Woodward & Markman, 1997). But the focus is still on how children are inherently limited in their interpretative strategies, rather than on how they may integrate different sources of information in different contexts.

As mentioned earlier, the idea of inflexible parsing strategies in adult comprehension has been challenged by many new studies showing how listeners integrate probabilistic information from multiple sources (Seidenberg, 1997). In research on early word learning as well, there is mounting evidence that infants use diverse sources of linguistic and non-linguistic information in making sense of new words, guided by learning biases that are construed more appropriately as preferences than as constraints (e.g., Bloom, 1993; Hollich, Hirsh-Pasek, & Golinkoff, 2000). As Nelson (1988) put it, "The connotation is quite different: Constraints imply restriction – a closing down of choice; whereas preference implies free, but biased, choice" (p. 228). These new models of early word learning draw on insights articulated years ago in the "competition model" of Bates and MacWhinney (1979, 1987, 1989), namely that multiple sources of information are available to the young language learner and that the influence of different information sources varies both as a function of their strength as cues in relation to other cues, and also as a function of the developmental level of the child. Research on word learning is just beginning to investigate the relative contributions of multiple cues on novel word interpretation by children at different ages. For example, many studies have shown how the shape of a novel object influences children's categorization and naming. Although in their everyday experience, children often experience new objects in motion surrounded by other objects, almost all experiments on the "shape bias" have used isolated static objects as experimental stimuli. However, when Smith (2005) presented 2-year-olds with dynamic stimuli, she found that movement influenced children's judgments as to which objects were similar. Findings like this will lead developmental researchers increasingly toward a different formulation of the question, one more in line with the probabilistic constraints approach to investigating language comprehension by adults (Seidenberg & MacDonald,



1999). To get beyond the question “Is there or is there not a shape bias?” (e.g., Cimpian & Markman, 2005), studies will begin to investigate when and under what circumstances shape is an important factor in object categorization, and how shape interacts with other perceptual features as well as linguistic and social cues in the referential context in guiding the child’s inference as to what a new word refers to.

## **6. LISTENING FOR MEANING IN SPEECH IN THE SECOND YEAR**

Infants’ early progress in developing language is often charted in terms of their increasing competence in understanding, producing, and learning individual words, an ability that is arguably shared in some respects with other species (Kaminski, Call, & Fischer, 2004; Seidenberg & Petitto, 1979). However, the ability to understand and use words flexibly in combination is a critical distinguishing feature of human language. Of course, it was Chomsky (1959) who pointed out long ago that multi-word sentences are much more than individual words strung together one-by-one. Grammatical sentences are made up of units of words that vary in size and are organized hierarchically in a large, but finite, set of complex ways that are not always obvious from the surface-level ordering of the words. While traditional views held that the language-learning child must be innately endowed with such grammatical knowledge, recent perspectives have continued to examine ways in which such proficiency can be constructed over the course of development. In this section, we review recent research on how infants develop impressive efficiency in understanding words in continuous speech across the second year, and on how they begin to use words in combination to express increasingly complex meanings through language.

### **6.1. The Development of Efficiency in Language Understanding**

To make sense of the rapidly spoken strings of words that make up the language children hear, they must learn to process fluent speech efficiently, “listening ahead” to anticipate what is coming next in the speech stream using different sources of linguistic and non-linguistic information. Many recent studies using on-line measures of comprehension with adults have shown that skilled listeners draw on multiple sources of knowledge to process speech with remarkable speed and efficiency (e.g., Tanenhaus et al., 1995). With the refinement of eye-tracking techniques for use with infants, it is now possible to monitor the time course of spoken language understanding by very young language learners as well. Using a looking-while-listening procedure with English-learning infants from 15 to 24 months of age Fernald, Pinto, Swingley, Weinberg and McRoberts (1998) found dramatic gains in the speed and accuracy of word recognition over the second year. In this procedure, infants look at pictures of familiar objects while listening to speech naming one of the objects. Fifteen-month-olds responded inconsistently and shifted their gaze to the appropriate picture only after the offset of the target word, while 24-month-olds were faster and more reliable, initiating a shift in gaze before the target word had been completely spoken. A recent longitudinal study following infants from 12 to 25 months found that on-line measures of efficiency in speech processing were correlated with numerous more traditional measures of lexical

and grammatical development (Fernald et al., 2006). Moreover, analyses of growth curves showed that children who were faster and more accurate in on-line comprehension at 25 months were those who showed faster and more accelerated growth in expressive vocabulary across the second year. Success at word recognition in degraded speech is also correlated with vocabulary size in the second year (Zangl et al., 2005), further evidence that speech processing efficiency is related to other dimensions of early language development.

A possible benefit of the increase in processing efficiency over the second year is that it enables infants to identify words more quickly based on partial phonetic information, rather than waiting until the word is complete. However, one consequence is that the young language learner is increasingly confronted with problems of temporary ambiguity. When Allopenna et al. (1998) presented adults with objects that included candy and a candle and asked them to *Pick up the can-*, they waited to hear the next speech sound before orienting to the appropriate object. That is, they postponed their response until the final syllable of the target word made it clear which object was the intended referent. The child who hears *Where's the doll?* in the presence of a doll and a dog is also faced with a temporary ambiguity, given that *doll* and *dog* overlap phonetically and thus are indistinguishable for the first 300 ms or so. Swingley, Pinto, and Fernald (1999) found that 24-month-olds in this situation also delayed their response by about 300 ms until disambiguating information became available. Even when they heard *only* the initial phonemes in familiar words (e.g., the isolated first syllable of *baby* or *kitty*), 18-month-olds were able to use this limited information to identify the appropriate referent (Fernald, Swingley, & Pinto, 2001). Further evidence for early use of phonetic information in a probabilistic fashion comes from studies by Swingley and Aslin (2000, 2002) showing that even younger infants can identify familiar words when they are mispronounced, but respond more strongly to the correct than to the incorrect version (e.g., *baby* vs. *vaby*).

Children also become increasingly attentive to prosodic and morphosyntactic regularities in speech that enable them to anticipate upcoming content words in the sentence, also relying on probabilistic information (Fernald & Hurtado, 2006). For example, 2-year-olds expect an object name to follow an unstressed article (Zangl & Fernald, under review). When an uninformative adjective occurs instead (e.g., *Where's the pretty CAR?*), they “listen through” the prenominal word and wait for the noun before responding; however, if the adjective is novel and accented, they are more likely to misinterpret the unknown word as a potential object name (Thorpe & Fernald, 2006). That is, when the word preceding the target name is stressed as well as lexically ambiguous (e.g., *Where's the ZAV car?*) it becomes relatively more noun-like, and 26-month-olds are more likely to respond accordingly by searching for a novel referent as soon as they hear *zav*, rather than waiting for the subsequent word that names the target object. This tendency of English-learning 2-year-olds to “false alarm” in response to stressed novel words preceded by the article *the* shows that they are integrating multiple probabilistic cues to predict what kind of word is coming next. Such studies using on-line measures of children's comprehension as the spoken sentence unfolds reveal a critical dimension of emerging language competence that was impossible to monitor with precision using

off-line methods. As children learn to interpret words in combination, they develop efficiency in integrating distributional, lexical, prosodic, and other available sources of information, enabling them to make sense of words that are known while avoiding costly interference from unfamiliar words in the sentence that are not yet known.

## 6.2. Emerging Awareness of Relations among Words

By two years of age, most children are demonstrating impressive skill at interpreting the speech that they hear around them. Several studies using preferential listening techniques (e.g., Gerken, Wilson, & Lewis, 2005; Gomez, 2002) as well as neurophysiological responses (e.g., Friederici, 2005) show that children in the second year are increasingly attentive to regularities in speech that are relevant to the grammatical structure of the language they are learning. They have also built up a considerable repertoire of words in their production vocabularies, and are beginning to use two- or three-word combinations (e.g., *mommy sock*). Soon, however, utterances increase in length and complexity in various ways. Children add more verbs, adjectives, and other predicates to their working vocabularies, and substantively increase their use of prepositions, articles, and other closed-class forms that do grammatical work, including the productive use of inflectional morphemes (e.g., English past tense *-ed*). At the same time, there is also sizeable variation in exactly when and how children move into more grammatically complex utterances in their every day language use. Indeed, while some children are reported to use primarily multi-word phrases and many closed-class forms by 24 months, other children are still primarily using nouns in single-word constructions (e.g., Bates et al., 1988; Bates & Goodman, 1999).

Who are the children who are more advanced in grammar at this age? Based on the norming data from the *CDI: Words and Sentences*, children with the highest grammar scores were also those children with the largest reported production vocabularies ( $r = 0.85$ ) (Bates et al., 1994). In the same data set, Marchman and Bates (1994) found that size of verb vocabulary was concurrently related to the number of reported overregularizations of the English past tense inflection (e.g., *daddy goed*), accounting for significant variance over and above chronological age. These “mistakes” are typically viewed as a major milestone in the development of grammatical rule-based knowledge. Links between lexical development and grammar have also been reported longitudinally. Following 27 children, Bates et al. (1988) found that the best predictor of grammatical sophistication at 28 months (as measured by mean length of utterance, MLU) was size of vocabulary 10 months earlier. Bates and Goodman (1997) cite similar relationships in a sample of children followed monthly from 12 to 30 months.

Other researchers have targeted children at the extremes in acquisition (e.g., late vs. early talkers), revealing that children who were delayed in early vocabulary production were later delayed in the use of grammatical forms (Paul, 1996, 1997; Rescorla & Schwartz, 1990; Rescorla, Roberts, & Dahlsgaard, 1997, 2000; Thal & Tobias, 1994; Thal & Katich, 1996; Marchman & Armstrong, 2003) and that particularly precocious

children display grammar abilities that are commensurate with their vocabulary, even though they are considerably younger than peers at their same level (Thal, Bates, Zappia, & Oroz, 1996; Thal, Bates, Goodman, & Jahn-Samilo, 1997). Similar lexical–grammar links have been found in children with focal brain injury (e.g., Bates et al., 1997; Marchman, Miller, & Bates, 1991; Marchman, Wulfeck, & Saccuman, 2003; Thal et al., 1991), and Williams syndrome (e.g., Singer-Harris, Bellugi, Bates, Jones, & Rossen, 1997). More recently, studies have documented that lexical development and grammar are related to a similar degree in children learning more than one language, with grammatical abilities robustly linked to lexical level in the same, but not the other, language (Marchman, Martínez-Sussmann, & Dale, 2004). Finally, strong heritability of the *relation* between lexical and grammatical level has been documented in behavioral genetic studies of monozygotic and dizygotic twins (Dale, Dionne, Eley, & Plomin, 2000). In other words, even though genetic factors make a relatively weak contribution to each aspect of language assessed individually, the genetic factors that influence lexical growth are the same as those that influence grammatical growth.

These studies all point to the idea that vocabulary and grammar development are highly interdependent, a view at odds with the nativist assumption that grammatical knowledge is autonomous and emerges independent of lexical knowledge. In light of the extensive individual variation that is observed in early language development, it is striking that lexical and grammatical skill “hang together” so tightly over acquisition, especially when abilities that would seem to more likely to travel together are less strongly related (e.g., reported lexical comprehension and production). Such interdependence is quite natural, however, within a view of acquisition in which domain-general learning mechanisms guide the child’s construction of a working linguistic system simultaneously at many different levels, in this case, learning words and learning grammatical rules. As Bates and MacWhinney (1987) proposed many years ago, “the native speaker learns to map phrasal configurations onto propositions, using the *same learning principles and representational mechanisms* needed to map single words onto their meanings” (p. 163, emphasis added). This type of domain-general continuity is directly modeled in connectionist and dynamical systems accounts of language development (e.g., Plunkett & Marchman, 1993; Elman et al., 1996; van Geert, 1998), and is at the core of probabilistic constraint-based explanations of many other psycholinguistic and developmental phenomena (e.g., Elman et al., 1996; Elman, Hare & McRae, 2005; Harm & Seidenberg, 2004; Tomasello, 2003). Interestingly, enhanced reliance on domain-general continuity has gained credibility in several frameworks in modern-day linguistics (e.g., Bresnan, 2001; Croft, 2001; Goldberg, 1995; Langacker, 1987). Finally, several recent studies have focused on ruling out indirect explanations for lexical–grammar links, for example, that lexical and grammatical relations derive from common influences from the environment or general cognitive or linguistic intelligence (e.g., Dale et al., 2000; Dionne, Dale, Boivin, & Plomin, 2003; Marchman et al., 2004).

Clearly, there is much more to be said about early vocabulary and grammar development. Studies are continuing to map out in more and more precise ways how those domain-general mechanisms might operate (e.g., Bates & Goodman, 1999; Tomasello,

2003; Elman, 2004; Naigles, 1996), and how those relations might change in character over development (Dionne et al., 2003; Tomasello, 2003). Yet, the picture that is gaining mounting empirical support portrays language acquisition as a gradual and continual process of mapping various types of linguistic entities onto communicative functions, using mechanisms that are shared across many different levels of the linguistic system.

## 7. CONCLUSION

Researchers who cross-cut the fields of cognitive science, developmental psychology, and psycholinguistics, have directly questioned key premises of the standard view of language acquisition (e.g., Bates & Goodman, 1999; Tomasello, 2003). Instead of a nativist and modular view, these researchers prefer to characterize language acquisition as a process of coordinating and integrating cognitive, linguistic and communicative information in the context of interaction with the physical and social world. Such “usage-based,” approaches (e.g., Tomasello, 2001, 2003; Goldberg, 1995; Bates & MacWhinney, 1987, 1989; Elman, 2004) generally adopt the idea that the complex and intricate linguistic knowledge that children have emerges gradually over the course of human interaction. The constraints that guide the building of linguistic systems are not necessarily specific to the task of language acquisition, but reflect a set of general learning mechanisms that come together in ways that are particularly good at building grammars from the speech that children hear. These perspectives can be traced back to somewhat different intellectual roots than those guiding the standard nativist view. For example, the interactionist perspectives of Piaget and Bruner (Piaget, 1952; Bruner, 1983) and principles derived from the modern incarnations incorporated in connectionist (e.g., Elman et al., 1996) or dynamical systems approaches (e.g., Thelen & Smith, 1994; Smith & Thelen, 1993) provide key theoretical and methodological tools. Through these perspectives the mechanisms of language learning are envisioned as considerably more interactive, incremental, and powerful than previously thought. In addition, the construction grammar or usage-based approaches in linguistics (e.g., Langacker, 1987; Givón, 1997) have offered a characterization of adult linguistic competence that is considerably more probabilistic, piecemeal, and “child centered” than traditional linguistic approaches (e.g., Seidenberg & MacDonald, 1999; Elman, 2004). Over the last several years, empirical studies demonstrating the remarkable abilities that young language learners bring to the task of acquisition have continually provided new insights into the powerful set of processing and representational mechanisms that characterize human cognition. Researchers have also embraced the study of individual differences and crosslinguistic research as ways to expand our view of how these mechanisms come into play during the real-time task of language learning (Slobin, 1997), i.e., how the child manages to “get there from here.” These perspectives provide the logical and empirical basis for shifting the focus of study from a child who learns words and builds a grammar in relative isolation to a child whose early life is filled with rich and diverse language-learning experiences that may vary in important ways from child to child, from family to family, and from linguistic community to linguistic community. As in other areas of current research on language processing and use by fluent adults, new research on language acquisition is moving away from the assumption that relatively

inflexible default processing strategies are the only way to account for the amazing complexity of early language learning. This brief review of research on how infants begin to make sense of speech sounds over the first two years of life has focused on emerging perspectives in the field. The skill of the human infant in integrating cues from multiple sources of probabilistic information is evident at every developmental level, as the infant first discerns regularities in patterns of meaningless sounds, then begins to appreciate words and their meanings and to build up more complex meanings through understanding and using words in combination.

## REFERENCES

- Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.
- Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, *67*, 3135–3153.
- Baldwin, D. A., & Moses, L. J. (1996). The ontogeny of social information gathering. *Child Development*, *5*, 1915–1939.
- Barrett, M. D. (1986). Early semantic representations and early word usage. In: S. A. Kuczaj, & M. D. Barrett (Eds), *The development of word meaning*. New York: Springer.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, I., & Volterra, V. (1979). *The emergence of symbols*. New York: Academic Press.
- Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Bates, E., Dale, P., & Thal, D. (1994). Individual differences and their implications for theories of language development. In: P. Fletcher, & B. MacWhinney (Eds), *Handbook of child language*. Oxford: Blackwell.
- Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *12*(5/6), 507–586.
- Bates, E., & Goodman, J. (1999). On the emergence of grammar from the lexicon. In: B. MacWhinney (Ed.), *The emergence of language* (pp. 29–79). Mahwah, NJ: Lawrence Erlbaum.
- Bates, E., & MacWhinney, B. (1979). A functionalist approach to the acquisition of grammar. In: E. Ochs, & B. Schieffelin (Eds), *Developmental pragmatics* (pp. 167–209). New York: Academic Press.
- Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In: B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum, 157–194.

- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In: B. MacWhinney, & E. Bates (Eds), *The crosslinguistic study of sentence processing* (pp. 3–76). New York: Cambridge University Press.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21(1), 85–124.
- Bates, E., Thal, D., Trauner, D., Fenson, J., Aram, D., Eisele, J., & Nass, R. (1997). From first words to grammar in children with focal brain injury. In: D. Thal, & J. Reilly (Eds), *Special issue on Origins of Communication Disorders, Developmental Neuropsychology*, 13(3), 275–343.
- Benedict, H. E. (1979). Early lexical development: comprehension and production. *Journal of Child Language*, 10, 183–200.
- Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, 117, 21–33.
- Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. Cambridge, MA: MIT Press.
- Bloom, L. (1993). *The transition from infancy to syntax: Acquiring the power of expression*. New York, NY: Cambridge University Press.
- Bloom, L., Lightbown, L., & Hood, L. (1975). Structure and variation in child language. *Monographs for the Society for Research in Child Development*, (Vol. 40, Serial No. 160).
- Bloom, P. (2000). *How children learn the meanings of words: Learning, development, and conceptual change*. Cambridge, MA: The MIT Press.
- Bloom, P. (2004). Can a dog learn a word? *Science*, 304(5677), 1605–1606.
- Bloom, P., & Markson, L. (2001). Are there principles that apply only to the acquisition of words? A reply to Waxman and Booth. *Cognition*, 78, 89–90.
- Bornstein, M. H., & Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities, *Infancy*, 7(3), 299–316.
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. -Y., Pascual, L. P., (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75(4), 1115–1139.
- Braine, M. (1994). Is nativism sufficient? *Journal of Child Language*, 21(1), 9–31.
- Braine, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, (Vol. 41, Serial No. 164).
- Brent, M. R., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1–2), 93–125.

- Bresnan, J. (2001). *Lexical functional syntax*. Malden, MA: Blackwell.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In: J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.
- Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2, 1–19.
- Bruner, J. (1983). *Child's talk*. New York: Norton.
- Bybee, J. (1995). Diachronic and typological properties of morphology and their implications for representation. In: L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 225–246). Hillsdale, NJ: Erlbaum
- Caselli, M. C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, 26, 69–111.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experiences. *Cognition*, 87(2), B69–B77.
- Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, M., Renlund, M., Aaltonen, O., Eerola, O., & Naatanen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, 82, 53–58.
- Cheyney, D. L., & Seyfarth, R. M. (1997). Why animals don't have language. *Tanner lectures on human values*, New York: Cambridge University.
- Choi, S., & Bowerman, M. (2001). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41(1–3), 83–121.
- Choi, S., & Gopnik, A. The acquisition of verbs in Korean: A crosslinguistic study. *Journal of Child Language*, 22, 497–529.
- Chomsky, N. (1959). Review of B. F. Skinner's *Verbal Behavior*. *Language*, 35, 16–58.
- Chomsky, N. (1975). *Reflections on language*. New York: Pantheon Books.
- Chomsky, N. (1981). *Lectures on government and binding*. New York, NY: Foris.
- Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30(3), 637–669.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13(2–3), 221–268.
- Cimpian, A., & Markman, E. M. (2005). The absence of a shape bias in children's word learning. *Development Psychology*, 41(6), 1003–1019.
- Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, 64(1), 1–37.



- Clark, E. V. (2003a). Language and representations. In: D. Gentner, & S. Goldin-Meadow (Eds), *Language in mind: Advances in the study of language and thought* (pp. 17–24). Cambridge, MA: MIT Press.
- Clark, E. V. (2003b). *First language acquisition*. New York, NY: Cambridge University Press.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, 96(3), 233–262.
- Dale, P. S., Dionne, G., Eley, T. C., & Plomin, R. (2000). Lexical and grammatical development: A behavioral genetic perspective. *Journal of Child Language*, 27, 619–642.
- Dapretto, M., & Bjork, E. L. (2000). *The development of word retrieval abilities in the second year and its relation to early vocabulary growth* *Child Development*, 71(3), 635–648.
- Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development*, 74(2), 394–412.
- Dromi, E. (1987). *Early lexical development*. New York: Cambridge University Press.
- Eimas, P. D. & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4(1), 99–109.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303–306.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71–99.
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 7, 301–306.
- Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences*, 9(3), 112–117.
- Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press/Bradford Books.
- Elman, J. L., Hare, M., & McRae, K. (2005). Cues, constraints, and competition in sentence processing In: M. Tomasello, & D. I. Slobin (Eds), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111–138). Mahwah, NJ: Erlbaum.
- Fernald, A. (2002). *How infants develop expectations about what's coming next in speech*. Paper presented at the 15th annual CUNY conference on human sentence processing, New York, NY, March.

- Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9(3), F33–F40.
- Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, 64, 637–656.
- Fernald, A., & McRoberts, G. W. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In: J. Morgan, & K. Demuth (Eds), *From signal to syntax*. Hillsdale, NJ: Erlbaum.
- Fernald, A., McRoberts, G. W., & Swingley, D. (2001). Infants' developing competence in understanding and recognizing words in fluent speech. In: J. Weissenborn, & B. Hoehle (Eds), *Approaches to bootstrapping in early language acquisition*. Benjamins: Amsterdam.
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116.
- Fernald, A., Pinto, J., Swingley, D., Weinberg, A., & McRoberts, G. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9, 72–75.
- Fernald, A., Swingley, D., & Pinto, J. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, 72(4), 1003–1015.
- Fenson, L., Dale, P. A., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. J. (1994). Variability in early communicative development. *Monograph of the Society for Research in Child Development*, (Serial No. 242, Vol. 59, 5).
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.
- Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Frazier, L. (1987). Structure in auditory word recognition. *Cognition. Special Issue: Spoken word recognition*, 25(1-2), 157–187.
- Frazier, L. & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, 26, 505–526.
- Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Sciences*, 9, 481–488.
- Friedlander, B. (1970). Receptive language development in infancy: Issues and problems. *Merrill-Palmer Quarterly*, 16, 7–52.
- Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, 40, 621–632.

- Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes. Special Issue: Language Acquisition and connectionism*, 13(2–3), 269–306.
- Genter, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In: M. Bowerman, & S. Levinson (Eds), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, UK: Cambridge University Press.
- Gentner, D., (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In: S. A. Kuczaj II (Ed.), *Language development: Vol. 2, Language, thought, and culture*. Hillsdale, NJ: Erlbaum.
- Gerken, I., Wilson, R., & Lewis, W. (2005). Infant can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249–268.
- Givón, T. (1997). Grammatical relations: An introduction. In: T. Givón (Ed.), *Grammatical relations: A functional perspective*. Amsterdam: John Benjamins.
- Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. In: F. J. Newmeyer (Ed.), *Language: Psychological and biological aspects*. Cambridge: Cambridge University Press.
- Gold, E. M. (1967). Language Identification in the limit. *Information and Control*, 16, 447–474.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldfield, B., & Reznick, S. (1990). Early lexical acquisition: Rate, content and the vocabulary spurt. *Journal of Child Language*, 17, 171–183.
- Goldfield, B., & Snow, C. (1985). Individual differences in language acquisition. In J. Gleason (Ed.), *Language development*. Columbus, OH: Merrill Publishing Co.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431–436.
- Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, 58(6), 1523–1531.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87, B23–B34.
- Halle, P. E., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, 17(2), 119–129.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111(3), 662–720.
- Harris, P., Barrett, M. D., Jones, D. & Brookes, S. (1988). Linguistic input and early word meaning. *Journal of Child Language*, 15, 77–94.

- Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: P. H. Brookes.
- Hertwig, R., & Todd, P. M. (2003) More is not always better: The benefits of cognitive limits. In D. Hardman, & L. Macchi (Eds), *Thinking: Psychological perspectives on reasoning, judgment and decision making*. (pp. 213–231) West Sussex: Wiley & Sons.
- Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monograph of the Society for Research in Child Development*, (Vol. 65, Serial No. 262).
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74, 1368–1878.
- Hoff, E., & Naigles, L. (2002). How children use input in acquiring a lexicon. *Child Development*, 73, 418–433.
- Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, 19, 603–629.
- Huttenlocher, J., Haight, W., Bryk, A., & Seltzer, M. (1991) Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236–248.
- Huttenlocher, J. & Smiley, P. (1987). Early word meanings: The case of object names. *Cognitive Psychology*, 19, 63–89.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45, 337–374.
- Jackson-Maldonado, D., Thal, D. J., Marchman, V., Newton, T., Fenson, L., & Conboy, B. (2003). *El Inventario del Desarrollo de Habilidades Comunicativas: User's guide and technical manual*. Baltimore, MD: Brookes Publishing Co.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548–567.
- Jusczyk, P. W. (1998). Constraining the search for structure in the input. *Lingua*, 106, 197–218.
- Jusczyk, P. (1997). Finding and remembering words: Some beginnings by English-learning infants. *Current Directions in Psychological Science*, 6(6), 170–174.
- Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29 (1), 1–23.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64(3), 675–687.
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory & Language*, 32, 402–420.

- Jusczyk, P. W., Rosner, B. W., Cutting, J. E., Foard, F., & Smith, L. B. (1977). Categorical perception of non-speech sounds by two-month-old infants. *Perception and Psychophysics*, *21*, 50–54.
- Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for “fast mapping”. *Science*, *304*(5677), 1682–1683.
- Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55–68.
- Kersten, A. W. & Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, *44*, 250–273.
- Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*, 1168–1179.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–841.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lieven, E. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In: C. Galloway, & B. Richards (Eds), *Input and interaction in language acquisition*. Cambridge: Cambridge University Press.
- Lieven, E., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, *24*(1), 187–219.
- MacWhinney, B. & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language*, *17*, 457–472.
- Maitel, S. L., Dromi, E., Sagi, A., & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, *27*, 43–67.
- Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants’ recognition of the sound patterns of their own names. *Psychological Science*, *6*(5), 314–317.
- Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children’s syntax: The ontogenesis and representation of syntactic categories. In: K. Nelson (Ed.), *Children’s Language: (Vol. 2, pp. 128–114)*. New York: Gardner Press.

- Marchman, V., & Armstrong, E. (2003). Productive use of inflectional morphology by “late-blooming” and typically-developing toddlers: A follow-up at 3½ years. Poster presented at the biennial meeting of the Society for Research in Child Development (SRCD), Tampa, FL, April.
- Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21(2), 339–366.
- Marchman, V., & Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research*, 45(5), 983–997.
- Marchman, V., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, 7(2), 212–224.
- Marchman, V., Miller, R., & Bates, E. (1991). Babble and first words in infants with focal brain injury. *Applied Psycholinguistics*, 12(1), 1–22.
- Marchman, V., Wulfeck, B., & Saccuman, C. (2003). Productive language use in children with early focal brain injury and specific language impairment. *Brain and Language*, 88(2), 202–214.
- Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46(1), 53–85.
- Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.
- Markman, E. M., & Abelev, M. (2004). Word learning in dogs? *Trends in Cognitive Sciences*, 8, 479–481.
- Markman, E. M., & Hutchinson, J. E. (1984). Children’s sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, 16, 1–27.
- Markman, E. M., & Wachtel, G. F. (1988). Children’s use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, 20(2), 121–157.
- Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47, 241–275.
- Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385, 813–815.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477–500.
- Mervis, C., & Tomasello, M. (1994). The instrument is great, but measuring comprehension is still a problem. In: L. Fenson, P. A. Dale, J. S. Reznick, E. Bates, D. Thal, & S. J. Pethick (Eds). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, (Serial No. 242, Vol. 59, (5).

- Mervis, C. B., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: A response to Goldfield & Reznick. *Journal of Child Language*, 22(2), 461–468.
- Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117.
- Morgan, J., Shi, R., & Allopenna, P. (1996). Perceptual bases or rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In: J. L. Morgan, & K. Demuth (Eds). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Hillsdale, NJ: Erlbaum Associates, Inc.
- Morgan, J. L., & Demuth, K. (Eds). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. xi, 487pp. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*, 74(1), 221–237.
- Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, 6(4), 413–429.
- Naigles, L. G. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58(2), 221–251.
- Naigles, L. G., & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: Preferential-looking in a study of dog, cat, and cow. *Journal of Child Language*, 22(1), 19–46.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, (Vol. 38, Serial No. 149).
- Nelson, K. (1988). Constraints on word learning? *Cognitive Development*, 3, 221–246.
- Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In: E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.
- Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, 76, 763–782.
- Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology*, 5, 5–21.
- Paul, R. (1997). Understanding language delay: A response to van Kleeck, Gillam, and Davis. *American Journal of Speech-Language Pathology*, 6, 40–49.
- Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43, 93–120.
- Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language*, 53, 560–573.

- Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Piaget, J. (1952). *The origins of intelligence in children*. New York: International University Press.
- Piaget, J. (1962). *Play, dreams and imitation in childhood*. New York: Norton.
- Pine, J. M., & Lieven, E. (1990). Referential style at 13 months: Why age-defined cross-sectional measures are inappropriate for the study of strategy differences in early language development. *Journal of Child Language*, 17, 625–631.
- Pinker, S. (1979). Formal models of language learning. *Cognition*, 7, 217–283.
- Pinker, S. (1991). Rules of language. *Science*, 253, 530–535.
- Pinker, S. (1994). *The language instinct: how the mind creates language*. New York: Harper Collins.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.
- Pinker, S. (2003). *The Blank Slate: The modern denial of human nature*. New York, NY: Viking.
- Plunkett, K. (1992). The segmentation problem in early language acquisition. *Journal of Child Language*, 20(1), 43–60.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43–102.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 21–69.
- Plunkett, K., Sinha, C. G., Möller, M. F., & Strandsby, O. (1992) Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4, 293–312.
- Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review*, 19(1–2), 9–50.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Rescorla, L. (1980). Overextension in early language development. *Journal of Child Language*, 7, 321–335.
- Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, 54(4), 587–599.
- Rescorla, L., & Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics*, 11, 393–407.
- Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research*, 40, 556–566.



- Rescorla, L., Roberts, J., & Dahlsgaard, K. (2000). Late talkers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language*, 27, 643–664.
- Rochat, P. (Ed.). (1999). *Early social cognition: Understanding others in the first months of life*. Mahwah, NJ: Erlbaum.
- Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72(1), 67–109.
- Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.
- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484–494.
- Saffran, J. R., Werker, J. F., & Werner, L. A. (2006) The infant's auditory world: hearing, speech, and the beginnings of language. In: R. Siegler, & D. Kuhn (Eds), *Handbook of Child Psychology*, Vol. 2: *Cognition, Perception and Language* (6th ed.). New York: Wiley.
- Schieffelin, B., & Ochs, E. (1986). *Language socialization across cultures*. New York: Cambridge University Press.
- Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275(5306), 1599–1603.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23(4), 569–588.
- Seidenberg, M. S., & Pettito, L. A. (1979). Signing behaviour in apes. *Cognition*, 7, 177–215.
- Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.
- Shi, R., Werker, J., & Morgan, J. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72, B11–B21.
- Shi, R., & Werker, J. F. (2001). Six-month old infants' preference for lexical words. *Psychological Science*, 12(1), 70–75.
- Singer-Harris, N., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting profiles of language development in children with Williams and Down syndromes. In: D. Thal, & J. Reilly (Eds), *Developmental Neuropsychology*, 13(3), 45–370.
- Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson, & D. I. Slobin (Eds), *Studies of child language development*. New York: Holt, Rinehart & Winston.
- Slobin, D. I. (1997). *The crosslinguistic study of language acquisition (Vol. 5): Expanding the contexts*. Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, L. B. (2005) Action alters shape categories. *Cognitive Science*, 29, 665–679.

Smith, L. B., & Thelen, E. (2003) Development as a dynamic system. *TRENDS in Cognitive Science*, 7, 343–348.

Snow, C., & Ferguson, C. (1977). *Talking to children: Language input and interaction*. Cambridge, MA: Cambridge University Press.

Sokolov, J., & Snow, C. (1994). The changing role of negative evidence in theories of language development. In: C. Gallaway, & B. Richards (Eds), *Input and interaction in language acquisition*. Cambridge, MA: Cambridge University Press.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior & Development*, 21(2), 181–200.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50(1), 86–132.

Swingley, D., & Aslin, R. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.

Swingley, D., & Aslin, R. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484.

Swingley, D., Pinto, J., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73–108.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin. *Journal of Child Language*, 24, 535–565.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the “noun bias” in context: A comparison of English and Mandarin. *Child Development*, 70, 620–635.

Thal, D., Bates, E., Goodman, J., & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late- and early-talking toddlers. *Developmental Neuropsychology*, 13(3), 239–274.

Thal, D., Bates, E., Zappia, M., & Oroz, M. (1996). Ties between lexical and grammatical development: Evidence from early talkers. *Journal of Child Language*, 23(2), 349–368.

Thal, D., & Katich, J. (1996). Does the early bird always catch the worm? Predicaments in early identification of specific language impairment. In: K. Cole, P. Dale, & D. Thal (Eds), *The measurement of communication and language: Vol 6, Assessment*. pp. 1–28. Baltimore, MD: Brookes Publishers.

Thal, D., & Tobias, S. (1994). Relationships between language and gesture in normally developing and late-talking toddlers. *Journal of Speech and Hearing Research*, 37, 157–170.

- Thal, D., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991). Early lexical development in children with focal brain injury. *Brain and Language*, *40*, 491–527.
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.
- Thiessen, E. D., & Saffran, J. R. (2004). Spectral tilt as a cue to word segmentation in infancy and adulthood. *Perception and Psychophysics*, *66*(5), 779–791.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: Bradford Books/MIT Press.
- Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds “listen through” ambiguous adjectives in fluent speech. *Cognition*, *100*, 389–433.
- Tomasello, M. (1995). Pragmatic contexts for early verb learning. In M. Tomasello & W. E. Merriman (Eds) *Beyond names for things: Young children's acquisition of verbs* (pp. 115–146). Hillsdale, NJ: Erlbaum.
- Tomasello, M. (2001). The item-based nature of children's early syntactic development. In: M. Tomasello, & E. Bates (Eds), *Language development: The essential readings* (pp. 169–186). Malden, MA: Blackwell.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology*, *30*, 639–650.
- Tomasello, M., & Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees, *Monographs of the Society for Research in Child Development*. (Vol. 70, Serial No. 279).
- Trueswell, J. C., & Tanenhaus, M. K. (2005). *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
- Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, *105*, 634–677.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Weizman, Z. O., & Snow, C. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*, 265–279.
- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development. Blackwell handbooks of developmental psychology*. (pp. 167–187). Malden, MA: Blackwell Publishing.
- Werker, J. F., Fennell, C.T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63.

Woodward, A. L., & Markman, E. M. (1997). Early word learning. In: W. Damon, D. Kuhn, & R. Siegler (Eds), *Handbook of child psychology (Vol. 2), Cognition, perception, and language*. New York: Wiley.

Zangl, R., & Fernald, A. (under review). How children process determiners in continuous speech: Developmental changes from 18 to 36 months.

Zangl, R., Klarman, L., Thal, D. J., Fernald, A., & Bates, E., (2005). Dynamics of word comprehension in infancy: Development in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*, 6(2), 179–208.

This page intentionally left blank

## Chapter 28

### Acquisition of Syntax and Semantics

*Stephen Crain and Rosalind Thornton*

#### 1. INTRODUCTION

There are two main views about the nature of language development. These views can be traced back to the ‘nature versus nurture’ debate about how knowledge in any domain is acquired. The ‘nativist’ perspective dates back to Plato’s dialogue ‘The Meno.’ ‘This view emphasizes the contributions of human nature to the acquisition of knowledge. It is supposed on the nativist approach to language that children are biologically fitted, as part of the human genome, with a theory of ‘Universal Grammar’ (e.g., Chomsky, 1965, 1975, 1986). Universal Grammar contains the ‘core’ principles of language, i.e., principles that are manifested in all human languages. In addition, Universal Grammar spells out particular ways in which human languages can vary; these points of variation are called parameters. Taken together, the principles and parameters of Universal Grammar establish the boundary conditions on what counts as a possible human language. Children navigate within these boundaries in the course of language development. Of course, experience determines which particular language children acquire, but nativists argue that much of the process of language acquisition is biologically driven, rather than being ‘data driven.’ The nativist approach views language learning as the by-product of a task-specific computational mechanism, with a structure that enables children to rapidly and effortlessly acquire any human language, without formal instruction and despite considerable differences in linguistic experience. Universal linguistic principles are not learned by the computational mechanism, but are implicit in the structure of the mechanism itself – i.e., these are in the Universal Grammar. This implicit (or built in) knowledge explains how learners come to know more about language than they observe from experience. This is the nativist’s solution to ‘Plato’s Problem.’<sup>1</sup>

---

<sup>1</sup> In Plato’s dialogue, *The Meno*, the protagonist, Socrates, demonstrates to Meno that a young slave knows more about geometry than he could have learned from experience. By extension, ‘Plato’s Problem’ refers to any gap between experience and knowledge.

The alternative ‘nurture’ approach views language development as the product of domain general learning mechanisms. These mechanisms embody general learning processes that are not specially tailored to acquire any particular kinds of facts about the world. Like knowledge in other domains, knowledge of language is accrued in a piecemeal fashion, based on statistical regularities in the input. According to the experience-dependent account, child language matches the input more or less, with more frequently attested constructions being mastered earlier in the course of language development. Gradually, more and more complex structures are composed, until the language of the child approximates that of an adult in the same linguistic community. Tomasello (2000) sums up this approach as follows:

When young children have something to say, they sometimes have a set expression readily available and so they simply retrieve linguistic schemas and items that they have previously mastered (in their own productions or in their comprehension of other speakers) and then “cut and paste” them together as necessary for the communication situation at hand ... (p. 77).

The nature versus nurture debate has intensified in recent years. For a few decades, linguists working within the theory of Universal Grammar pointed out the difficulty learners faced in mastering many facets of language. In the 1980s and 1990s, a great many experimental studies of children’s adherence to linguistic universals were reported in the literature, leading to a picture of language development that was consistent with Universal Grammar.

Recently, there has been a shift in the opposite direction. More and more, it seems, developmental psycholinguists are exploring the possibility that linguistic facts can be learned without the kinds of abstract or implicit principles that have been proposed in the theory of Universal Grammar. Two developments have prompted this change in direction. One is the discovery that children are able to effectively learn certain linguistic properties based on statistical regularities in the input. For example, Saffran, Aslin, and Newport (1996) showed that 8-month-old children could exploit statistical learning to extract information about transitional probabilities from the input. Infants inferred the existence of word boundaries between three-syllable pseudowords (nonsensical combinations of syllables). Those three-syllable sequences that crossed a word boundary were not treated by the child subjects as a ‘word’ during the post-test phase of the study, because there was a lower probability for such sequences to be repeated if they crossed a word boundary than if they were part of a ‘word.’

The second development concerns the nature of the input available to children. The linguistic input had been assumed to be quite impoverished and, therefore, insufficient to support language learning without assistance from Universal Grammar (Chomsky, 1980). It has recently been argued, however, that the input contains relevant features in sufficient abundance to support statistically based acquisition of several seemingly complex facts about language. We will discuss this issue in the next section of the chapter.

Critics of statistical learning have pointed out limitations in statistical learning mechanisms that exploit transitional probabilities. For example, Yang (2004) showed

that statistical learning mechanisms cannot reliably segment sequences of monosyllabic words, though such sequences make up the majority of the input that is directed to children. In a series of papers, Marcus (1998, 1999; Marcus, Vijayan, Rao, & Vishton, 1999) has shown that statistical learning mechanisms are ill suited to learning many properties of languages (also see Smith, 1996). At the same time, the arsenal of arguments and evidence in support of nativism, and against the experience-dependent ('data driven') approach to language development, has also continued to grow. Evidence in favour of the nativist perspective takes several different forms. First, experimental investigations have shown that children do not violate core linguistic principles, even in cases where they might be tempted to violate such principles if they were to adopt general-purpose learning algorithms (Section 3). Second, the nativist approach is reinforced by the observation that children learn 'deep' linguistic principles that tie together apparently disparate facts about language; this is another aspect of children's linguistic competence that is not plausibly a product of experience (Section 4). Third, it has been demonstrated that children know 'hidden' aspects about the meanings of certain sentences; again, it is unlikely that these aspects of meaning are learned from experience (Section 5). Finally, studies have revealed that children follow the natural seams (parameters) of natural language even when child language differs from that of adults. Some features of children's nonadult linguistic behaviour, moreover, are quite unexpected on an experience-dependent account of language development (Section 9).

In this chapter, we discuss these arguments and review some of the results from experimental investigations of child language. The experimental findings should be influential in the current debate about the nature of language development because, as Tomasello (2000) asserts, "Choosing between the alternative is, or should be, an empirical matter ..." (p. 67). We agree. To get started, we will describe how both the experience-dependent approach and the nativist approach attempt to deal with one of the most fundamental features of language – its dependence on structure.

## 2. STRUCTURE DEPENDENCE

Much of the current debate in the literature focuses on the nature of linguistic operations. The example of Yes/No questions is frequently cited. At issue is the relation between declarative sentences (on the left-hand side of the arrow) and their Yes/No question counterparts (on the right-hand side). For every ordinary declarative sentence in English, there is a corresponding Yes/No question, so these structure are obviously related. But how?

1. (a) Bill can play the sax.  $\Rightarrow$  Can Bill play the sax?  
 (b) The sky is blue.  $\Rightarrow$  Is the sky blue?

As Chomsky (1971, 1975) observed, a simple 'structure-independent' hypothesis yields the correct results for much of the input that children receive. For example, the following structure-independent hypothesis will generate the Yes/No questions in (1):



Structure-independent rule. To form a Yes/No question, move the *first* verbal element {is, can, has, ...} of the declarative statement to the front.

The inadequacy of this structure-independent operation is revealed when it is applied to complex examples with a modifying clause (*who is beating a donkey*, as in (3)). The first *is* appears in the modifying clause; if it is moved to the front, the result is an unacceptable Yes/No question in (3).

2. Declarative: the farmer who is beating a donkey is mean.
3. Yes/No question: is the farmer who \_ beating a donkey is mean?

To produce the correct Yes/No question corresponding to (2), the auxiliary verb 'is' following the entire subject phrase *the farmer who is beating a donkey* is moved to the front, yielding (4).

4. Yes/No question: is the farmer who is beating a donkey \_ mean?

A rough formulation of the structure-dependent rule that gives the right results is something like the following:

Structure-dependent rule. Move the auxiliary verb in the main clause to a sentence-initial structural position.

Chomsky (1971) maintained that children would never adopt structure-independent hypotheses, even if the data available to children were consistent with both structure-independent and structure-dependent rules. In other words, children would not be expected to make certain kinds of mistakes in forming Yes/No questions at any stage in language development. So, for example, they are not expected to produce questions like (3): *Is the farmer who beating a donkey is mean?* In an elicited production study, Crain and Nakayama (1987) evoked Yes/No questions from 30, 3- to 5-year-old children, to see if they ever made such mistakes. Although children made certain kinds of errors, they never produced questions that were consistent with structure-independent rules. (On the other hand, the kinds of nonadult responses children made were consistent with the continuity assumption; see Section 9).

It has frequently been claimed by advocates of the experience-dependent approach that nativists assume that "no evidence exists that would enable a three-year-old to unlearn" mistaken structure-independent rules, if children were to initially adopt such rules (Cowie, 1999; also see MacWhinney, 2004; Pullum & Scholz, 2002). But no reasonable nativist would endorse such a strong claim about all possible evidence. The following passage from Chomsky (1975, p. 31) is often quoted as the basis of this conclusion about nativism:

A person may go through a considerable part of his life without ever facing relevant experience, but he will have no hesitation in using the structure-dependent rule, even if all of experience is consistent with [the structure-independent rule].

As this passage makes clear, Chomsky is not claiming that nobody *ever* has relevant experience. The issue concerns the robustness of evidence, not its existence, and children's use of the evidence, regardless of its quantity. As Lasnik and Crain (1985) note, if relevant data are not available in sufficient quantities, then at least some (and perhaps many) children won't come by them, and these children will not converge on an adult grammar. But this is contrary to fact. All (normal) children converge on a system of linguistic principles that is equivalent to that of adults. Therefore, if convergence depends on there being evidence, then it must be available in abundance. Suppose to the contrary, that evidence falsifying the structure-independent hypothesis for forming Yes/No questions is not available to children before they reach their third birthday. Then many children should be observed to make structure-independent errors. But Crain and Nakayama (1987) did not find any evidence that children were adopting structure-independent rules. So either children never form structure-independent hypotheses, or there is abundant evidence *available to and used by* very young children. According to Cowie, "... something like the requisite guarantee can be provided when one reflects on the sheer size of the data sample to which a learner has access." (p. 219). If the requisite evidence includes sentences like (4), however, then the evidence is not readily available to children. A search of the input to English-speaking children turned up only one example of a structure like (4) out of about 3 million utterances (reported by MacWhinney, 2004, using the CHILDES database; see MacWhinney, 2000).

Advocates of the experience-based approach have therefore proposed other sources of evidence for children. One example is (5), which is assumed to be derived from the declarative sentence represented in (6).

5. Where's the other dolly that was in there?  
 6. [ the other dolly that was<sub>AUX</sub> in there ] [ is<sub>AUX</sub> where ]

Notice that the representation that is assigned to the declarative sentence in (6) is partitioned into a main clause [*is<sub>AUX</sub> where*] and a relative clause [*the other dolly that was<sub>AUX</sub> in there*]. To form the corresponding wh-question, (5), the main clause [*is<sub>AUX</sub> where*] is moved to the front (and its internal parts are inverted). Despite the absence of sentences like (7) in the input to children, if questions like (5) are readily available in the input, then these questions would be subject to a similar analysis (compare examples (6) and (8)). In forming both kinds of questions, the 'is<sub>AUX</sub>' in the relative clause appears first in the declarative sentence, but it remains in place, whereas the 'is<sub>AUX</sub>' in the main clause is moved to the front in order to form the corresponding question.

7. Is the farmer who is beating the donkey mean?  
 8. [ the farmer who is<sub>AUX</sub> beating the donkey ] [ is<sub>AUX</sub> mean ]

As MacWhinney (2004) acknowledges, this experience-based account "requires children to pay attention to relational patterns, rather than serial order as calculated from the beginning of the sentence" (p. 891). So we should ask what distinguishes the experience-based account from the nativist account offered by Chomsky. The difference is that the

experience-based account could, in principle, have learned to move the first 'is' to the front, as in the structure-independent rule described earlier. In Chomsky's view, children are incapable of any such structure-independent analyses. Regardless of the input, according to the nativist account, children are compelled to impose a symbolic analysis onto the utterances they experience. According to the experience-based account, on the other hand, children have no such predisposition; the system children acquire depends on the statistical regularities of the input.

It remains to determine whether or not questions like (5) (*Where's the other dolly that was in there?*) are available in sufficient quantity in the input to children, to ensure that every child converges on a grammar that conforms to structure-dependent operations. In the present case, Legate and Yang (2002) conclude that the input does not suffice. They report the results of a search through transcriptions of the input to two young children, Nina and Adam (in the CHILDES database). The input to Nina consists of 14 critical *wh*-questions out of 20,651 questions overall. There were just four critical examples out of 8889 questions in the input to Adam. The paucity of critical input for these children bears out Chomsky's conjecture that a child could "go through a considerable part of his life without ever facing relevant experience." Moreover, such low frequencies of relevant input make it unlikely that every child encounters the requisite evidence by the age at which they are found to adhere to structure dependence. This is a problem for the experience-based account because, without relevant input, some (perhaps many) children would be expected to commit structure-independent errors, such as (3), but this is contrary to the findings of experimental research (Crain & Nakayama, 1987).

### 3. AVOIDING ERRORS: INNATE CONSTRAINTS VERSUS CONSERVATISM

#### 3.1. A Constraint on Reference

Another distinguishing feature of the two approaches to language development is how they explain the kinds of sentences children refrain from producing, and the kinds of meanings that children do not assign to sentences. One case in point is the reference of ordinary pronouns. Note that in the examples in (9) and (10), the pronoun *he* may or may not refer to the individual called *the Ninja Turtle*. To indicate these dual referential possibilities, we will adopt the following notation: two expressions refer to the same individual(s) only if they have the same index. So, (9) and (10) are ambiguous, because the pronoun *he* can have the same index as *the Ninja Turtle* ('1'), but one of these expressions can also be assigned an index '2' which the other expression lacks; in that case, the two expressions are said to be disjoint in reference or noncoreferential.

9. The Ninja Turtle<sub>1</sub> danced while he<sub>1/2</sub> ate pizza.
10. While he<sub>1</sub> ate pizza, the Ninja Turtle<sub>1/2</sub> danced.

Consider another sentence, (11), which also contains the pronoun *he* and the expression *the Ninja Turtle*. Unlike examples (9) and (10), (11) is unambiguous. Intuitively, the pronoun *he* cannot refer to the Ninja Turtle, but must refer to some other male individual.

In other words, coreference (as indicated by the assignment of the same number) is not permissible in the sentence in (11); it has the reading in (11a), but not the reading indicated in (11b).

11. He danced while the Ninja Turtle ate pizza.
- a. He<sub>1</sub> danced while the Ninja Turtle<sub>2</sub> ate pizza.
  - b. \*He<sub>1</sub> danced while the Ninja Turtle<sub>1</sub> ate pizza.

There are two ways of describing the possibilities for referential interpretations of pronouns. Each of these options has been taken by one of the two approaches to language development. One way is to list the various possibilities for coreference. This is the strategy taken by the experience-dependent approach. Adopting this strategy, the list includes some way of representing the positive instances of coreference between pronouns and other expressions, so examples like (9) and (10) would be represented (somehow) in the list. Nothing would be said about the case in (11b), because this is not an instance of coreference.

The alternative strategy is to formulate a negative principle representing those cases in which coreference is prohibited, such as (11b). Nothing is said about any of the other cases, such as (9), (10) and (11a). On grounds of parsimony, Lasnik (1976) argued for the second strategy, because the list of cases where coreference is possible adds up to a huge inventory of linguistic representations, whereas a single generalisation can explain mandatory non-coreference, with cases of coreference left open. Negative linguistic principles are known as constraints. So a constraint prevents coreference between pronouns and referring expressions in sentences like (11b).

Constraints are frequently invoked in arguments for nativism for the following reason. Suppose for the sake of argument that children's grammars embody constraints as negative statements; in the present example, the constraint is a prohibition against certain coreference possibilities, as illustrated in (11b). It seems unlikely that children could 'learn' such negative facts from experience, because parental speech rarely if ever includes explicit negative evidence and learning constraints would seem to require negative evidence (see, e.g., Bowerman, 1998; Brown & Hanlon, 1970; Marcus, 1993). Acquisition in the absence of decisive evidence is one of the main hallmarks of innate specification of knowledge.

Another hallmark of innate specification is early emergence (Crain, 1991). Developmental psycholinguists have investigated the time-course of the acquisition of constraints in pursuit of the early emergence hallmark of innateness. Of course, even innate principles need not emerge early in the course of development. Just as some properties of physical development are biologically timed to appear long after birth (e.g., a second set of teeth), certain aspects of linguistic knowledge might become operative only at a certain maturational stage of development (Borer & Wexler, 1987). But the earlier complex principles emerge in child grammars, the more difficult it would be for the experience-dependent approach to explain the facts, because early emergence of knowledge compresses the evidential basis for learning.

The question of children's knowledge of the constraint on coreference was pursued in a comprehension experiment by Crain and McKee (1985). In this experiment, children encountered sentences like (12) in circumstances appropriate to both interpretations. On one interpretation of (12), the pronoun *he* and the referring expression *the Ninja Turtle* have the same index, 1. This is called the *backwards anaphora* interpretation. It is 'backwards' in the sense that the pronoun comes first. More typically a pronoun follows the expression with which it is anaphorically linked. On the alternative interpretation of (12), *the Ninja Turtle* is not coindexed with the pronoun. The pronoun is said to have *extrasentential* reference on this reading; it refers to an individual who is not mentioned in the sentence.

12. While he<sub>1</sub> ate pizza, the Ninja Turtle<sub>1/2</sub> danced.

The experimental procedure used in the Crain and McKee (1985) study was the Truth-Value Judgment task. As the name suggests, this procedure requires subjects to judge the truth or falsity of a sentence, according to its fit to the context. Two experimenters are needed to conduct the Truth-Value Judgment task. One experimenter uses toys and props to act out a situation corresponding to one interpretation of the target sentence. A second experimenter manipulates a puppet; we often use Kermit the Frog as the puppet. Following each situation, Kermit the Frog says what he thought happened on that trial. When Kermit the Frog accurately describes something that happened in the story, the child is instructed to reward him, say with a strawberry. Sometimes Kermit does not pay close attention, however, and he says the wrong thing. In that case, the child is instructed to give Kermit something to remind him to pay closer attention, say a rag. These procedures make it fun for children to attend to Kermit's statements. Without the rag ploy children are reluctant to say that Kermit has said anything wrong. Note that both (a) the events corresponding to the meaning of the target sentence, and (b) the target sentence itself are provided for the children. This allows unparalleled experimental control and at the same time reduces extraneous processing demands that are present in comprehension tasks in which children are required to act out events themselves.

For ambiguous sentences such as (12), the same sentence was presented on two separate occasions, in two contexts. In one context for (12), the Ninja Turtle was dancing and eating pizza; in the other, someone else ate pizza while the Ninja Turtle was dancing. Kermit uttered the same sentence following both situations. The results were that children accepted the backward anaphora reading about two-thirds of the time, in appropriate contexts. The extrasentential reading of the pronoun was accepted only slightly more often. Only one of the 62, 2- to 5-year-old children (mean age 4,2) interviewed in the Crain and McKee study consistently rejected the backward anaphora interpretation.

To test children's knowledge of the constraint against coreference, there was another condition in the experiment. In this condition, sentences like (13) were presented in situations corresponding to the meaning that is ruled out by the non-coreference constraint. For (13), the situation was one in which the Ninja Turtle danced and ate pizza at the same time. If children adhered to the constraint prohibiting coreference, they were expected to reject (13) as an accurate description of this situation.

13. He<sub>1</sub> danced while the Ninja Turtle<sub>2</sub> ate pizza.

In fact, the child subjects judged sentences like (13) to be false almost 90% of the time. In the context for sentence (13), it was clear that some other salient (male) character did not dance while the Ninja Turtle ate pizza. This made it reasonable for children to give a “No” response, provided that their grammars made a “Yes” response inappropriate. This characteristic of the task is called *plausible dissent*. The need for plausible dissent in experiments is discussed in detail in Crain and Thornton (1998). The findings show that even 2- and 3-year olds prohibit backwards anaphora only when structural conditions (involving c-command) dictate that they should. It is important to appreciate that this experiment provides further evidence that children do not rely on their linguistic experience in making judgements about the appropriate mappings of sentences with their meanings. Since there is nothing in children’s experience to tell them which sentence/meaning pairs are NOT allowed, there is no way to learn the structural constraint prohibiting coreference. This ‘negative statement’ must be part of children’s grammars.

It has been proposed that the same constraint that prohibits coreference in sentences like (14) also governs coreference relations in some discourse contexts. Consider the short discourses shown in (15) and (16).

14. \*He<sub>1</sub> sent the letter to Chuckie’s<sub>1</sub> house.

15. Speaker A: I know where he<sub>1</sub> sent the letter.

Speaker B: Me too. To Chuckie’s<sub>2</sub> house

16. Speaker A: I know where he<sub>1</sub> sent the letter.

Speaker B: Me too. \*To Chuckie’s<sub>1</sub> house

It is intuitively clear that in (14) the pronoun *he* cannot refer to Chuckie, but must refer to some other salient male in the conversational context. The judgments about coreference, then, are similar to the judgements for sentences like (13). This raises the possibility that the same constraint governs both linguistic phenomena. A recent proposal to this effect was made by Merchant (2005) (cf. Hankamer, 1979; Morgan, 1973, 1989). The idea is that part of the structure of the statement by Speaker A is reconstructed by Speaker B, but subsequently deleted. This is illustrated in (17).

17. Speaker A: I know where he<sub>1</sub> sent the letter.

Speaker B: Me too. He<sub>1</sub> sent the letter to Chuckie’s<sub>2</sub> house.

Even though only a fragment answer (*To Chuckie’s house*) is actually produced by Speaker B, it has the same propositional content as a full sentence.

In a recent study, Conroy and Thornton (2005) presented both full sentences and discourse sequences to 20 English-speaking children (mean age 4,6), to see whether children made similar judgements in response to both complete sentences and discourse sequences. On a typical trial of the relevant experimental condition, one of the characters, Tommy, was preparing to send a letter to Chuckie’s house, but then decided against it. In response, Chuckie sent a letter to his own house. Against this backdrop, half of the time,

children heard a complete sentence (*He sent the letter to Chuckie's house*), and half of the time, they heard a discourse like the one in (17). The main finding was that children rejected both the full sentences and the discourse sequences an equal proportion of the time (86% rejections, as compared to 89% rejections, respectively). This finding is consistent with the hypothesis that the same negative constraint against coreference underlies children's responses to both phenomena. On the experience-dependent approach, there is no reason to expect children's responses to both phenomena to coincide, but there is nothing in the approach that rules out this possibility either.

### 3.2. A Constraint on Contraction

Another example of a constraint governs where contraction may and may not occur. In English, this constraint prevents the verbal elements *want* and *to* to be contracted to form *wanna* in certain kinds of sentences, although *wanna*-contraction is permitted most of the time. Examples (18–21) illustrate permissible contractions. Example (22a) illustrates an impermissible contraction.

18. a. Who does Arnold *wanna* make breakfast for?  
b. Who does Arnold *want to* make breakfast for?
19. a. Does Arnold *wanna* make breakfast for Maria?  
b. Does Arnold *want to* make breakfast for Maria?
20. a. Why does Arnold *wanna* make breakfast?  
b. Why does Arnold *want to* make breakfast?
21. a. I don't *wanna* make breakfast for Arnold or Maria.  
b. I don't *want to* make breakfast for Arnold or Maria.
22. a. \*Who does Arnold *wanna* make breakfast?  
b. Who does Arnold *want to* make breakfast?

All of the questions in these example begin with *wh*-words (*who*, *what*, *why*, *where*, even *how*) and will be called *wh*-questions. According to a standard account of *wanna*-contraction, *wh*-questions are formed by movement of a *wh*-phrase from one position at an underlying level of representation to another position, on the surface, where it is pronounced. A further assumption of the account is that a record, which we abbreviate as *t* (for 'trace'), is left behind at the site of the origin of the *wh*-movement. In (23) the *wh*-phrase originates in the subject position of the embedded infinitival clause *want t to kiss Bill*. When the *wh*-phrase starts out between *want* and *to*, as in (16), the trace left behind by *wh*-movement blocks the contraction of *want* and *to*. This explains why (23b) is ruled out. The same account explains the unacceptability of (22a).

23. a. Who do you want *t* to kiss Bill?      Subject extraction  
b. \*Who do you wanna kiss Bill?

By contrast, in (24), the formation of the *wh*-question requires the movement of the *wh*-phrase from the object position of the embedded infinitival clause. In that case, the trace does not intervene between *want* and *to*, so *wanna*-contraction is permitted.

24. a. Who do you want to kiss *t*?      Object extraction  
 b. Who do you wanna kiss *t*?

These facts invite the following generalisation: Contraction of the verbal elements *want* and *to* is blocked if the trace of wh-movement intervenes between them. In declaratives, the constraint on contraction is irrelevant, so contraction is tolerated.

As examples (18) – (21) indicate, much of the evidence available to children learning English runs counter to the constraint. That is, contraction of *want* and *to* is licensed in general – (22) is an exception to the rule. If the grammars of English-speaking children lacked the constraint on contraction of *want* and *to* (across the trace of a moved wh-word, then child English would include more sentences than adult English does. In other words, without the constraint, children would *overgenerate*, and would produce sentences like (22a) and (24b) with illicit contraction of *want* and *to*.

Children who lack the constraint on contraction across a trace should permit contraction to a similar extent in both subject- and object-extraction questions. To test children's adherence to the constraint, an experiment was designed to elicit relevant questions from children (Thornton, 1990, 1996). This permitted a comparison of the proportion of contraction by children in questions like (24) with contraction in questions like (23). The finding was that the 21 children interviewed (mean age 4,3) produced contracted forms more than half the time (57%) in questions like (23), but the same children produced contracted forms less than 10% of the time in questions like (24), where contraction is outlawed by the constraint.

The linguistic constraint that prohibits *wanna*-contraction also applies to a variety of other constructions, but not in ways that can easily be determined on the basis of the primary linguistic data. For example, the constraint prohibiting contraction across a trace governs a linguistic phenomenon known as *is*-contraction. A good case can be made that *is* contracts to its right, despite the orthographic convention that links an 's with the word to its left. The paradigm in (25) shows that *is* can contract when there is no trace to its immediate right, as in (25b), but contraction is blocked when there is a trace to its immediate right, as in (25d).<sup>2</sup>

25. a. Do you know what that is doing *t* up there?  
 b. Do you know what that's doing *t* up there?  
 c. Do you know what that is *t* up there?  
 d. \*Do you know what that's *t* up there?

Having witnessed two applications of the constraint on contraction, it is important to ask how a learning-theoretic account could explain the generalisation that relates *wanna*-contraction and *is*-contraction. The constraint applies to linguistic phenomena that bear little superficial resemblance. In the *wanna*-contraction paradigm, the

<sup>2</sup> *Is*-contraction is also blocked if there is no linguistic material to the right of *is*, e.g., when it is in sentence final position.



constraint prevents contraction across the subject position of an embedded infinitival clause, whereas in the *is*-contraction paradigm, the same constraint prevents contraction across the object position in a tensed clause. Until a wide range of linguistic phenomena was considered, including both positive and negative data, linguists failed to see that the two phenomena were related. Assuming that language-learners do not have access to such complex arrays of positive and negative data, nativists conclude that language-learners must have an advantage over linguists, in knowing the linguistic constraint in advance of encountering the limited primary linguistic data to which they have access.

Returning to child language, the nativist is compelled to predict that children will adhere to the constraint on contraction across a trace in both constructions. Another 12 2-4-year-old children (mean age 3,8) participated in an elicited production experiment designed to assess their knowledge of the constraint that prohibits *is*-contraction. The finding was the complete absence of illicit productions. Illicit contraction is apparently prevented by the constraint. These mutually supporting findings suggest that the same constraint rules out (24b) and (25d).

### 3.3. Maintaining Records of Attested Structures

Experience-based accounts of language acquisition take a different stance on the acquisition of 'constraints.' Their approach is to list the positive cases. By supposing that children are 'conservative' learners, in the sense that their grammars are directly tied to experience, such accounts avoid the problem of learning negative constraints in the absence of negative evidence. Being conservative, learners never produce forms they do not encounter. For example, Pullum and Scholz (2002, p.16) use the conservative learning strategy to explain how the linguistic expressions that cooccur with subordinate words like *than* and *that* are learned, as summarized in (26).

26. They wanted more than was available.  
 a. ... 'more than + finite VP'  
 b. \*... 'more that + finite VP'

If the types of constituents that can occur as complements to subordinating words like *that* and *than* are learned piecemeal from positive examples, then the pattern '*more than + finite VP*' will be learned (after encountering examples like *They wanted more than was available*), but '*V + that + finite VP*' will not be learned, because no examples of that sort will ever be encountered.

As this quote makes clear, experience-based accounts are committed to piecemeal acquisition from positive example, so that the absence of generalisations beyond the input are explained. As Cowie (1999, p. 223) remarks:

the non appearance of a string in the primary linguistic data can legitimately be taken as constituting negative evidence.

But, as Crain and Pietroski (2001) point out, the conservative learner will have to keep detailed records of all kinds of grammatical distinctions in order to avoid potential pitfalls in sentence production. For example, the declarative in (27) has the wh-question counterpart in (28) (where *someone* has been turned into *who*, and moved to the front of the sentence). But the subordinate word *that* must not appear in the wh-question corresponding to (28), as the unacceptability of (29b) illustrates.

27. He is hoping that someone is coming to visit.  
 28. [ who is he hoping ] [ that \_ is coming to visit ]  
 29. a. Who is he hoping is coming to visit?  
 b. \*Who is he hoping that is coming to visit?

Consider next how children would represent the absence of *wanna*-contraction in the kinds of wh-questions discussed in the last section. Children would need to distinguish between word strings that differ in the nature of the wh-phrase because, as we saw, *wanna*-contraction is permitted in ‘why’ questions, but not in all ‘who’ or ‘what’ questions. In ‘who’ and ‘what’ questions, *wanna*-contraction is permitted if the verb in the clause following *wanna* is transitive, but only if the trace of the moved wh-phrase follows the verb, rather than precedes it (cf. examples (23) and (24)). In short, children must encode the distinction between subject and object position as well as between transitive and intransitive verbs. In the simplest case, children would require a statistical learning mechanism that operates on labelled strings that are six words long (wh-phrase, auxiliary verb, subject NP, *want*, *to*, verb).

To further reinforce the need for detailed record keeping on the experience-dependent account, let us look at the distribution of the expression *at all*. Other words with similar distributional patterns are *any*, *much* and *ever* – the class of such expressions is referred to as *negative polarity items*. Example (30) illustrates that the universal quantifier *every* licenses the negative polarity item *at all* in the subject phrase (e.g., *every politician who favors the rich* or *every politician in this room*), but negative polarity items are not permitted in the predicate phrase of such sentences. Suppose a learner who encountered (30a) formed the broader generalisation that *every* licenses negative polarity items in either position. Such a learner would overgenerate, i.e., she would produce the unacceptable (30b) as well as the acceptable (30a). Conservative learning is necessary to hold overgeneration in check.

30. a. Every politician who favours the rich *at all* is in this room.  
 b. \* Every politician in this room favours the rich *at all*.

Example (31) reveals that both ‘local’ and ‘distant’ negation (*not* or *n’t*) license negative polarity items. And (32) reveals that some linguistic contexts with negation license negative polarity items, but others do not.

31. a. Bush doesn’t believe that liberals favour the poor *at all*.  
 b. Bush believes that liberals don’t favour the poor *at all*.  
 32. a. The news that Bush won didn’t surprise the Supreme Court *at all*.  
 b. \* The news that Bush didn’t win surprised the Supreme Court *at all*.

If children are to avoid the kinds of overgeneration illustrated in the (b) examples of (29), (30) and (32), then they must keep track of all of the relevant distinctions in the linguistic contexts that license negative polarity items, and ones that do not. This prediction of the experience-based account seems highly implausible to the nativist, because the relevant distinctions that children would need to keep track of are so subtle and so numerous (see Crain & Pietroski, 2001, pp. 172–173). The kinds of record keeping that is needed to mimic linguistic constraints would seem to be beyond the capacity of certain statistical learning mechanisms, such as connectionist or parallel distributed processing networks. These networks rely on local regularities – i.e., changes in the “connection between one unit and another on the basis of information that is locally available to the connection” (Rumelhart & McClelland 1986, p. 214). According to Rumelhart and McClelland, such models “provide very simple mechanisms for extracting information from an ensemble of inputs without the aid of sophisticated generalizations or rule-formulating mechanisms.” Such models are evidently incapable of learning the kinds of linguistic facts that children learn, such as facts about the ‘displacement’ of *wh*-phrases, the consequences of *wh*-movement for contraction, and the permissible locations of negative polarity items.

#### 4. AN UNEXPECTED GENERALISATION

There are more arrows in the nativist’s quiver. Another reason for questioning the experience-based account of language acquisition is the lack of explanations for (a) the generalisations formed by children in the course of language development, and (b) cross-linguistic generalisations. As in any other science, progress is made in linguistics when apparently unrelated facts can be amalgamated. As the physicist Richard Feynman (2005, pp. 23–24) remarks:

The things with which we concern ourselves in science appear in myriad forms, and with a multitude of attributes. ... Curiosity demands that we ask questions, that we try to put things together and try to understand this multitude of aspects as perhaps resulting from the action of a relatively small number of elemental things and forces acting in an infinite variety of combinations. ... In this way we try gradually to analyze all things, to put together things which at first sight look different, with the hope that we may be able to *reduce* the number of *different* things and thereby understand them better.

In the previous section we observed that negative polarity items like *at all*, *much*, *ever* and *any*, are licensed in certain linguistic contexts, but not in others. For example, we saw that *at all* could appear in the subject phrase of the universal quantifier, but not in the predicate phrase. This and other asymmetries are illustrated in (33)–(35).

33. a. Every linguist who agreed with *any* philosopher is in this room.  
 b. \* Every linguist in this room agreed with *any* philosopher.

34. a. If *any* linguist enters the gym, then Geoff leaves.  
 b. \* If Geoff leaves, then *any* linguist enters the gym.
35. a. Geoff went to the gym before *any* linguist.  
 b. \* Geoff went to the gym after *any* linguist.

The point we made earlier about such asymmetries was that, on a data-driven approach, children could avoid producing the unacceptable (b) examples only by keeping careful records of the actually occurring expressions and (re)producing only the examples that were attested in the input. Otherwise, illicit examples could be generated.

There is another aspect of these asymmetries that we would draw to your attention. Another seemingly quite distinct phenomenon is manifested in the same linguistic contexts that permit *any*. The second phenomenon is the interpretation of disjunction (*or* in English). Although these linguistic phenomena are radically different in character, the fact that they are manifested in the same linguistic contexts argues that they should be amalgamated. It is worth asking how the alternative approaches to language development can achieve the amalgamation.

First, it will be helpful to describe the interpretation of disjunction in logic and in natural language. In classical logic, the logical expression for disjunction ‘ $\vee$ ’ is assigned truth conditions associated with inclusive-*or*. This means that formulas of the form  $A \vee B$  are true when  $A$  is true (but not  $B$ ), when  $B$  is true (but not  $A$ ), and when both  $A$  and  $B$  are true. Such a formula will be FALSE only when both of its disjuncts are false. If the original disjunction is negated, it will have the opposite truth conditions. So  $\neg(A \vee B)$  will be TRUE only when both of its disjuncts are false. Because disjunction is assigned the truth conditions of inclusive-*or*, a formula in which disjunction appears in the scope of negation  $\neg(A \vee B)$  will be true if and only if both  $A$  and  $B$  are false. Let us refer to these truth-conditions of disjunction under negation as the ‘neither ... nor ...’ interpretation. The ‘neither ... nor ...’ interpretation can also be rendered as a conjunction, where both of the disjuncts from  $\neg(A \vee B)$  and negated. This is stated in one of De Morgan’s laws, where the symbol ‘ $\Rightarrow$ ’ indicates logical entailment, and the symbol ‘ $\wedge$ ’ represents conjunction (English *and*).

$$\neg(A \vee B) \Rightarrow \neg A \wedge \neg B$$

This logical entailment will be referred to the ‘conjunctive’ entailment of disjunction in the scope of negation.

To some degree, natural language mirrors classical logic. For example, when disjunction appears in simple negative sentences in English, the interpretation is consistent with De Morgan’s law. Consider the sentence in (36). Adult speakers of English interpret (36) to entail (37).

36. John doesn’t speak French *or* Spanish.  
 37. John doesn’t speak French *and* John doesn’t speak Spanish.

In short, in English disjunction under negation yields the kind of conjunctive entailment described in De Morgan's law. It follows that the word for disjunction in English (*or*) has the truth-conditions associated with inclusive-*or* in classical logic.

De Morgan's law is just the tip of the iceberg, however. Disjunctive statements yield conjunctive entailments in a great many other linguistic contexts as well as in simple negative sentences (Chierchia, 2004). Example (38) shows that sentences with the universal quantifier *every* generate conjunctive entailments. Similarly, if disjunction is in the antecedent of a conditional, the result is logically equivalent to a conjunctive statement, as (39) illustrates. And (40) reveals that the preposition *before* yields conjunctive entailments.

38. a. Every student who speaks French or Spanish is in this room.  
 b.  $\Rightarrow$  every student who speaks French is in this room *and*  
 every student who speaks Spanish is in this room
39. a. If Ted *or* Kyle enters the gym, then Geoff leaves.  
 b.  $\Rightarrow$  if Ted enters the gym, then Geoff leaves *and*  
 if Kyle enters the gym, then Geoff leaves
40. a. Geoff went to the gym before Ted *or* Kyle.  
 b.  $\Rightarrow$  Geoff went to the gym after Ted *and*  
 Geoff went to the gym before Kyle

We have already witnessed these linguistic contexts, in the discussion of the contexts that license negative polarity items (repeated here).

41. Every linguist who agreed with *any* philosopher is in this room.  
 42. If *any* linguist enters the gym, then Geoff leaves.  
 43. Geoff went to the gym before *any* linguist.

As these examples indicate, conjunctive entailments of disjunction are generated in precisely the same linguistic contexts in which *any* is permitted. Moreover, conjunctive entailments are *not* enforced when disjunction is in the predicate phrase of the universal quantifier *every*, or when it is in the consequent clause of conditionals, or when it follows the preposition *after*. In short, wherever negative polarity items are not licensed, conjunctive entailments of disjunction are not generated. To illustrate, there is no conjunctive entailment in (44). To see this, note that (a) and (b) are not contradictory, which would be the case if (44a) made a conjunctive entailment.

44. a. Every student in this room speaks French *or* Spanish.  
 b. every student in this room speaks French or Spanish, but no one speaks both  
 c.  $* \Rightarrow$  every student in this room speaks French *and*  
 every student in this room speaks Spanish

With these parallels in mind, we can assert the following descriptive generalisation.

45. All (and only) those linguistics contexts that license *any* yield conjunctive entailments of disjunction.

The linguistic generalisation in (45) represents a challenge to the experience-based approach, because there is no apparent mechanism that would enable children to learn the phenomena under consideration. First, consider the asymmetry in the acceptability of (46a) as compared to (46b).

46. a. \* Every student read anything.  
 b. Every student who read anything passed the exam.

We have seen that a ‘conservative’ learner could avoid producing (46a) by keeping track of when negative polarity items like *any* occur, and using these items only in constructions where they have been attested in the input. In addition to the need for detailed records of attested constructions, if language learners are conservative in this way, admitting into their grammars only principles that generate expressions encountered in the linguistic environment, then there is a danger that they will *undergenerate*, such that their grammars will be weaker than adult grammars. Such learners would not achieve a state of linguistic competence that allows for production and comprehension of sentences never encountered. As Pinker (1990, p. 6) remarks, “... children cannot simply stick with the exact sentences they hear, because they must generalize to the infinite language of their community.”

Look next at the asymmetry in (47). Here, no particular word is at issue. The word *or* is permitted in both sentences. The relevant distinction is in the interpretation of *or*: in (47a) *or* is assigned a ‘not both’ interpretation; in (47b) it makes a conjunctive entailment. The relevant distinction in this case cannot be based on any distributional analysis of the occurrence or nonoccurrence of a particular (kind of) word; the distinction that must be drawn by the learner concerns the different interpretations that the same word (*or*) receives when it appears in different linguistic environments.

47. a. Every student wrote a paper or made a classroom presentation.  
 b. Every student who wrote a paper or made a classroom presentation passed the exam.

It is unclear how children figure out this interpretive distinction on the experience-dependent account. Worse yet is the fact that the linguistic contexts that permit *any* and license-conjunctive entailments are highly correlated. It is implausible, to say the least, that English-speaking children are somehow informed by their caretakers that the same linguistic contexts that license *any* also give rise to conjunctive entailments for *or*. Without triggering evidence, children may require assistance from principles that operate at the ‘core’ of natural language, rather than on the surface.

The alternative to learning is innate specification. According to Universal Grammar, there is a common property that governs the insertion of negative polarity items, and the

licensing of conjunctive entailments of disjunction. The common property is called *downward entailment*. A linguistic expression is downward entailing if it generates valid inferences from general claims about things, to specific claims about those things.<sup>3</sup> The examples in (48) demonstrate that all of the linguistic expressions under consideration have the defining property of downward entailment, since it is valid to substitute claims about sets of things (being a Romance language) with claims about subsets of those things (French, Spanish, Italian, etc.).

48. a. Every student who speaks a Romance language likes to travel.  
 ⇒ *every student who speaks French likes to travel*  
 b. If a student speaks a Romance language, she likes to travel  
 ⇒ *if a student speaks French, she likes to travel*  
 c. John went to Europe before learning a Romance language.  
 ⇒ *John went to Europe before learning French.*

So, downward-entailing expressions have three properties. They license downward entailments (i.e., inferences from general statements to specific statements), they create conjunctive entailments when they are combined with disjunction and they license negative polarity items like *any* and *at all*. This provides us with empirical tests to assess children's knowledge of downward-entailing operators.

## 5. DOWNWARD ENTAILMENT IN CHILD LANGUAGE

There have been a number of experimental studies on English-speaking children's interpretation of disjunction in the scope of negation (Chierchia et al., 2001; Gualmini et al., 2001; Gualmini & Crain, 2002, 2004; Crain et al., 2002). These studies have revealed that 4- to 5-year-old English-speaking children are aware of the conjunctive entailment of disjunction under negation. A representative example is an experiment by Crain et al. (2002) using the Truth-Value Judgment task (see Crain & Thornton, 1998, for extensive discussion of this task). On a typical trial of the experiment, sentence (49) was produced by a (wizard) puppet as a prediction about how events would unfold in a story.

49. The girl who stayed up late will *not* get a dime *or* a jewel.

It subsequently turned out that the girl who stayed up late received a jewel, but not a dime. English-speaking children (mean age 5,0) correctly rejected sentences like (49) 92% of the time in experimental contexts such as this. Children's stated reason for rejecting (49) was that the girl who stayed up late had received a jewel. It is evident that, in children's grammars, (49) entails that the girl would receive neither a dime nor a jewel. This is logically equivalent to the conjunction: the girl would not receive a dime *and* she would

<sup>3</sup> The pattern of inference is in a 'downward' direction due to the presence of these linguistic expressions. In the absence of a downward-entailing expression, the pattern of inference is typically 'upward' such that inferences are sanctioned from claims about sets of things to claims about supersets of those things. So, for example, *Mary bought a Ferrari* entails *Mary bought a car*, but not vice versa.

not receive a jewel. This conjunctive entailment follows from (49) only if the disjunction operator *or* is assigned the inclusive-*or* interpretation, as in classical logic. The fact that children interpreted *or* as inclusive-*or* is difficult to explain on the experience-dependent approach, because children have little direct evidence for the inclusive-*or* interpretation of disjunction. The majority of the input to children consists of *positive* statements. In most positive statements, the use of *or* does not conform to classical logic. Instead, the use of *or* implies exclusivity (the ‘not both’ reading) although it does not entail it.

The implicature of ‘exclusivity’ for *or* stems from the availability of another statement, with *and*, which is more informative (if both statements are true). The statement with *and* is more informative because a statement of the form *A and B* is true in a subset of circumstances that verify a statement of the form *A or B*. For this reason, the expressions *or* and *and* form a scale based on information strength, with *and* being more informative (stronger) than *or* in contexts that verify sentences with either expression. A pragmatic principle *Be Cooperative* (cf. Grice, 1975) entreats speakers to be as informative as possible. Upon hearing someone use the weaker term on the scale *or*, listeners infer that the speaker, who was being cooperative, was not in position to use the stronger term *and*. So the speaker is taken to imply the negation of the sentence with the stronger term: this yields the derived meaning: *A or B, but not both A and B*. Adult use of *or* is clearly governed by this *scalar implicature*; adults avoid using *A or B* in situations in which both A and B are true. Consequently, the vast majority of children’s experience is consistent with the conclusion that natural language disjunction is exclusive-*or*, and not inclusive-*or* (see Crain, Goro, & Thornton, 2006). It should come as no surprise, then, that the same children accepted sentences like (50) 87% of the time in a context in which the girl who stayed up late had received a jewel, but not a dime. This is the same context that resulted in children’s rejection of (49).

50. The girl who didn’t go to bed will get a dime *or* a jewel.

It ought to come as a surprise for the experience-based approach that children interpret natural language disjunction in accordance with classical logic despite the paucity of evidence for this interpretation in the input. The evidential ‘gap’ is even more extreme in languages like Japanese, Hungarian, Chinese and so on, where the adult use of disjunction violates De Morgan’s law even in simple negative sentences (See Section 9.2). Nevertheless, cross-linguistic studies of Japanese-speaking children and Chinese-speaking children demonstrate their steadfast adherence to De Morgan’s law for the conjunctive entailment of disjunction in simple negative sentences, which is possible only if these children interpret disjunction as inclusive-*or*.

Children’s knowledge of the asymmetry involving the universal quantifier *every* has also been demonstrated in the literature. Before we discuss the findings, it will be useful to clarify a few more semantic properties of the universal quantifier. The universal quantifier is a Determiner, like *no*, *some*, *both*, *the*, *three* and so forth. Structurally, Determiners combine with a noun (*student*) or a noun phrase (*student in this room*) to form a grammatical unit – like *every student* or *every student in this room*. The noun (phrase) that *every* combines with what is called its Restrictor (abbreviated by the subscript ‘R’ in



the schema in (51)). Once *every* combines with its Restrictor, the entire unit can then be combined with a predicate phrase (e.g., *swims, speaks French or Spanish*, etc.). The predicate phrase is called the Scope of the universal quantifier (abbreviated ‘S’ in (51)). If the disjunction operator *or* appears in the Scope of *every*, it has an ‘exclusive-or’ (‘not both’) interpretation. For example, the sentence *John speaks French or Spanish* implies that John speaks French OR John Speaks Spanish, but not both. As before, we attribute the ‘not both’ reading to a semantic/pragmatic implicature; this interpretation is not taken as evidence of an ambiguity in the meaning of *or* in English, or in any natural language.

51. Every<sub>R</sub>[ ..... ]<sub>S</sub>[.... or .....] = exclusive-or  
 Every<sub>R</sub>[ ... or ... ]<sub>S</sub>[.....] = conjunctive entailment

Example (52) shows that the negative polarity item *any* is permitted in the Restrictor of the universal quantifier *every*, but not in its Scope. This illustrates the descriptive generalisation that *any* may only appear in linguistic contexts that license the conjunctive entailment of disjunction.

52. Every<sub>R</sub>[ ... any ... ]<sub>S</sub>[... \*any ...]

Several studies have investigated the truth conditions children associate with disjunction in the Restrictor and in the Scope of the universal quantifier (e.g., Boster & Crain, 1993; Gualmini, Meroni, & Crain, 2003). Using the Truth-Value Judgment methodology, children were asked in these studies to evaluate sentences like those in (53) and (54), produced by a puppet, Kermit the Frog.

53. Every woman bought eggs or bananas.  
 54. Every woman who bought eggs or bananas got a basket.

Sentences like (53) were presented to children in a context in which some of the women bought eggs, but none of them bought bananas. The child subjects consistently accepted test sentences like (53) in this condition, showing that they assigned an exclusive-or interpretation to disjunction in the Scope of the universal quantifier, *every*. Children were presented with sentences like (54) in a context in which women who bought eggs received a basket, but not women who bought bananas. The child subjects consistently rejected the test sentences in this context. This finding is taken as evidence that children generated a conjunctive entailment of disjunction in the Restrictor of *every*. This asymmetry in children’s responses demonstrates their knowledge of the asymmetry in the two grammatical structures associated with the universal quantifier – the Restrictor and the Scope. Taken together, the findings are compelling evidence that children know that the Restrictor of *every* is downward entailing, but not its Scope.

There are only a handful of studies bearing on the development of polarity sensitivity in children, but what little is known is consistent with the conclusion that young children produce and avoid negative polarity items in the same linguistic contexts as adults do (O’Leary & Crain, 1994; Thornton, 1995; van der Wal, 1996). An experiment by O’Leary

and Crain is representative. These researchers used a Truth-Value Judgment task with an elicitation component. In the task, the puppet, Kermit the Frog, often produced false descriptions of the events that had taken place in the story. Whenever Kermit the Frog failed to accurately state what had happened in a story, children were asked to say ‘what really happened.’ The experimenter who was manipulating Kermit produced sentences like those in (55) and (56).

55. Kermit: Every dinosaur found *something* to write with.  
 Child: *No, this one didn't find anything to write with.*
56. Kermit: Only one of the reindeer found *anything* to eat.  
 Child: *No, every reindeer found something to eat.*

In the condition illustrated by (55), Kermit’s statement had a universal quantifier *every*, which does not tolerate negative polarity items, such as *anything*, in its Scope; instead, the (positive polarity) expression *something* was used. Eleven children (mean age 4;10) participated in the study. These children’s responses frequently contained the negative polarity item *anything* in linguistic contexts that license it. In another condition, illustrated in (56), Kermit’s statement contained the negative polarity item *anything*. However, in correcting Kermit, children consistently used the universal quantifier *every*, so the linguistic context forced children to avoid repeating the same item. These findings make it clear that children have mastered some, if not all, of the requisite knowledge of downward entailment, which underlies the appropriate use and avoidance of negative polarity items.

These findings are a challenge to the experience-dependent approach. On that approach, as we noted earlier, it is conceivable that children could master the facts about the distribution of negative polarity items, such as *any*, based on statistical properties of the input. Children would have to be exceedingly accomplished at keeping track of the linguistic environments that license such items, however, to avoid producing them in illicit environments. The challenge posed by the asymmetry in the interpretation of disjunction *or*, in the Restrictor versus the Scope of the universal quantifier *every* is more formidable, since the distinction is one of interpretation and does not involve the distribution of lexical material.

## 6. AN ABSTRACT STRUCTURAL PROPERTY: C-COMMAND

In addition to the ‘core’ semantic notion, downward entailment, a structural property cuts across all of the phenomena we have been discussing. This structural property is known as *c-command*.<sup>4</sup> For example, in order for disjunctive statements to license conjunctive entailments, the downward-entailing expression must *c-command* the disjunction operator. Moreover, in order to license negative polarity items, the down-

---

<sup>4</sup> An expression A *c-commands* another expression B in a phrase structure diagram if there is a path that extends above A to the first branching node, and then proceeds down to B.

ward-entailing expression must c-command the position where the item is introduced. In the (a) examples in (57) and (58), negation (*n't*) c-commands *at all*, and *or*, respectively; c-command is absent in the (b) examples. This explains why (57b) receives a 'not both' interpretation, rather than licensing a conjunctive entailment, as in (58a).

57. a. The news that Bush won *didn't* surprise the Supreme Court *at all*.  
 b. \* The news that Bush *didn't* win surprised the Supreme Court *at all*.  
 58. a. The news that Bush won *didn't* surprise Karl *or* Jeb.  
 b. The news that Bush *didn't* win surprised Karl *or* Jeb

The structural relation c-command extends to many other linguistic phenomena. For example, c-command governs the constraint on coreference that prevents the ordinary pronoun *he* from picking out the same individual as the expression *the Ninja Turtle* in (59) (previous example (11)). And another constraint invoking c-command explains why the reflexive pronoun *himself* must be coreferential with *the father*, and not with *the Ninja Turtle* or *Grover*, in (60).

59. He danced while the Ninja Turtle ate pizza.  
 60. Grover said that the father of the Ninja Turtle fed himself.

As these examples illustrate, the structural notion of c-command runs through the principles of Universal Grammar. This structural notion would have to emerge from the statistical regularities of the input on the experience-dependent account of language development, but this too seems highly implausible, since c-command governs a wide array of phenomena that concern meaning, and not just the form of sequences of natural language expressions.

## 7. UNIVERSAL LINGUISTIC PRINCIPLES

Another challenge for the experience-based account of language development is to explain why many of the phenomena we have been discussing appear in languages other than English (as well as in English). This is confirming evidence that the principles underlying these phenomena run deep beneath the surface. For example, the conjunctive entailments of disjunction under negation, and in sentences with the universal quantifier, are manifested in Chinese (61) and in Japanese (62), just as in English. It is likely that all languages exhibit the same linguistic behaviour. It may turn out that natural language disjunction is always inclusive-*or*, as in classical logic, and so the following generalisation may be advanced: universally, disjunction yields a conjunctive entailment when it is c-commanded by a downward-entailing operator.

### Conjunctive entailments for disjunction (*huozhe*) in Chinese

61. a. Mali *meiyou* shuo-guo Yuehan hui shuo fayu *huozhe* xibanyayu.  
 Mary not say-Perf John can speak French or Spanish  
 "Mary didn't say that John spoke French or Spanish"

- ⇒ Mary didn't say that John spoke French *and*  
 Mary didn't say that John spoke Spanish.
- b. Meige [hui shuo fayu huoze xibanyayu de] xuesheng dou tongguo-le kaoshi  
 every can speak French or Spanish DE student DOU pass-Perf exam  
 "Every student who speaks French or Spanish passed the exam"  
 ⇒ every student who speaks French passed the exam *and*  
 every student who speaks Spanish passed the exam.

Conjunctive entailments for disjunction (*ka*) in Japanese

62. a. Mary-wa [furansugo *ka* supeingo-wo hanas-u] gakusei-wo mi-nakat-ta  
 Mary-TOP French or Spanish-ACC speak-pres student-ACC see-neg-past  
 "Mary didn't see a student who speaks French or Spanish"  
 ⇒ Mary didn't see a student who speaks French *and*  
 Mary didn't see a student who speaks Spanish.
- b. [furansugo *ka* supeingo-wo hanas-u] dono gakusei-mo goukakushi-ta  
 French or Spanish-ACC speak-pres every student-MO pass-exam-past  
 "Every student who speaks French or Spanish passed the exam"  
 ⇒ every student who speaks French passed the exam *and*  
 every student who speaks Spanish passed the exam.

A common rejoinder by advocates of the experience-dependent approach is that linguistic universals are part of (an innately specified) logic, and are not specific contingent properties of natural language. For example, Goldberg (2003) contends that "cross-linguistic generalisations are explained by appeal to general cognitive constraints together with the functions of the constructions involved." Attributing cross-linguistic generalisations to general cognitive constraints, rather than to specific linguistic constraints, may be plausible in certain cases, such as the basic interpretation of disjunction as inclusive-*or*. However, other features of natural language resist such an explanation. We will mention just two counterexamples here.

First, we have already seen that the interpretation of disjunction across languages is complicated by properties specific to natural language and not manifested in systems of logic. Recall that most positive declarative sentences with disjunction carry an implicature of exclusivity (i.e., the 'not both' reading of disjunction). This scalar implicature appears to be a universal property of natural languages, though the computation of such implicatures emerges late in the course of language development (see Chierchia et al., 2001; Guasti et al., 2005). Setting aside such complications, the basic interpretation of disjunction can perhaps be chalked up to 'general cognitive constraints.' In any event, this implicature is cancelled in the examples in (61) and (62); hence disjunction licenses the conjunctive entailment associated with inclusive-*or*.

There is a second mismatch between logic and language in (61) and (62). The downward-entailing expressions in (61) and (62) (negation and the universal quantifier) appear outside the clause that contains disjunction. This is probably another linguistic universal. If negation and disjunction were clausemates, however, then the conjunctive

entailment is not enforced in languages like Japanese, Chinese, Hungarian and many others. In another class of languages, which includes English and German, among others, disjunction yields conjunctive entailments regardless of the placement of the downward-entailing operator with respect to the disjunction operator. In these languages, conjunctive entailments of disjunction arise so long as the downward-entailing operator c-commands disjunction. This cross-linguistic variation in where downward-entailing expressions are able to license conjunctive entailments is a parametric option for natural languages with no counterpart in logic. The existence of such parameters, which partition languages in ways that logic does not even suggest, constitutes one of the strongest arguments for Universal Grammar and against the experience-dependent approach to language development (see Crain, Gualmini, & Pietroski, 2005, for further discussion).

The next linguistic phenomenon we describe is the interpretation of focus operators, such as *only* in English. This construction is of particular interest, for several reasons. First, the linguistic phenomenon is (more-than-likely) universal. Second, it has no counterpart in logic. Finally, the interpretation of focus operators involves the generation of a hidden meaning component that has no apparent evidential basis in children's experience, so this phenomenon represents yet another challenge to the experience-based approach to language development.

## 8. HIDDEN ENTAILMENTS

As proposed by Horn (1969), the meaning of a sentence with the focus operator *only*, such as (63), can be decomposed into two conjoined propositions. The first proposition pertains to the focus element, *Bruce*. The content of this proposition is the truth-conditional meaning of the original sentence, absent the focus operator, *only*. So, (64) represents the first meaning component of (63). We will call this its *presupposition*.

63. Only Bruce speaks a Romance language.  
 64. Bruce speaks a Romance language.

The second meaning component is a proposition that is entailed by (63). The content of this proposition comes from the focus operator *only*. The entailment is that the property being attributed to the individual in focus (speaking a Romance language) is not a property of anyone else in the conversational context. So, the second meaning component of (63) can be represented as (65). We will call this the *assertion*.<sup>5</sup>

65. For all  $x$  [ $x \neq$  Bruce], it is not the case that  $x$  speaks a Romance language.

Now let us ask if *only* is downward entailing. As we saw, downward-entailing expressions endorse inferences from claims about sets of things to claims about subsets of those things. The entailment from a set to its subsets does not hold for sentences with the focus operator *only*, however. Consider example (66).

66. Only Bruce speaks a Romance language.  
 67. #Only Bruce speaks French.

English speakers typically deny that (66) entails (67), on the grounds that Bruce might speak Spanish or Italian, and not French. Based on observations such as this, von Fintel (1999) argues that the first meaning component of *only*, the presupposition, does not contain a downward-entailing operator. This explains why the standard diagnostic for downward entailment fails. For example, *Bruce speaks a Romance language* does not entail *Bruce speaks French*. What about the second meaning component, the assertion? It turns out that, in this meaning component, sentences with the focus operator *only* do license inferences from a set to its subsets. The example in (66) entails (68).

68. For all  $x$  [ $x \neq$  Bruce ], it is not the case that  $x$  speaks French  
 (All the others being contrasted with Bruce do not speak French).

So, the second meaning component of the focus operator *only* apparently does contain a downward-entailing expression. We assume that the downward-entailing expression is negation, or the semantic equivalent of negation. Thus, although sentences with *only* (e.g., (66)) lack an overt downward-entailing expression, they contain a *covert* downward-entailing expression. The covert downward-entailing expression appears in the assertion. The acquisition of the covert meaning component of sentences with a focus operator represents another challenge to be confronted by the experience-dependent approach to language development.

Another test of downward entailment, as we saw, is the licensing of conjunctive entailments when disjunction appears in the scope of a downward-entailing expression. Let us apply this test to sentences with the focus operator *only*. Consider (69).

69. Only Bruce speaks French or Spanish.  
 70. #Only Bruce speaks French *and* only Bruce speaks Spanish.  
 71. For all  $x$  [ $x \neq$  Bruce ], it is not the case that  $x$  speaks French *and* it is not the case that  $x$  speaks Spanish

Note first that (69) does not license the conjunctive entailment in (70). There are circumstances in which (66) is true, but where (70) is false, such as the circumstances in which Bruce speaks just French, or just Spanish, but not both. So, disjunction in the presupposition of sentences with *only* does not create conjunctive entailments, hence *only* is

---

<sup>5</sup> There is general agreement that (65) is entailed by (63), but the status of the meaning component represented in (64) is more controversial: Horn (1969, 1996) and von Fintel (1999) argue that it is a presupposition, whereas Atlas (1993, 1996) and Herburger (2000) take it to be the assertion of the sentence. We will adopt Horn's terminology, and will refer to the first proposition, about the element in focus, as the 'presupposition'; and we will refer to the second proposition, about the individuals being contrasted with the element in focus, as the 'assertion.' We intend no theoretical commitments by adopting this terminology.

not downward entailing in this meaning component. By contrast, (69) entails the conjunctive statement in (71). Hence *only* passes another test for being downward entailing, but just in one of its meaning components. We might say that the focus operator *only* is ‘partially’ downward entailing.

Recent experimental research has sought to determine whether or not children know that sentences with the focus operator *only* contain a hidden downward-entailing operator (negation, or its semantic equivalent). As noted earlier, 4- to 5-year-old children appear to know that *or* licenses conjunctive entailments in certain downward-entailing contexts, e.g., under negation, and in the Restrictor of the universal quantifier *every*. So, children’s interpretation of *or* can be used to assess their knowledge of the semantics of *only* (Goro, Minai, & Crain, 2006). It seems unlikely that there is relevant evidence in the input about the entailment of sentences with *only*. On the other hand, if children do not acquire knowledge of the entailment from experience, then children should have to access this knowledge regardless of differences in the language they are learning.

With these objectives in mind, experiments were conducted with English-speaking children (using sentences with *only...or*) and with Japanese-speaking children (using ones with *dake ...ka...*). The research strategy was to investigate their interpretations of disjunction in the overt meaning component and in the covert downward-entailing component of sentences with the focus operator *only/dake*. One of the test sentences is given in (72).

72. a. Only Bunny Rabbit will eat a carrot or a green pepper.  
 b. Usagichan-dake-ga ninjin ka piiman-wotaberu-yo.  
 rabbit-*only*-NOM carrot *or* green pepper-ACC eat-dec

Under the decomposition analysis, the meaning of (72) can be partitioned into the two conjoined propositions in (73).

73. a. *Presupposition*. Bunny Rabbit will eat a carrot or/ka a green pepper  
 b. *Assertion*. Everyone other than Bunny Rabbit will not eat a carrot or/ka a green pepper

Within the *presupposition* component, the disjunction operator *or* yields *disjunctive* truth conditions: *Bunny Rabbit will eat a carrot or will eat a green pepper*. Suppose, first, that children assign the correct interpretation to *or* within the *presupposition* component.

74.

Condition 1		
	Carrot	Green pepper
Winnie the Pooh	*	*
Bunny Rabbit	✓	*
Cookie Monster	*	*

If so, children should avoid the conjunctive entailment of disjunction in the presupposition, so they should accept (72) in the situation where Bunny Rabbit ate a carrot but not a green pepper. The entire truth conditions are schematically represented in (74).

By contrast, within the assertion meaning component of (69), *or* appears in a downward-entailing environment. Therefore, it licenses the conjunctive entailment – that

75.            Condition 2

	Carrot	Green pepper
Winnie the Pooh	*	*
Bunny Rabbit	✓	*
Cookie Monster	*	✓

everyone else will not eat a carrot and they will not eat a green pepper. Consequently, if children assign the correct interpretation to *or* within the assertion, they should reject (72) in the situation represented in (75) on the grounds that Cookie Monster ate a green pepper (while, again, Bunny Rabbit ate a carrot but not a green pepper).

Summarising, if English/Japanese children assign the inclusive-*or* interpretation to *or/ka*, then they should accept the test sentences in Condition 1, but they should reject them in Condition 2. By contrast, if children assign a different semantics to *or/ka*, then they could also accept the test sentence in Condition 2. We conducted experiments with English-speaking and Japanese-speaking children, to compare their linguistic behaviour. The experiments in English and Japanese were identical in design, with only minimal changes in some of the toy props. The experiment employed the Truth-Value Judgment task. There were two experimenters. One of them acted out the stories using the toy props, and the other manipulated the puppet, Kermit the Frog. While the story was being acted out, the puppet watched along with the child subject. In each trial, the story was interrupted – after the introduction of the characters and a description of the situation – so that the puppet could make a prediction about what he thought would happen. Then, the story was resumed, and its final outcome provided the experimental context against which the subject evaluated the target sentence, which had been presented as the puppet’s prediction. The puppet repeated his prediction at the end of each story, and then the child subject was asked whether the puppet’s prediction had been right or wrong. Twenty-one English-speaking children (mean age 5,0) participated in the experiment, and 20 Japanese-speaking children (mean age 5,4).

The main finding was that both English-speaking children and Japanese-speaking children consistently accepted the test sentences in Condition 1, and consistently rejected the test sentences in Condition 2. The two groups of children showed no significantly different behaviour in interpreting disjunction within sentences containing a focus operator, *only* versus *dake*. Most crucially for our purpose, the high rejection rate in Condition 2 shows that children assigned conjunctive entailments to disjunction in the



assertion component of the test sentences. This, in turn, suggests that they assigned the same semantics to the disjunction operator in each language, despite the differences in input. Children's consistent rejections of the test sentences in Condition 2 provide evidence that they are computing the covert meaning component that is associated with focus operators. As we saw, the covert meaning component expresses a (negative) proposition about a set of individuals that are being contrasted with the element in focus. The findings clearly establish children's ability to compute such contrast sets, although this ability has been questioned by some researchers (cf. Paterson, Liversedge, Rowland, & Filik, 2003).

In our experiments, there was no evidence of a significant effect of input on the acquisition of disjunction. Both English-speaking children and Japanese-speaking children were able to compute the derived logical truth conditions of disjunction. The experience-dependent account (e.g., Tomasello, 2000, 2003) would be hard-pressed to explain the findings of the present studies. In particular, it is hard to see how the experience-dependent account could explain the fact that the same lexical item is interpreted in two different ways in the same sentence. No straightforward learning algorithm would do the trick. It is difficult to see how 'cut and paste' operations, like those proposed on the experience-dependent account, could be used to explain the dual interpretations of a single expression of disjunction in sentences with focus operators. We also leave it as a challenge to such models to account for the absence of any impact of input characteristics on the outcome of acquisition.

## 9. THE CONTINUITY HYPOTHESIS

According to the nativist perspective, children are expected to sometimes follow developmental paths to the adult grammar that would be very surprising from a data-driven perspective. Of course, normal children eventually internalise grammars that are equivalent to those of adults. But a child who has not yet achieved a dialect of English can still be speaking a natural language – albeit one that is (metaphorically) a foreign language, at least somewhat, from an adult perspective. And interestingly children often do exhibit constructions that are not available in the local language – but ones that are available in other adult languages. This is unsurprising if children are free to try out various linguistic options (compatible with Universal Grammar) before 'setting parameters' in ways that specify a particular natural language grammar. This proposal about the course of language development is referred to as the Continuity hypothesis (Pinker, 1984; Crain, 1991; Crain & Pietroski, 2001). According to one version of the Continuity hypothesis, child language can differ from the local adult language only in ways that adult languages can differ from each other (Crain, 1991; Crain & Pietroski, 2001). The idea is that at any given time, children are speaking a possible human language, just not necessarily the particular language that is being spoken around them. Such mismatches between child and adult language are seen to be among the strongest arguments for a Universal Grammar. We first discuss one example of continuity in syntactic development, and then we discuss an example of continuity in semantic development.

### 9.1. An Example of Continuity in Syntactic Development

In most English *wh*-questions (i.e., questions that begin with *wh*-words: *why*, *what*, *where*, *who*, etc.) the *wh*-question word must be immediately followed by an inflected auxiliary verb (i.e., a tensed form of *be*, *do*, *can*, *have*, etc.). Hence, the examples in (76a–d) are acceptable, whereas the examples in (76e–h) are not acceptable.<sup>6</sup>

76. a. Why are you here?  
 b. What do you want to do?  
 c. Where is he going?  
 d. Who don't you want to win the game?  
 e. \*Why you are here?  
 f. \*What you want to do?  
 g. \*Where he is going?  
 h. \*Who you don't want to win the game?

In Italian, the *wh*-word corresponding to English *why* is *perché*. Italian *perché* differs from other Italian *wh*-words in simple questions (for analysis, see Rizzi, 1997). As the example in (77) illustrates, the adverb *già* as well as an entire subject phrase (*I tuoi amici*) can intervene between *perché* and the inflected verb (*hanno*). No linguistic material can intervene with other *wh*-words in Italian.<sup>7</sup>

77. Perché (I tuoi amici) già hanno finito il lavoro?  
 Why (the-pl your friends) already have-3pl finished the-sg work  
 “Why (your friends) already have finished the work?”

However, in complex *wh*-questions with *perché*, the intervention of short adverbs or a subject phrase is prohibited in questions like (78) (if the question is asking about the reason for someone's resignation).

78. Perché ha detto che si dimetterà?  
 Why have-3sg said that self resign-3sg/future  
 “Why did he say that he would resign?”

So, complex questions like (78) pattern the same way in both English and in Italian, whereas the simple questions differ, at least for the question words corresponding to ‘why.’ In both languages, the inflected verb must immediately follow the question word in the complex question.

In studies of child English, it has frequently been noted that children produce nonadult *why*-questions. More specifically, children's simple *why* questions are often

<sup>6</sup> The exception is *how come* (see footnote 7).

<sup>7</sup> The exception is *come mai* (how come in English; see footnote 6).

followed by a subject phrase, as in (76e–h). Moreover, such nonadult *why*-questions persist in children's speech well after they consistently produce adult questions with other *wh*-words. Adopting the Continuity hypothesis, Thornton (2004) suggested that children of English-speaking adults initially treat the question-word *why* in the same way as Italian adults treat *perché* (cf. de Villiers, 1991). If this is correct, an English-speaking child should differ from English-speaking adults in the way she forms simple *why*-questions, but the child should parallel English-speaking adults in producing well-formed long-distance *why*-questions. From a data-driven perspective, this pattern is surely not anticipated. Since simple questions are more frequent in the input, these should become adult-like in advance of more complex questions, all other things being equal.

By age 3, AL produced adult-like *wh*-questions for all *wh*-words except for *why*, correctly inverting 87% of the time. In the same time period, from 2 to 3-years, *why*-questions were inverted only 40% for positive questions and even less for negative ones. The non-inversion with *why* persisted until AL was 5 and a half years old, as illustrated in (79).

79. Why the monster goed away and never comed back? (3,3)  
 Why the lights are on in my school? (4,10)  
 Why daddy took the broken-down car? (5,3)

From the time AL was 3 to 5 and a half years of age she produced 83 complex *wh*-questions. There were 62 questions with *wh*-phrases other than *why*, and these questions were correctly inverted 100% of the time. Only 4 of the 21 complex questions with *why* were non-adult. Adult-like complex *wh*-questions are illustrated in (80).

80. What do you think is under daddy's chair? (3,5)  
 How do you think he can save his wife and her at the same time? (4,9)  
 Why do you think Santa's not coming this year? (3,10)  
 Why do you think mummy would not wanna watch the show? (4,6)

In short, the production data suggest that an English-speaking child analyses *why*-questions like the corresponding questions are analyzed in Romance languages, such as *perché* in Italian. In producing simple *why*-questions, moreover, AL was ignoring abundant evidence in the input indicating a mismatch between her grammar and that of adult speakers in the same linguistic community. However, AL adhered to the grammatical principles that govern all natural languages, producing adult-like complex *why*-questions, but nonadult simple *why*-questions. See Thornton (2004) for several further parallels between AL's *why*-questions and those of adult speakers of Italian; see Rizzi (1997) for an analysis of questions in Italian.

## 9.2. An Example of Continuity in Semantic Development

An example of continuity in semantic development is based on an observation by Goro (2004) who notes that, in Japanese, simple negative sentences with disjunction

do not license conjunctive entailments. We noted earlier that Japanese does indeed generate conjunctive entailments for disjunction. It turns out, however, that simple negative sentences lack the conjunctive entailments associated with de Morgan's laws, at least for adults. In Japanese, for example, the translation of the English sentence *Max's computer did not come with Ichat or Isync* asserts that Max's computer didn't come with Ichat or it didn't come with Isync; the 'not both' reading, rather than the 'neither' reading.

As a further example, adult Japanese-speakers interpret (81) to mean that the pig didn't eat the carrot *or* didn't eat the pepper. Despite the appearance of *ka* under negation in surface syntax, *ka* is interpreted by adults as if it has scope over negation.

81. *Butasan-wa ninjin ka pi'iman-wo tabe-nakat-ta*  
 pig-TOP pepper or carrot-ACC eat-NEG-PAST  
 Literally: "The pig didn't eat the pepper or the carrot"  
 Meaning: "The pig didn't eat the pepper or the pig didn't eat the carrot"

Based on considerations of language learnability, Goro (2004) hypothesised that Japanese-speaking children would, nevertheless, interpret the disjunction operator *ka* as licensing conjunctive entailments in simple negative sentences like (81).<sup>8</sup> The prediction was that Japanese-speaking children would interpret such sentences in the same way as English-speaking children and adults, despite the absence of this interpretation for adult speakers and, hence, the absence of evidence for this interpretation in the input to children.

In brief, Goro's proposal is that the semantics of natural language disjunction is innately specified as inclusive-*or*. However, the interaction of disjunction with negation is subject to cross-linguistic variation, as proposed by Szabolcsi (2002). In one class of languages, including English and German, disjunction may be interpreted under local negation, whereas it must be interpreted outside the scope of local negation in another class of languages, including Japanese and Hungarian, regardless of its surface position in such languages. To adopt some technical terminology, Goro proposed that disjunction is a *positive polarity item* in Japanese (like *some* in English), but not in English. By definition, a positive polarity item must be interpreted as if it were outside the scope of negation, rather than in its scope. In Japanese, then, the disjunction operator *ka* appears to have the truth conditions associated with exclusive-*or* (not both) in simple negative sentences, whereas *or* creates conjunctive entailments (neither) in the corresponding sentences of English.

Goro's next observation was that the alternative values of the 'positive polarity parameter' for disjunction stand in a subset/superset relation, with English exemplifying the

---

<sup>8</sup> To convey the 'neither' interpretation, Japanese employs sentences with a *~mo ~mo* construction, which is semantically similar to the use of conjunction (*and*) in English.

subset value of the parameter, and Japanese exemplifying the superset value. He reasoned that this situation would lead to a ‘subset problem’ unless children acquiring Japanese initially select the parameter value corresponding to English. Since adult speakers do not make such entailments, it is unlikely that children learn to make them based on the adult input. The reason children should appear more logical than adults in Japanese, Goro suggested, is that children adhere to a principle of language acquisition: the *semantic subset principle* (Crain, Ni, & Conway, 1994).

The semantic subset principle enforces an ordering on the values of certain parameters, where one value makes a sentence true in a subset of the circumstances that make it true on the other value. The semantic subset principle compels children to adopt the subset value of the parameter as their initial interpretation; this value is abandoned only on the basis of positive evidence in the local language. If children adopted the superset value instead, they would generate sentences that are not in the local language, in addition to sentences in the local language. This raises a familiar learnability problem: in the absence of negative evidence, it is difficult to see how children would purge their grammars of the means for generating sentences that are not acceptable in the local language. To avoid this problem, the semantic subset principle orders the value of parameters.

To investigate this solution to the ‘logical problem of language acquisition,’ Goro and Akiba (2004) examined Japanese children’s interpretation of negated disjunctions in sentences like (81) using the Truth Value Judgment task. They interviewed 30 Japanese-speaking children (mean age 5,3) as well as a control group of Japanese-speaking adults. On a typical trial, subjects were asked to judge whether or not (81) was an accurate description of a situation in which the pig had eaten the carrot but not the green pepper. The findings were precisely as anticipated. Japanese-speaking adults uniformly accepted the target sentences (such as (81)), whereas children rejected them 75% of the time. The findings are even more compelling once the data from four children, who responded like adults, were set aside. The remaining 26 children rejected the target sentences 87% of the time.

The pattern of responses by Japanese-speaking children are difficult to explain on a ‘data driven’ account of language development, since Japanese-speaking children interpreted negated disjunctions as licensing conjunctive entailments, whereas Japanese-speaking adults did not. On the other hand, the findings are consistent with the Continuity hypothesis, according to which child language is expected to diverge from the local adult language, but only in ways that adult languages can differ from each other (see, e.g., Crain, 1991, 2002; Crain, et al. in press; Crain & Pietroski, 2001, 2002; Thornton, 1990, 2004). If children acquired the semantics of the disjunction operator from experience using general-purpose learning algorithms, the fact that *ka* in Japanese receives a ‘not both’ interpretation in both positive and simple negative sentences would be expected to affect the acquisition process. Specifically, adult input with *ka* in both positive sentences and in simple negative sentences could mislead Japanese children, prompting them to conclude that *ka* is a ‘non-logical’ connective, i.e., one that does not

obey De Morgan's laws. Fortunately, this does not happen. The theory of Universal Grammar anticipates that children learning any language should interpret disjunction as inclusive-*or*, regardless of the input children encounter. The fact that this was found in a language in which the input from adults violates De Morgan's laws provides further evidence for the continuity assumption (see Jing, Crain, & Hsu, 2005, for a similar analysis of child Chinese).

## 10. CONCLUSION

In this chapter, we have tried to give readers a flavour of some past and current research in the acquisition of syntax and semantics. The presentation of research findings was framed within what is arguably the most central debate in the field: the degree to which human language acquisition is 'data driven' or experience dependent, and the degree to which it is determined by the human genome. Of course, no definitive answer was offered here, one way or the other. At most we have pointed out concepts and consequences of both the nativist approach and the experience-dependent approach to language development. In our view, the experience-dependent approach is implausible in several respects. First, its viability depends on the abilities of children to keep highly detailed records of attested structures in the input. Second, it seems incapable of explaining both language-specific and cross-linguistic generalisations, both in syntactic and in semantic development. Third, it fails to explain how children acquire the 'hidden' meanings of sentences with focus operators, such as *only*. Fourth, it fails to explain the universal mastery of certain aspects of syntax (*c-command*) and semantics (downward entailment, inclusive-*or*). Fifth, it lacks an account of children's nonadult linguistic behaviour, both in syntax and in semantics. Until these challenges are met, the nativist approach appears more convincing. But, this is an empirical matter, after all. Only the future will tell if the nativist approach or the experience-dependent approach is closer to the truth.

## REFERENCES

- Atlas, J. D. (1993). The importance of being "Only": Testing the neo-Gricean versus neo-entailment paradigms. *Journal of Semantics*, 10, 301–318.
- Atlas, J. D. (1996). 'Only' noun phrases, pseudo-negative generalized quantifiers, negative polarity items, and monotonicity. *Journal of Semantics*, 13, 265–328.
- Borer, H., & Wexler, K. (1987). The maturation of syntax. In: T. Roeper, & E. Williams (Eds), *Parameter setting*. D. Reidel, Dordrecht, Holland.
- Boster, C., & Crain, S. (1993). On children's understanding of *every* and *or*. In: C. Smith (Ed.), *Conference Proceedings: Early Cognition and the transition to Language*, Presented by College of Liberal Arts, Center for Cognitive Science, Department of Psychology, and Department of Linguistics, University of Texas, Austin, TX.

- Bowerman, M. F. (1998). The “no negative evidence” problem: How do children avoid constructing overly general grammar? In: J. A. Hawkins (Ed.), *Explaining language universals*. Oxford: Blackwell.
- Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In: J. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In: A. Belletti, & L. Rizzi (Eds), *Structures and beyond*. Oxford: Oxford University Press.
- Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. *Proceedings of the 25th Boston University conference on language development*, Somerville, MA, Cascadilla Press. (pp. 157–168).
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA, The MIT Press.
- Chomsky, N. (1971). *Problems of knowledge and freedom*. New York: Pantheon Books.
- Chomsky, N. (1975). *Reflections on language*. New York: Pantheon Books.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- Conroy, A., & Thornton, R. (2005). Children’s knowledge of Principle C in discourse. *Tokyo Conference on psycholinguistics*. Tokyo: Hituzi Publishing Company.
- Cowie, F. (1999). *What’s within: Nativism reconsidered*. New York: Oxford University Press.
- Crain, S. (1991). Language acquisition in the absence of experience. *Behavioural and Brain Sciences*, 4, 597–650.
- Crain, S. (2002). The continuity assumption. In: I. Lasser (Ed.), *The process of language acquisition*. (3–24). New York: Peter Lang.
- Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In: S. Berman, J. Choe, & J. McDonough (Eds), *Proceedings of 16th North Eastern Linguistics Society*, Amherst, MA, GLSA.
- Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, 63, 522–543. Reprinted in Otero, C. (Ed.), *Noam Chomsky: Critical assessments*. London, England: Routledge.
- Crain, S., & Pietroski, P. (2001). Nature, nurture and Universal Grammar. *Linguistics and Philosophy*, 24, 139–185.
- Crain, S., & Pietroski, P. (2002). Why language acquisition is a snap. *The Linguistic Review*, 19, 163–183.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Crain, S., Gardner, A., Gualmini, A., & Rabbin, B. (2002). Children's command of negation. *Proceedings of the 3rd Tokyo conference on psycholinguistics*, Tokyo, Hituzi Publishing Company. (pp. 71–95).

Crain, S., Goro, T., & Thornton, R. (2006). Language acquisition is language change. *Journal of Psycholinguistic Research*.

Crain, S., Gualmini, A., & Pietroski, P. (2005). Brass tacks in linguistic theory: Innate grammatical principles. *Proceedings of the 1st annual AHRB conference on innateness and the structure of the mind*, University of Sheffield, England (pp. 175–197).

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing, and modularity. In: C. Clifton, L. Frazier, & K. Rayner (Eds), *Perspectives on sentence processing* (pp. 443–467). Hillsdale, NJ: Lawrence Erlbaum.

deVilliers, J. (1991). Why questions? In: T. L. Maxfield, & B. Plunkett (Eds), *Papers in the acquisition of wh: Proceedings of the UMass Roundtable*, University of Massachusetts Occasional papers, Amherst, MA (pp. 155–174).

Feynman, R. P. (2005). *Six easy pieces*. Perseus Books, New York: Basic Books.

Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, 7(5), 219–224.

Goro, T. (2004). *The emergence of Universal Grammar in the emergence of language: The acquisition of Japanese logical connectives and positive polarity*. Unpublished manuscript, University of Maryland at College Park.

Goro, T., & Akiba, S. (2004). The acquisition of disjunction and positive polarity in Japanese. *Proceedings of the 23rd West Coast conference on formal linguistics*, Summerville, MA, Cascadilla Press (pp. 251–264).

Goro, T., Minai, U., & Crain, S. (2006) Bringing out the logic in child language. In: L. Bateman, & C. Ussery (Eds), *Proceedings of the Thirty-Fifth Annual Meeting of the North East Linguistics Society*. South Carolina: BookSurge Publishing.

Grice, P. (1975). Logic and conversation. In: P. Cole, & J. Morgan (Eds), *Syntax and semantics 3: Speech acts*, (pp. 41–58). New York: Academic Press.

Gualmini, A., & Crain, S. (2002). Why no child or adult must learn De Morgan's laws. *Proceedings of the 24th annual Boston University conference on language development*, Summerville, MA, Cascadilla Press (pp. 367–378).

Gualmini, A., & Crain, S. (2004). Operator conditioning. *Proceedings of the 26th annual Boston University conference on child language development*, Cascadilla Press, MA.



- Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. *Proceedings of semantics and linguistic theory 11*, New York University.
- Gualmini, A., Meroni, L., & Crain, S. (2003). an asymmetric universal in child language. *Proceedings of Sinn und Bedeutung VI*, Arbeitspapiere des Fachbereichs Sprachwissenschaft, Konstanz, Germany (pp. 136–148).
- Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, 20, 667–696.
- Hankamer, J. (1979). *Deletion in coordinate structures*. New York: Garland.
- Herburger, E. (2000). *What counts: Focus and quantification*. Cambridge, MA: MIT Press.
- Horn, L. (1969). A presuppositional approach to *only* and *even*. *Proceedings of the Chicago Linguistic Society*, 5, 98–107.
- Horn, L. (1996). Presupposition and implicature. In: S. Lappin (Ed.), *Handbook of contemporary semantic theory*, (pp. 299–319). Oxford: Blackwell.
- Jing, C., Crain, S., & Hsu, C. (2005). The interpretation of focus in Chinese: Child vs. adult language. *Proceedings of the 6th Tokyo conference on psycholinguistics*. Hituzi Publishing Company, Tokyo.
- Lasnik, H. (1976). Remarks on coreference. *Linguistic Analysis*, 2, 1–22.
- Lasnik, H., & Crain, S. (1985). On the acquisition of pronominal reference. *Lingua*, 65, 135–154.
- Legate, J. A., & Yang, C. D. (2002). Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review*, 19, 151–162.
- MacWhinney, B. (2000). *The CHILDES project” Tools for analysing talk*. 3rd Edn., Mahwah, NJ: Erlbaum Associates.
- MacWhinney, B. (2004). A multiple process solution to the logical problem of child language. *Journal of Child Language*, 31, 883–914.
- Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46, 53–85.
- Marcus, G. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37, 243–282.
- Marcus, G. F. (1999). Language acquisition in the absence of explicit negative evidence: Can simple recurrent networks obviate the need for domain-specific learning devices? *Cognition*, 73, 293–296.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science*, 283, 77–80.

- Merchant, J. (2005). Fragments and ellipsis. *Linguistics and Philosophy*, 27, 661–738.
- Morgan, J. (1973). Sentence fragments and the notion ‘sentence’. In: B. B. Kachru, R. Lees, Y. Malkiel, A. Pietrangeli, & S. Saporta (Eds), *Issues in linguistics: Papers in honor of Henry and Rene’e Kahane* (pp. 719–752). Chicago, IL: University of Illinois Press.
- Morgan, J. L. (1989). Sentence fragments revisited. In: B. Music, R. Graczyk, & C. Wiltshire (Eds), *Papers from the 25th annual regional meeting of the Chicago Linguistic Society, part two: Parasession on language in context*, Chicago, IL (pp. 228–241).
- O’Leary, C., & Crain, S. (1994). Negative polarity (a positive result) and positive polarity (a negative result). *Presented at the 18th annual Boston University conference on language development*, Boston, MA.
- Paterson, K. L., Liversedge, S. P., Rowland, C., & Filik, R. (2003). Children’s comprehension of sentences with focus particles. *Cognition*, 89, 263–294.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA.: Harvard University Press.
- Pinker, S. (1990). Language acquisition. In: D. N. Osherson, & H. Lasnik (Eds), *An invitation to cognitive science* (Vol. 1, pp. 107–133). Cambridge, MA: MIT Press.
- Pullum, G., & Scholz, B. (2002). Empirical assessment of the stimulus poverty argument. *The Linguistic Review*, 19, 9–50.
- Rizzi, L. (1997). The fine structure of the left periphery, In: L. Haegeman (Ed.), *Elements of grammar*, Dordrecht: Kluwer Academic Publisher.
- Rumelhart, D. E., & McClelland, J. L., (1986). On learning the past tenses of English verbs. In: J. L., McClelland, D. E. Rumelhart, & The PDP Research Group (Eds), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Psychological and Biological Models. Cambridge, MA: MIT Press.
- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 19–26.
- Smith, N. (1996). Structural eccentricities, *Glott International*, 2, 7–8.
- Szabolcsi, A. (2002). Hungarian disjunctions and positive polarity. In: I. Kenesei, & P. Siptar (Eds), *Approaches to Hungarian*, (Vol. 8, pp. 217–241). Budapest: Akademiai Kiado.
- Thornton, R. (1990). *Adventures in long-distance moving: The acquisition of complex Wh-questions*. Doctoral dissertation, University of Connecticut, Storrs.
- Thornton, R. (1995). Children’s negative questions: A production/comprehension asymmetry. In: J. Fuller, H. Han, & D. Parkinson (Eds), *Proceedings of eastern states conference on linguistics*, CLC Publications, Cornell, New York.

- Thornton, R. (1996). Elicited production. In: D. McDaniel, C. McKee, & H. S. Cairns (Eds), *Methods for assessing children's syntax*. Cambridge, MA: MIT Press.
- Thornton, R. (2004). Why continuity. In: A. Brugos, L. Micciulla, & C. E. Smith (Eds), *Proceedings of the 28th Boston University conference on language development*, Cascadia Press, Somerville, MA (pp.620–632).
- Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, 11, 61–82.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- van der Wal, S. (1996). *Negative polarity items and negation: Tandem acquisition*. Groningen Dissertations in Linguistics 17, University of Groningen.
- von Stechow, K. (1999). NPI-licensing, Strawson entailment, and context dependency. *Journal of Semantics*, 16, 97–148.
- Yang, C. D. (2004). Universal Grammar, statistics or both. *Trends in Cognitive Science*, 8, 451–456.

## Chapter 29

### Learning to Read\*

*Richard K. Wagner, Shayne B. Piasta, and Joseph K. Torgesen*

Interest in reading on the part of psychological researchers can be traced nearly to Wundt's laboratory in Leipzig in the late 1800s. Almost a century ago, this interest culminated in the publication of Huey's *The Psychology and Pedagogy of Reading* in 1908 (Rayner & Pollatsek, 1989). Of significance in the present context is that interest in the teaching and learning of reading—the pedagogical part of Huey's treatise—arose side by side with interest in understanding the perceptual and cognitive machinery that makes reading possible. After several decades of research activity, the topic of learning to read essentially was set aside in the 1920s with the behavioral revolution in American psychology, emerging again only in the 1960s with the emerging prominence of the field of cognitive psychology. Although interest in learning to read produced a steady stream of knowledge from the 1960s on, recent years have been characterized by a virtual flood of knowledge. This rapid increase in knowledge resulted from the infusion of many millions of research dollars into reading research in response to a growing appreciation by policymakers of the vital importance of reading to active participation in 21st century society.

The degree to which policy and science have become intertwined has become a distinguishing feature of research on learning to read. Adam's (1990) influential and popular tome *Beginning to Read* was produced in response to a Congressional mandate to provide guidance “as to how schools might maximize the quality of phonic instruction in beginning reading programs” (p. 29). The similarly influential and popular *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffith, 1998) resulted from a request by the United States Departments of Education and Health and Human Services to the National Academy of Sciences to study how reading difficulties might be prevented. (The National Academy of Sciences was established by Congress early in the country's history for the purpose of informing the government on matters of national interest: two of its first charges were to study how a currency could be developed that would resist counterfeiting, and how wooden battleships might be protected from cannonballs—the solution was to reinforce the sides with iron cladding. Soon after the

---

\*Preparation of this manuscript was supported by NICHD Grant P50 HD052120 and by IES Grant R305G030104.

National Academy report, a National Reading Panel (2000) was established by an act of Congress, with rumors suggesting the motivation for creating the panel was unhappiness with some parts of the National Academy's report. Rayner, Foorman, Perfetti, Pesetsky, & Scidenberg (2001) article titled "How Psychological Science Informs the Teaching of Reading" was not written at the request of a government agency, but was published in the journal *Psychological Science in the Public Interest*, which was a new effort by the American Psychological Society to apply psychological research to issues of public interest. Today, the three authors of this chapter work at the Florida Center for Reading Research, a center funded in part by research dollars associated with Reading First and the No Child Left Behind Act of 2002, two recent Federal efforts improve the reading skills of the nation's children.

The explosive growth in knowledge about reading due in part to increased Federal funding for reading research has occurred primarily in three areas. The first, not surprisingly, encompasses how elementary-school age children learn to read. The second area of growth is centered on how prereaders "learn to learn to read." By this we mean knowledge about developing language and literacy skills that we now recognize as rudimentary precursors of full-blown reading. For many years, the consensus was that learning to read began with the initiation of formal reading instruction in first grade, provided that children were "ready" by virtue of possessing basic perceptual and motor skills; such readiness skills were demonstrated by performing tasks under the watchful eye of school psychologists such as using scissors and skipping (not at the same time, of course!). The new view is that developing knowledge and skills that can be observed reliably in children as young as three serve as essential precursors that make learning to read possible, and sometimes even easy. The third area of explosive growth is in knowledge about why children fail to learn to read, an all too frequent occurrence despite advances in teaching practice. Our chapter correspondingly is divided into three major sections that address, in turn, learning to read, learning to learn to read, and failing to learn to read.

## 1. LEARNING TO READ

Reading refers to understanding a message from a writer. Doing so requires decoding the script the message has been written in. Most scripts are informative about both the meaning and pronunciation of the words used to convey the message. Our focus will be on learning to read the script of written English, and for not entirely jingoistic reasons: Much of what is known about learning to read is about learning to read alphabetic scripts in general, and English in particular. In addition, the major findings from the literature on learning to read other scripts are more similar to the findings regarding learning to read English than you would expect, based on how different scripts appear to be. The development of reading skill, and the characteristics of individuals who fail to learn to read well, are remarkably similar across European languages (Zeigler & Goswami, in press). When our world view is broadened to include Asian languages, some differences are found in the relative importance of key underlying skills in learning to read, but the fact that some individuals fail to learn to read appears to be universal regardless of the nature of the written language to be mastered (McBride-Chang et al., in press).

## 1.1. Phonology, Morphology, and Orthography

All existing writing systems represent aspects of their corresponding spoken language (Rayner et al., 2001). As such, the task for the beginning reader is one of connecting between the orthographic representations or marks on the printed page or computer screen and the morphological and phonological representations that correspond, respectively, to the meanings and pronunciations of words that beginning readers already carry around in their heads. To understand what learning to read entails, it is helpful to review some basic facts about phonology, morphology, and orthography that will prove relevant to the present context.

*Phonology.* For the purpose of understanding the role of phonology in learning to read, it is helpful to distinguish four levels at which speech can be represented (Crowder & Wagner, 1992). These four levels form a hierarchy. From bottom to top, they are the acoustic, the phonetic, the phonological, and the morphophonemic levels.

At the *acoustic* level, speech is represented by continuous waves of acoustic energy. This acoustic energy can be observed on a spectrogram, which displays acoustic energy in terms of frequency over time. The most important observation made is that virtually none of the spacing or breaks we perceive between words, syllables, or sounds within syllables is visible. That we perceive distinct words and parts of words is due to our perceptual and cognitive machinery, as opposed to a characteristic inherent in the physical signal itself.

At the *phonetic* level, speech is represented by phones, which are the universe of individual sounds made by speakers of all languages. At the phonetic level, the sounds of the ‘t’ in the words ‘top,’ ‘stop,’ and ‘pot’ are represented by three distinct phones. To confirm that these indeed are different sounds, hold your hand in front of your mouth while you say the three words. You will feel a relatively strong burst of air while pronouncing the sound of the ‘t’ in ‘pot’; a somewhat less strong burst of air while pronouncing the sound of the ‘t’ in ‘top’; and very little air movement while pronouncing the sound of the ‘t’ in ‘stop.’

At the *phonological* level, speech is represented by abstract phonemes, which refer to sound distinctions that signal changes in meaning in a given language. The sounds represented by the ‘f’ in ‘fan’ and the ‘p’ in ‘pan’ are different phonemes, which lead to different meanings associated with the words ‘fan’ and ‘pan.’ In contrast, the phones represented by the sounds of the ‘t’ in ‘top,’ ‘spot,’ and ‘pot’ all are representations of the /t/ phoneme, and hence are referred to as allophones of the phoneme /t/. The abstract character of phonemes derives from the fact that they represent categories of phones that signal meaning in a given oral language.

Finally, at the *morphophonemic* level, speech is represented by strings of phonemes that also represent morphemes or units of meaning. These strings are referred to as morphophonemes. Just as allophones were phones associated with a single phoneme, allomorphs are associated with a given morpheme or meaning unit. Examples of allomorphs

are the ‘sign’ part of the words ‘sign’ and ‘signature.’ Written English is morphophonemic in that pronunciations are represented for the most part but with compromises so as to convey meaning. Thus, the written words SIGN and SIGNATURE share the spelling SIGN despite the fact that the pronunciations of this segment differ in the two words.

*Morphology.* Meaning is represented in oral languages such as English morphologically. Morphology refers to the composition of a word with respect to the morphemes or ‘minimal meaningful elements’ (Bloomfield, 1933). Morphemes include word roots, suffixes, prefixes, and inflections (parts of words that indicate number, person, tense, or case) (Arnbak & Elbro, 2000). The suffix ‘er’ is a morpheme that denotes “one who does something,” as in the words ‘teacher,’ ‘preacher,’ and ‘bookmaker.’ Morphological knowledge becomes particularly useful as children become skilled readers. They encounter many new words, the meanings of which must be inferred from analysis of the word and the context in which it is found. Fortunately, many of these new words are morphologically related to known words. For example, a child who encounters the word EVIDENTIARY might infer its meaning in part by analogy to the known word EVIDENCE).

*Orthography.* Orthography refers to the system of marks used to represent pronunciation and meaning in writing. English orthography consists of the 26 upper- and lower-case letters, the numerals 0 through 9, punctuation marks, and a small number of other symbols (e.g., # to represent ‘pound,’ and more recently, @ to represent ‘at’ in e-mail addresses). Three major kinds of orthography or writing systems can be found today (Crowder & Wagner, 1992; DeFrancis, 1989; Gelb, 1952; Rayner et al., 2002; Rayner & Pollatsek, 1989). Alphabetic writing systems rely on a relatively small number of orthographic units or letters that map roughly onto the phonological level of speech representation. The fact that the sounds of the ‘t’ in ‘top,’ ‘spot,’ and ‘pot’ all are allophones of the /t/ phoneme is captured in the use of the single letter t to represent each of them. Examples of alphabetic writing systems include the Chinese *pin yan* system, English, German, Italian, Korean, and Spanish. Alphabets vary in the consistency of relations between letters and sounds (i.e., the number of sounds associated with a given letter). Alphabets with consistent mappings, which are referred to as shallow orthographies, include Italian and Dutch. Alphabets with inconsistent mappings, which are referred to as deep orthographies, include English (Frost, Katz, & Bentin, 1987). Syllabaries, the second kind of writing system, have orthographic units that correspond to syllables rather than phonemes. An example of a syllabary is the Japanese Kana. Finally, morpho-syllabic systems have orthographic units that represent syllables that also are morphemes. Examples of morpho-syllabic writing systems include Chinese and the Japanese Kanji system.

## **1.2. Phonological Awareness and Learning to Read Words: Reciprocal Developmental Relations**

Phonological awareness refers to an awareness of and access to the sound structure of one’s oral language (Anthony & Francis, 2005; Jorm & Share, 1983; Wagner & Torgesen,

1987). The pronunciations of words can be represented as strings of phonemes. All spoken English words can be represented using the phonemes presented in Table 1. Pronunciations vary regionally. What is typically thought of as American English is the dialect (called Mid-western American English) spoken by national newscasters. Consonants are produced by obstructing the escaping air using different means (e.g., lips, teeth, tongue on roof of mouth), with vocal chords either vibrating (voiced) or not (unvoiced). Vowels are produced by vibrating the vocal chords with no obstruction, but varying the shape of the vocal tract. Of the nearly 10 trillion possible combinations of 40 phonemes, only a relatively small number actually occur in spoken language, and many of these combinations occur in multiple words (Wagner et al., 1997). Thus, ‘cat,’ ‘rat,’ and ‘hat’ each consist of three phonemes, the first of which is different and the latter two are identical in the three words. These facts are represented by their spellings. Each has a different initial letter and identical medial and final letters. To a child with phonological awareness, the English writing system will appear to be sensible way of representing spoken words in print. A child lacking such awareness will find the English writing system to be much more arbitrary.

Just as we described a hierarchy of levels at which speech can be represented, there are differences in performance on phonological awareness tasks depending on the linguistic unit one must access and manipulate (Anthony & Francis, 2005; Anthony et al., 2003). One commonly used phonological awareness task is elision (Wagner, Torgesen, & Rashotte, 1999). A word or nonword is presented to an individual who is asked to repeat it. Then, the individual is asked to cut out a particular phonological segment and pronounce what remains (e.g., “Say past.” “Now say past without saying /s/”). Linguistic complexity refers to the fact that it is easier to manipulate larger phonological segments

Table 1  
Phonemes of Mid-western American English.

Consonants									
Stops	Voiced	<i>bat</i>	<i>dog</i>	<i>goat</i>					
	Voiceless	<i>pig</i>	<i>turtle</i>	<i>cat</i>					
Fricatives	Voiced	<i>dove</i>	<i>these</i>	<i>zebra</i>	<i>measure</i>				
	Voiceless	<i>farm</i>	<i>thistle</i>	<i>snake</i>	<i>sheep</i>	<i>hippopotamus</i>			
Affricatives	Voiced	<i>jump</i>							
	Voiceless	<i>chirp</i>							
Nasals		<i>mouse</i>	<i>night</i>	<i>wing</i>					
Liquids		<i>lamb</i>	<i>rooster</i>						
Glides		<i>yak</i>	<i>wombat</i>						
Vowels									
Front		<i>zebra</i>	<i>kit</i>	<i>hen</i>	<i>cat</i>				
Middle		<i>tern</i>	<i>pup</i>	<i>about</i>	<i>hood</i>				
Back		<i>dart</i>	<i>cod</i>	<i>pork</i>	<i>poodle</i>				
Diphthongs		<i>cow</i>	<i>python</i>	<i>goat</i>	<i>koi</i>	<i>deer</i>	<i>poor</i>	<i>ferret</i>	<i>snake</i>



than smaller ones. Preschool nonreaders are able to perform elision on compound words (e.g., “Say starfish.” “Now say starfish without saying fish.”), syllables (e.g., “Say ringing.” “Now say ringing without saying ‘ing’”), and to some extent, on onset-rime units within a syllable (e.g., “Say rant.” “Now say rant without saying /r/”). The onset part of a syllable is the initial consonant or consonant cluster. The rime part is the vowel and remaining consonants. Although an onset can be an individual phoneme, and it thus can appear that pre-school nonreaders are able to access and manipulate individual phonemes, elision tasks involving onset-rime can be performed by manipulating the more accessible rime unit. Preschool nonreaders typically are not able to perform elision on individual phonemes (e.g., “Say toad.” “Now say toad without saying the /d/”), and isolating a phoneme from a cluster of phonemes is more difficult still (e.g., “Say past.” “Now say past without saying the /s/ sound”).

Two streams of converging evidence support the view that phonological awareness plays a causal role in learning to read. First, causal modeling of longitudinal correlational data support such a role. Wagner et al. (1997) reported a 5-year longitudinal study of 216 children who were followed from kindergarten through fourth grade. This study used latent variables, with multiple observed indicators for each latent variable, in order to capture common variance among observed indicators. Because measurement error is by definition specific to individual observed indicators and latent variables consist only of common variance, latent variables can provide a clearer picture of true relations among constructs that is less distorted by measurement error, particularly when the indicators are administered on different occasions (thereby preventing time sampling error variance from being common among indicators) and vary in task format, strategy, and response (reducing the introduction of method variance shared by the observed indicators into the latent variable).

The three phonological constructs included in the study were phonological awareness, phonological memory, and rapid naming (Wagner & Torgesen, 1987). The construct of phonological awareness already has been described. The three measures of phonological awareness used in this study were phoneme elision (“Say the word cup. Now tell me what word would be left if I said cup without saying /k/”); sound categorization, a task requiring picking the phonologically odd one out of a quadruplet of items (fun, *pin*, bun, gun); phoneme segmentation (children listened to a word and then were asked to “Tell me each sound that you hear in the order that you hear it”), and blending phonemes presented in isolation into either words or nonwords. Phonological memory refers to coding information into a sound-based representation system for temporary storage (Baddeley, 1982, 1986; Conrad, 1964). The measures of phonological memory were digit span and memory for sentences. Rapid naming was included as a measure of the efficiency with which phonological codes can be retrieved from permanent memory. The measures of rapid naming required that children name series of digits or letters as rapidly as they were able to.

In addition to the phonological variables, vocabulary was included as a control variable as was the autoregressive effect of word-level reading at the previous time point.

Thus, predictors in the causal model of word-level reading in third grade were first-grade measures of the three phonological constructs, vocabulary, and word-level reading. The autoregressor was included to rule out an artifactual cause of a spurious causal influence. It is possible that the only causal variable operating on the outcome of word-level measured in third-grade is word-level reading in first grade. In other words, if you already are at the top of your class in reading in first grade, you are likely to continue to be above your peers because you are starting with an advantage. If the autoregressor of first-grade reading is omitted from the model, any variable included in the model that is correlated with the omitted first-grade reading variable will show a spurious causal influence.

The main results addressing a potential causal influence of phonological processing on subsequent word reading are presented in Table 2. The values in the table are structure coefficients, representing the causal influence of each exogenous variable (cause) independent of the other variables in the model. Phonological awareness exerted an independent causal influence on word-level reading at each of the three time intervals examined. Phonological memory did not exert an independent causal influence. Rapid naming exerted a causal influence initially, but it was developmentally limited. The growing influence of the autoregressor reflects the increasing stability of individual differences in word-level reading over time (i.e., the idea that a good reader in first grade is likely to be a good reader in third grade).

This study also examined a possible causal influence going in the other direction, namely, from print to phonological processing. These results are presented in Table 3. Here the outcome was subsequent phonological awareness and the exogenous or causal variable was letter knowledge. The results suggest an independent causal influence of letter knowledge on subsequent phonological awareness.

Table 2  
Independent causal influences of phonological processing and control variables on subsequent word-level reading.

Exogenous variable	Time period		
	K–2nd grade	1st–3rd grade	2nd–4th grade
<i>Phonological processing latent variables</i>			
Phonological awareness	.37***	.29*	.27***
Phonological memory	.12	–.03	.07
Rapid naming	.25*	.21*	.07
<i>Control variables</i>			
Vocabulary	.10	.22***	–.01
Autoregressor	.02	.27*	.57***

\* $p < .05$ , and \*\*\* $p < .001$ .

Source: Adapted from Wagner et al. (1997).

Table 3  
 Reciprocal causal influences of letter knowledge on subsequent phonological awareness.

Exogenous variable	Time period	
	K-2nd grade	1st-3rd grade
Letter knowledge	.23**	.12*
<i>Control variables</i>		
Vocabulary	.19*	.17**
Autoregressor	.43***	.70***

\* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$ .

Source: Adapted from Wagner et al. (1997).

In sum, causal influences between developing phonological processing variables and word-level reading variables are bi-directional: Individual differences in early phonological awareness play a causal role in subsequent individual differences in word-level reading skills. Conversely, individual differences in early letter knowledge play a causal role in subsequent individual differences in phonological awareness. Similar results have been reported by others (de Jong & van der Leij, 2002; Parrila, Kirby, & McQuarrie, 2004; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Perfetti, Beck, Bell, & Hughes, 1988).

The second stream of support for reciprocal causal relations between phonological awareness and word-level reading comes from intervention studies and studies of illiterate adults and poor readers. Support for a causal influence of phonological awareness on subsequent word reading comes from studies that train phonological awareness and then look for effects on word-level reading (e.g., Ball & Blachman, 1991; Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1991; Lundberg, Frost, & Peterson, 1988; National Reading Panel, 2000). For example, an early instructional study (Lundberg, Frost, & Peterson, 1988) showed that phonological awareness training, by itself, can have a significant impact on subsequent word-level reading and spelling skills. In this study, students engaged in a variety of activities to stimulate phonemic awareness and were not explicitly taught about letter sounds during the study. The activities were all conducted orally, and included, among others, rhyming games, games that required matching words on initial or final sounds, blending activities in which the teacher broke words into segments and the children had to guess the word, and activities involving counting phonemes. The study followed the children during their first year of reading instruction and found that students who had received the training in phonemic awareness acquired basic word identification (phonemic decoding) and spelling skills more readily than students who had not received the training.

Since that study was conducted, a number of other studies have found that instruction that directly illustrates the relevance of phonological training to reading and spelling activities consistently produces larger gains than phonemic awareness training by itself.

(Blachman, Ball, Black, & Tangel, 1994; Byrne & Fielding-Barnsley, 1995; Cunningham, 1990; Fuchs, et al., 2001; Hatcher & Hulme, 1999). This makes good sense, because the combination of letter-sound training and phonemic awareness training allows children to engage in simple activities that directly link phonemic awareness to word reading and spelling. For example, simple spelling can be illustrated by asking students to first identify the phonemic segments in a word, and then represent those segments with “letter tiles” they have learned to associate with those phonemes. A popular activity that combines reading and spelling, and that can be built on beginning levels of phonemic awareness and letter knowledge, involves ‘chaining’ words. The child might be shown a word like *mat*, and asked to pronounce the sound of each letter separately, and then blend them together using phonemic awareness skills. The teacher might then ask the child to make another word by substituting a different sound, represented by a different letter, at the beginning or end of the word.

Interventions aimed at improving phonological awareness and phonemic decoding may take either preventative or remedial forms. For instance, Torgesen and colleagues (1999) provided intensive phonological awareness and decoding training to students who began kindergarten with poor phonological skills. The training was in addition to the regular classroom instruction; students received 20 minutes of individual tutoring four times a week for two and a half years. The phonological awareness/synthetic phonics (PASP) group received the Auditory Discrimination in Depth program (Lindamood & Lindamood, 1984), in which students were taught to associate phonemes with how each is articulated (e.g., /p/ is a “lip popper”). Other phonemic awareness activities, such as using manipulatives to represent the individual phonemes heard in spoken words, were introduced, and students were eventually taught to decode via synthetic phonics. The effects of such training were compared to a second intervention group that received an embedded phonics (EP) program. These students were taught sight words and letter-sound correspondences, and lessons regarding phonics were taught through reading and writing activities. Eventually, students were encouraged to use their implicit phonics knowledge to phonemically decode words in basal readers. The two intervention groups were also compared to an active control group, where the tutoring received by students consisted of activities used with their regular classroom instruction reading programs, and a no-treatment control group, where students received only their regular classroom. Torgesen et al. (1999) concluded that PASP was slightly more effective in promoting students’ phonological awareness, phonemic decoding, and word reading than EP instruction, although both types of instruction led to gains over the control groups. In fact, the at-risk students who received the PASP intervention achieved word-level reading skills in the middle of the normal range by the end of second grade, demonstrating the effectiveness of phonological training in preventing reading difficulties.

Torgesen et al. (2001) provides an example of phonological training as remediation. The study employed the same interventions described above, but sampled only students (8–10 years of age) who had already demonstrated serious reading difficulties. The individual tutoring replaced these students’ regular resource room instruction for a total of 67.5h (two 50 minute sessions per day, 5 days a week). Generally, the interventions were

equally effective in facilitating substantial growth in accuracy and rate of reading words, nonwords, and connected text; any differences between instructional groups at posttest favored the PASP method, but these differences were slight and no longer reliable at a 2-year follow-up. When progress made via PASP or EP instruction was compared to previous years' reading growth in special education classes, both methods of teaching phonemic decoding skills were successful in remediating the reading difficulties of many students: 40% of students no longer required special education services and returned to the general education classroom.

Finally, support for the reciprocal causal influence of reading on phonological awareness comes from studies of illiterate adults. In one study, illiterate adults were unable to add or delete initial consonants from a spoken word. However, recently literate adults from the same community were able to do the task (Morais, Cary, Alegria, & Bertelson, 1979). Similarly, Chinese readers who have not learned to read an alphabet perform poorly on phonological awareness tasks (Read, Zhang, Nie, & Ding, 1986). Poor readers also show deficits in phonological awareness, a topic that will be addressed in the next section.

Some of the evidence that is cited in support of a causal relation between phonological awareness and learning to read is suspect because of alternative possible explanations (Castles & Coltheart, 2004). For example, if both phonological awareness and letter knowledge are trained in an intervention, as is commonly done, can we be sure it is the phonological awareness training that is responsible for a boost in word reading? Maybe the combination of training phonological awareness and letter knowledge actually is a proxy for phonics instruction, something we already know to be effective. On the other hand, it might be the case that one reason for the effectiveness of phonics instruction is that it produces a by-product of improved phonological awareness. Although many individual studies can be challenged on various grounds, the overall pattern of results, and in particular the longitudinal correlational studies that have included an autoregressor effect, support the existence of a causal influence of phonological awareness on word reading (Hulme, Snowling, Caravolas, & Carroll, 2005).

### **1.3. Oral Language Vocabulary and Reading Comprehension: A Second Case of Developmental Reciprocal Relations**

The research just reviewed dealt with word-level reading, measured by asking individuals to decode individual words or nonwords. But of course the purpose of reading is comprehension, and most reading for comprehension involves reading connected text for the meaning it conveys. Here, vocabulary knowledge appears to play an important role based on the magnitude of correlations between measures of vocabulary and measures of reading comprehension (Nagy, Berninger, Abbott, Vaughan, & Vermulen, 2003). Over years of development, children show remarkable growth in both vocabulary and reading comprehension. Might there be a causal relation between the development of vocabulary and reading comprehension?

Three alternative views of relations between vocabulary and reading comprehension have been proposed by Anderson and Freebody (1981; cited in Nagy, *in press*). First, an instrumental hypothesis states that simply knowing more words makes you better at reading comprehension, for the obvious reason that if you do not know the meanings of some words in a passage, your comprehension is likely to suffer. Second, a knowledge hypothesis states that individual differences in both vocabulary and reading comprehension are caused by individual differences in conceptual knowledge: if you have a lot of conceptual knowledge, you are likely to know more vocabulary words and also to be better able to comprehend what you read. Third, an aptitude hypothesis states that vocabulary and reading comprehension are correlated because individual differences in both are caused, at least in part, by individual differences in a fundamental ability such as general verbal ability.

More recently, Nagy (*in press*) proposed an updated version of the aptitude hypothesis. According to this view, individual differences in metalinguistic awareness (the ability to reflect upon language that can be manifested as phonological awareness, morphological awareness, or syntactic awareness) are causally related to individual differences in both vocabulary and reading comprehension. Relatedly, Carlisle (*in press*) argued that individual differences in one form of metalinguistic ability, naming morphological processing, results in greater breadth and depth of word knowledge, which in turn facilitates reading comprehension.

Structural equation modeling, and its special case of path analysis, can be used to test alternative causal models such as these. Wagner, Muse, and Tannenbaum (*in press*) tested four alternative causal models of developmental relations between vocabulary and reading comprehension by fitting path-analytic models to longitudinal data provided by 216 students for whom annual assessments of vocabulary and reading comprehension were available from second through fifth grade. The first model proposed that individual differences in reading comprehension exert a causal influence on vocabulary development. Many new vocabulary words are learned by inferring meaning from context when reading as opposed to being taught them directly. Children who are better at reading comprehension are more likely to figure out the meanings of new vocabulary words. If so, there will be a causal influence of individual differences in reading comprehension on subsequent individual differences in vocabulary. The second model proposed that individual differences in vocabulary exert a causal influence on the development of reading comprehension skills. A large vocabulary provides a richer comprehension experience when reading, which over time improves reading comprehension. The third model proposed that causal relations between vocabulary and reading comprehension are bi-directional. This model suggests that effects described in both models one and two are operating simultaneously. A fourth model proposed that vocabulary and reading comprehension are correlated because they jointly are caused by some third variable. Examples of this model include Anderson and Freebody's (1981) knowledge and aptitude hypotheses, and Nagy's (*in press*) metalinguistic awareness hypothesis.

Each path-analytic model tested in the study included both vocabulary and reading comprehension as variables at two time points. The measure of vocabulary was the

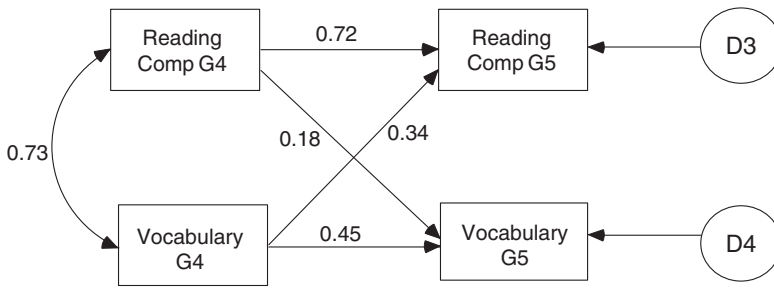


Figure 1. Path analysis of development of vocabulary and reading comprehension from fourth to fifth grade. (Adapted from Wagner, Muse, & Tannenbaum, in press).

Vocabulary subtest from the Stanford-Binet. The measure of reading comprehension was Passage Comprehension from the Woodcock-Johnson. These models also included an autoregressor variable. Results for the time period of fourth — fifth grade are presented in Figure 1. In this figure, the four coefficients of interest are on the arrows in the center of the figure. Each of these path coefficients is significant. The causal influence of fourth-grade reading comprehension on fifth-grade vocabulary is represented by the path coefficient of .18. The causal influence of fourth-grade vocabulary on fifth-grade reading comprehension is represented by the path coefficient of .34. The autoregressive effect for reading comprehension is represented by the path coefficient of .72, and that for vocabulary is represented by the path coefficient of .45. The results for the time periods second—third grade, and third—fourth grade, were highly similar to the results presented in Figure 1.

In summary, the bi-directional model was supported for every time point examined. The case of reciprocal relations between vocabulary development and reading comprehension then mirrors that of relations between phonological awareness and word reading. In both cases, developing oral and written language skills appear to influence each other in a mutually beneficial way.

## 2. LEARNING TO LEARN TO READ

For most of the past century, the “readiness” model has dominated thinking about pre-reading children becoming able to be taught to read (Rayner et al., 2001). Readiness has been conceptualized as a maturational construct. When children arrived in kindergarten or first grade, school psychologists and others administered readiness tests to identify children who were not ready to learn to read. The cure for children who were not deemed ready to read essentially was to wait until they were ready. Delayed entry into formal schooling, or repeating kindergarten or first grade were common strategies to delay formal reading instruction until a child had matured sufficiently. Transitional classes, with demands somewhere between kindergarten and first grade, were sometimes provided as an alternative to simply repeating kindergarten.

The readiness model has now been displaced by an emergent literacy model. Emergent literacy refers to the developmental precursors of conventional reading and writing (Sulzby, 1989; Sulzby & Teale, 1991; Whitehurst & Lonigan, 1998, 2001). In this model, prereaders are believed to be developing important precursor knowledge and abilities that facilitate learning to read. The key constructs in emergent literacy turn out to mirror the literacy-related constructs of phonology, orthography, and morphology discussed in the previous section. At the prereader level, the relevant constructs are phonological sensitivity, print knowledge, and oral language.

## 2.1. Preschool Forms of Phonology, Morphology, and Literacy

*Phonological sensitivity.* Phonological sensitivity refers to an initial form of phonological awareness characterized by being able to access and manipulate larger phonological units but not individual phonemes. Rudimentary acquisition of at least the prosody (tonal patterns and rhythm) of a young child's native language appears to begin prenatally, based on the fact that 2-day-old neonates show preference for their native language (Moon, Cooper, & Fifer, 1993). By 1 month of age, infants demonstrate that they are able to perceive and discriminate speech sounds (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Early experience with their native language tunes the developing infant's phonological system to the speech sounds it encounters. The trade-off for this developing specialization is that 1-year-old children show a decrement in their ability to distinguish non native phonemes (Werker & Tees, 1999).

Phonological sensitivity appears to be a fairly stable individual differences dimension, at least when measured after 3 years of age (Lonigan, Burgess, & Anthony, 2000). What varies developmentally is the linguistic complexity of the phonological segments that a young child is able to access and manipulate. In keeping with the picture described in the previous section, young children move through a developmental progression of being able to demonstrate rudimentary awareness of succeeding smaller segments of speech. These segments, in developmental order, are (1) words in compound words; (2) syllables within words; (3) onset-rime units within syllables; (4) phonemes within rimes; and (5) phonemes within phoneme clusters. Relative success at larger units of linguistic complexity portend relative success at smaller ones later in development (Anthony & Francis, 2005; Anthony et al., 2003)

*Print awareness.* The strongest predictor of subsequent word reading is knowledge of the alphabet at school entry (Adams, 1990; Stevenson & Newman, 1986). In addition to the obvious importance of knowing the names and sounds of letters in learning to read (Bond & Dykstra, 1967; Mason, 1980), letter knowledge appears to play an indirect role in learning to read via facilitating the development of phonological awareness (Burgess & Lonigan, 1998; Wagner, Torgesen, & Rashotte, 1994; Wagner et al., 1997).

Although the importance of letter knowledge is established, that of more rudimentary forms of print knowledge is less well established. For example, young prereaders differ in knowledge about basic concepts about print, including knowing whether a book is



upside down, where the beginning of a book is, and the difference between pictures and print (Clay, 1979). What is not clear is whether this knowledge plays an important role in the emergent stage of reading development, or whether it is merely a by-product of differences in early reading experience that do not, in and of themselves, affect learning to read. Although performance on measures of concepts about print do predict later reading performance, their predictive power is subsumed by the stronger relations between letter knowledge and later reading when both letter knowledge and concepts about print are used as simultaneous predictors (Lonigan et al., 2000).

*Oral language.* A precursor of the ability to communicate with words is gesturing by young infants (Lock, 1978; Namy & Waxman, 1998; Woodward & Hoyne, 1999). Referential gestures signal understanding of intention and symbolic representation that precede true verbal reference. Joint attention between an infant and a caregiver appears to facilitate vocabulary development (Bloom, 2000; Tomasello, 2003). Evidence of understanding word meanings can be found prior to an infant's first birthday, and by the time of school entry, a child will know several thousands of words (Beck & McKeown, 1991).

## **2.2. Effectiveness of Preschool Intervention**

A notable characteristic of prereaders' performance on measures of phonological sensitivity, print knowledge, and oral language is that of tremendous variability across children. The degree of variability may reflect the greater variability of learning environments children are in before they enter the relatively more homogeneous environment that characterizes public schooling. Strong correlations exist between performance in each area—phonological sensitivity, print awareness, and oral language—and measures of poverty or socioeconomic status (Hart & Risley, 1995; Helburn, 1995; Phillips & Lonigan, in press; Smith, Blank, & Collins, 1992). Children living in poverty are more likely to be behind in phonological sensitivity and print knowledge (Bowey, 1995; MacLean, Bryant, & Bradley, 1987; Lonigan, Burgess, Anthony, & Barker, 1998; Raz & Bryant, 1990), and in oral language (Juel, Griffith, & Gough, 1986; Hart & Risley, 1995; Lonigan & Whitehurst, 1998).

Do preschool interventions have any impact on young children's phonological sensitivity, print knowledge, or oral language? Some evidence supports the use of a form of shared reading called dialogic reading (Whitehurst & Lonigan, 1998). Typical storybook reading involves the teacher or adult reading with the child listening passively. In dialogic reading, the child is encouraged to become the storyteller. The adult asks questions, and adds information to increase the child's ability to provide a rich account of the story being 'read.' When compared to an equivalent amount of traditional shared reading, dialogic reading has been shown to produce gains in oral language (Arnold, Lonigan, Whitehurst, & Epstein, 1994; Lonigan, Anthony, Bloomfield, Dyer, & Samwel, 1999; Lonigan & Whitehurst, 1998; Valdez-Menchaca & Whitehurst, 1992; Whitehurst, Arnold, et al., 1994, 1999). When dialogic reading is modified so as to emphasize phonological

sensitivity and print knowledge in addition to meaning, it appears to promote development in these areas compared to a control condition of traditional preschool curriculum (Whitehurst, Epstein et al., 1994; Lonigan, 2003). Other more direct intervention programs also have been shown to accelerate the development of phonological sensitivity (Byrne & Fielding-Barnsley, 1991). However, one should not assume that gains found in research studies will scale-up easily. Large-scale studies involving the NICHD Early Child Care Research Network (2003) provide evidence that preschool teacher and classroom quality moderate gains in cognitive and language scores, and the average quality of preschool education is not high, though it may be improving.

### 3. FAILING TO LEARN TO READ

Reading disability, reading impairment, and developmental dyslexia are essentially interchangeable terms that refer to unexpectedly poor performance in reading (Piasta & Wagner, in press; Wagner, 2005). How expected level of reading performance is quantified can vary. Two common approaches are to compare individuals' reading performances to normative data from their age-matched peers or by comparing their reading performance to their oral language or general cognitive ability. The qualifier 'unexpectedly' refers to not being explained either by lack of an opportunity to learn (i.e., ineffective instruction) or by other potential causes. Poor readers whose impairments are primarily due to sensory, motor, or emotional impairments, mental deficiency, economic or cultural disadvantage, or inadequate reading instruction are excluded from consideration (Lyon, Shaywitz, & Shaywitz, 2003). Implicit in these exclusionary and discrepancy criteria is the assumption that the origin of dyslexia is constitutional, due to neurobiological factors which are intrinsic to the individual (Adams & Bruck, 1993; Bruck, 1990; Gough & Tunmer, 1986; Perfetti, 1985, 1986; Siegel, 2003; Stanovich, 1994).

In many states, formal identification for purposes of becoming eligible for special education services requires a discrepancy between aptitude, as measured by intelligence tests or achievement tests outside the realm of reading (e.g., mathematics), and reading skill. However, current research-based definitions omit a discrepancy requirement, largely on the basis of the fact that the most appropriate intervention for beginning readers who are at risk for reading failure does not vary depending on whether or not the child's reading is discrepant from aptitude (for recent research disputing the need for such an aptitude-achievement discrepancy, see Fletcher, Francis, Rourke, Shaywitz, & Shaywitz, 1992; Fletcher, Morris, & Lyon, 2003; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Lyon et al., 2001, 2003; Shaywitz, Fletcher, Holahan, & Shaywitz, 1992; Shaywitz & Shaywitz, 2003; Siegel, 2003; Stanovich, 1994; Stanovich & Siegel, 1994); and the recent reauthorization of the Individuals with Disabilities Education Act no longer requires such a discrepancy for meeting federal Special Education eligibility for learning disabilities (see Fletcher, Coulter, Reschly, & Vaughn 2004). Developmental dyslexia is distinguished from acquired dyslexia, in that developmental dyslexia refers to a failure to acquire sufficient reading skills, whereas acquired dyslexia refers to impaired reading in formerly normal readers due to brain injury or illness.

### 3.1. Popular Misconceptions of Developmental Dyslexia

Although the term dyslexia is in common use, the key features of developmental dyslexia are misunderstood by most laypersons (Wagner & Muse, in press). The most popular misconception of dyslexia is that it involves seeing mirror images of words or letters. Thus, individuals with dyslexia are reported to read 'WAS' as 'SAW,' or to confuse the letters 'b' and 'd.' Support for this misconception was provided by an early theory of dyslexia by Orton that suggested failure to establish cerebral dominance was responsible for the 'backwards' reading observed in people with dyslexia (Hallahan & Mock, 2003). It is the case that children with developmental dyslexia, who commonly are identified in second or third grade, indeed can be observed to confuse 'WAS' for 'SAW' and 'b' for 'd.' However, these kinds of reversal errors are among the easiest kinds of errors to make. They can be observed everyday in kindergarten and first-grade classrooms among normally developing readers. Reading 'WAS' as 'SAW' rather than as 'WAS' requires that the word is read from left to right, an arbitrary fact of reading English that is not true of other writing systems such as Hebrew. Confusions between letters such as 'b' and 'd' are understandable as well. The letters are visually confusable (to beginning readers, the letter 'b' can be seen as a stick and a ball, whereas the letter 'd' is a ball and a stick) and phonologically (i.e., both are stop consonants). The critical piece of evidence is that second-grade readers with dyslexia make no more reversal errors than do younger normal readers who are matched at the same level of reading (Crowder & Wagner, 1992; Werker, Bryson, & Wassenberg, 1989). What explains the popularity of this mistaken view is that teachers and parents of second-grade readers only see children with reading problems making these errors. Teachers and parents of beginning readers know that such errors are quite common. Overall, the notion of dyslexia as representing a visual-perceptual deficit has not been supported (see reviews in Rayner, 1998; Stanovich, 1982).

A second erroneous yet commonly held view about dyslexia is that it results from deficient eye-movements. Reading requires highly sophisticated and coordinated eye-movements (Crowder & Wagner, 1992; Rayner & Pollatsek, 1989). As you read the words on this page, it seems that you are moving your eyes smoothly across the page. However, this perception is far from reality, as can be observed with a simple experiment. Ask a friend to read directly across from you, holding a book low enough so that you can observe your friend's eyes across the top of the book. What you will see if you look carefully is that your friend's eyes indeed move across the page in a series of small, but observable jerky movements. The jerky movements are called saccades. During saccades, the eyes are moving too fast to see letters or words clearly. Nearly all information is acquired during the fixations or brief pauses between saccades.

If you were to perform the same informal experiment on an individual with reading impairment, you may notice that the individual's eyes move much more erratically. However, erratic eye movements might be a by-product rather than a cause of poor reading. This turns out to be the case. The eye-movements of individuals with dyslexia do not move across the page as smoothly as do those of normal readers simply because they are having trouble reading the words. When normal readers are given material that is as

difficult for them to read as is grade-level reading material for individuals with reading impairment, the eye movements of the normal readers look like those of individuals with dyslexia. Conversely, when individuals with dyslexia are given easy reading material they can read well, their eye movements are indistinguishable from those of normal readers. In addition, interventions based on eye-movement training produce gains in performance on eye-movement tasks but do not produce gains in actual reading (Crowder & Wagner, 1992).

### 3.2. Phonological Bases of Dyslexia

Most individuals with dyslexia have difficulty that is apparent in word-level reading (e.g., Adams, 1990; Snow et al., 1998; Stanovich, 1982; Vellutino, 1979). Although many individuals with dyslexia also are impaired in reading comprehension, the impaired comprehension appears to be more of a by-product of the primary impairment in word-level reading early on (Aaron, 1989; Stanovich & Siegel, 1994). That is, they are unable to decode words in a passage, which correspondingly affects their ability to comprehend the meaning of the passage. However, if left untreated, poor word-level reading is likely to affect comprehension and vocabulary development long-term. Additional support for the view that the primary impairment for most individuals with dyslexia is manifest in their single-word reading is the fact that adults who have compensated for their reading difficulty and no longer are impaired at reading comprehension nevertheless continue to struggle with word recognition (Bruck, 1988, 1990, 1993; Scarborough, 1984).

For the vast majority of individuals with dyslexia, their poor word-level reading begins with a phonological rather than a visual problem, and the initial poor performance may be worsened by ineffective instruction (Spear-Swerling & Sternberg, 1996; Torgesen, 1999; Wagner, 2005; Wagner & Garon, 1999; Wagner & Torgesen, 1987). Individuals with dyslexia perform poorly on measures of phonological awareness, phonological decoding, and recognition of sight words, compared to younger reading-age matched controls (Ehri, 1998; Fox, 1994; Siegel & Faux, 1989).

Perhaps the most obvious impairment common to individuals with dyslexia is their difficulty in decoding pronounceable nonwords or pseudowords (see Rack, Snowling, & Olson, 1992, for a review). Children with dyslexia continue to struggle to read pseudowords even once they have demonstrated knowledge of similar orthographic patterns in real words (Siegel & Faux, 1989). Additional support for a primarily phonological deficit comes from intervention studies that rely on phonological awareness training and phonics instruction produce gains in at-risk or dyslexic readers (e.g., Ball & Blachman, 1991; Brady, Fowler, Stone, & Winbury, 1994; Byrne & Fielding-Barnsley, 1989, 1991, 1993, 1995; Byrne, Fielding-Barnsley, & Ashley, 2000; Ehri, Nunes, Stahl, & Willows, 2001; Ehri, Nunes, Stahl, & Willows, 2001; Foorman et al., 2003; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Foorman, Francis, Novy, & Liberman, 1991; Hatcher, Hulme, & Ellis, 1994; Lovett, Steinbach, & Frijters, 2000; Rashotte, MacPhee, & Torgesen, 2001; Schneider, Ennemoser, Roth, & Kuespert, 1999; Schneider, Roth, & Ennemoser, 2000;

Torgesen, 2005; Torgesen et al., 1999, 2001; Vellutino, Scanlon, & Tanzman, 1998; Wise, Ring, & Olson, 1999; see also Adams, 1990; Bus & van Ijzendoorn, 1999; Chall, 1967/1983; National Reading Panel, 2000; Snow et al., 1998). However, this evidence in support of a phonological deficit as the cause of dyslexia must be qualified by acknowledging the fact that roughly 3% of the poorest readers do not respond to interventions that have been tried to date (Torgesen, 2000, 2002).

### **3.3. Are there Phonological and Orthographic Forms of Developmental Dyslexia?**

The traditional view of word-level reading has been that a reader can get from print to meaning in either of two ways (see Adams, 1990, for a review). The first, which may be termed phonological decoding, involves translating graphemes (in English, these individual letters or letter combinations) into phonemes (sounds) using a set of grapheme phoneme correspondence (GPC) rules. According to the traditional view, phonological decoding is what makes it possible for us to produce a pronunciation when confronted with a word that is new to us, or even a pseudoword. The second way to get from print to meaning is a more direct, orthographic-based approach that results from repeated associations of letter strings and meanings. This traditional model has been replaced by either modified dual-route models or neural net-based models that decode both familiar words and pseudowords using associations between print and pronunciations learned from a corpus of real words (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989).

Research on acquired dyslexia, poor reading due to brain injury or illness, has provided examples of individuals with distinct deficits in either orthographic or phonological processing (Baddeley, Ellis, Miles, & Lewis, 1982; Bryant & Impey, 1986; Coltheart, 1983). Individuals who are unable to read pseudowords but are unimpaired in regular and exception word reading are described as phonological dyslexics. They are impaired in phonological processing, which forces them to rely only on orthographic information to identify words and leaving them unable to phonologically decode words whose print forms are unfamiliar (i.e., pseudowords). Individuals who are able to read pseudowords but are unable to read irregular real words are described as surface dyslexics. They have a primary deficit in orthographic processing and are characterized by strong phonological decoding skills but poor exception word reading. It should be noted that the extreme examples of either phonological or surface dyslexia that have been featured prominently in the literature on acquired dyslexia are the rare exception rather than the rule. Most individuals with acquired dyslexia are properly classified as ‘mixed’ as they exhibit some deficits in both phonological and orthographic processing.

Based on the acquired dyslexia literature, researchers began to ask whether there might be phonological and surface forms of developmental dyslexia (Baddeley et al., 1982; Bryant & Impey, 1986; Coltheart, 1987; Snowling, 1983). These studies have explored two approaches to classify individuals based on their performance at decoding pseudowords and

irregular real words (Stanovich, Siegel, & Gottardo, 1997). One approach is based on hard criteria. This requires normal or better processing in either the phonological or orthographic domain, and subnormal performance in the other domain. The other approach is based on soft criteria. This approach relies on a relative imbalance in performance in the two domains without the requirement of average or better performance in at least one domain.

For example, Castles and Coltheart (1993) compared the pseudoword and irregular real-word reading of students with dyslexia and a chronological age-matched sample of normal readers. Upon analysis of performance in both domains, 15% of dyslexic individuals met the hard classification criteria for the phonological subtype and 19% met the hard criteria for the surface subtype. Using the soft classification criteria, 55% and 30% of the sample met the requirements for phonological and surface subtypes, respectively, and 6% of the sample showed a dual deficit, with impairments in both phonological and orthographic processing.

The Castles and Coltheart (1993) study has been criticized because of its reliance on a chronological-age rather than reading-age matched control group (Manis, Seidenberg, Doi, McBride-Chang, & Peterson, 1996; Stanovich, Seigel, & Gottardo, 1997; Stanovich, Siegel, Gottardo, Chiappe, & Sidhu, 1997). Because relative competence in single-word reading of pseudowords versus exception words may vary with absolute level of reading skill, a reading-age match control group is required (Stanovich, 1988; Stanovich, Nathan, & Zolman, 1988). Given the wide range in both chronological age and reading abilities within Castles and Coltheart's (1993) sample, Stanovich and colleagues (Stanovich, Seigel, Gottardo et al., 1997) reanalyzed the original Castles and Coltheart (1993) data using a reading-age matched control group of normal readers to establish the empirical relations between phonological and orthographic abilities. Upon reanalysis, 38% of the sample of poor readers met the soft phonological subtype criteria, while only two (5.0%) met the soft surface subtype criteria. Stanovich, Seigel, Gottardo et al. (1997) concluded that while the phonological subtype represents a deviant pattern of processing consistent with conceptualizations of reading disability, the surface subtype shows a processing pattern similar to that seen in younger normal readers, suggesting a developmental lag in these reading disabled individuals' acquisition of orthographic skill.

Manis et al., (1996) looked at hard and soft subtypes within a developmental dyslexic group and both a chronological-age matched and a reading-matched control group. Comparison of these disabled readers with their chronologically age matched peers resulted in 10% of the sample meeting the hard classification criteria for each subtype. Using the soft classification scheme, 33% of the dyslexic group was categorized as the phonological subtype, 30% as the surface subtype, and 10% met the criteria for inclusion in both subtypes. However, when compared to the younger, reading-age matched group, no surface subtype individuals remained.

Further findings by Stanovich and colleagues (Stanovich & Siegel, 1994; Stanovich, Seigel, Gottardo et al., 1997) support the notion of individuals who meet the phonological subtype criteria show discrepant performance, whereas individuals who meet the surface

subtype criteria cannot be distinguished from a general delay in development. They performed subtype analyses using the soft criteria with a sample of readers with dyslexia, normal readers matched on chronological age, and younger normal readers matched on reading age. Of this group of impaired readers, 23% met the phonological subtype criteria and 19% met the surface subtype criteria when compared to chronological age-matched peers. Using the reading-age-matched controls, 15% of the dyslexic group were classified as members in the phonological subtype, but none met the criteria for the surface subtype.

Stanovich et al. (1997) extended this line of research on dyslexic subtypes to a younger sample of children with dyslexia (third-grade students with reading scores below the 25th percentile) using the soft classification scheme and both chronological age-matched and reading-age-matched comparisons. Using the chronological age-matched controls, they found that 25%, 22%, and 28% of the dyslexic sample could be classified as members of the phonological subtype, surface subtype, or both subtypes, respectively. Once reading ability was controlled through comparison to the younger normal reader group, 25% of the dyslexic group was identified in the phonological subtype and only one disabled reader (less than 2% of the dyslexic sample) met the surface subtype criteria. Finally, Manis et al. (1999) carried out a 2-year longitudinal study of subtypes of dyslexia. The results again supported a phonological subtype but not a surface subtype of developmental dyslexia.

In summary, although both phonological and surface subtypes may be discriminated within the population of poor readers, only the phonological subtype is congruent with the characterization of dyslexia as an unexpected, specific impairment in word reading processes. Once reading level is controlled, the surface subtype virtually disappears, consistent with the notion that orthographic deficits are best conceptualized as the results of developmental lag.

#### 4. CONCLUSIONS

In conclusion, we examined three aspects of learning to read: learning to read, learning to learn to read, and failing to learn to read. Beginning with learning to read, three constructs important in understanding learning to read were identified: phonology, morphology, and orthography. Learning to read is best characterized by a mutually beneficial interaction between oral language skills (i.e., phonology and morphology) and written language skills (i.e., orthography). Reciprocal causal relations exist between the development of phonological awareness and word-level reading. Similar reciprocal causal relations exist between the development of vocabulary and reading comprehension.

Turning to the case of learning to learn to read, a readiness model has been replaced by an emergent literacy model that features prereading versions of the same three key constructs: phonological sensitivity, oral language, and print knowledge. A notable characteristic of prereaders' performance on measures of phonological sensitivity, print knowledge, and oral language is tremendous variability across children, which probably

reflects the greater variability in literacy environments before formal schooling begins. Some support exists for intervention at the preschool level for remediating deficits in phonological sensitivity, print knowledge, and oral language.

For the case of failing to learn to read, an impairment in phonological processing appears to be the basis of most cases of developmental dyslexia. Despite considerable interest, well-controlled studies do not support the existence of a second subtype of dyslexia based on impaired orthographic processing.

## REFERENCES

- Aaron, P. G. (1989). Qualitative and quantitative differences among dyslexic, normal, and nondyslexic poor readers. *Reading and Writing, 1*, 291–308.
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Adams, M. J., & Bruck, M. (1993). Word recognition: The interface of educational policies and scientific research. *Reading and Writing, 5*, 113–139.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In: J. T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews*. Newark, DE: International Reading Association.
- Anthony, J., & Francis, D. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*, 255–259.
- Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Bloomfield, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82*, 65–92.
- Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*, 470–487.
- Arnbak, E., & Elbro, C. (2000). The effects of morphological awareness training on the reading and spelling skills of young dyslexics. *Scandinavian Journal of Educational Research, 44*, 229–251.
- Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: Replication and extension to a videotape training format. *Journal of Educational Psychology, 86*, 235–243.
- Baddeley, A. (1982). Reading and working memory. *Bulletin of the British Psychological Society, 35*, 414–417.
- Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.



- Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition*, *11*, 185–199.
- Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, *26*, 49–66.
- Beck, I. L., & McKeown, M. G. (1991). Conditions of vocabulary acquisition. In: R. E. Barr, M. L. E. Kamil, P. Mosenthal, & P. Pearson (Eds), *Handbook of Reading Research*, (Vol. 2, pp. 789–814). Hillsdale, NJ: Lawrence Erlbaum.
- Blachman, B., Ball, E., Black, S., & Tangel, D. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms: Does it make a difference? *Reading and Writing: An Interdisciplinary Journal*, *6*, 1–17.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Bloomfield, L. (1933). *Language*. Oxford, England: Holt.
- Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, *2*, 5–142.
- Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, *87*, 476–487.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, *301*, 419–421.
- Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. *Annals of Dyslexia*, *44*, 26–59.
- Bruck, M. (1988). The word recognition and spelling of dyslexic children. *Reading Research Quarterly*, *23*, 51–69.
- Bruck, M. (1990). Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology*, *26*, 439–454.
- Bruck, M. (1993). Word recognition and component phonological processing skills of adults with childhood diagnosis of dyslexia. *Developmental Review*, *13*, 258–268.
- Bryant, P. E., & Impey, L. (1986). The similarities between normal readers and developmental and acquired dyslexics. *Cognition*, *24*, 121–137.
- Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and pre-reading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology*, *70*, 117–141.
- Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, *91*, 403–414.

- Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology, 81*, 313–321.
- Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83*, 451–455.
- Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*, 104–111.
- Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*, 488–503.
- Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92*, 659–667.
- Carlisle, J. F. (in press). Fostering morphological processing, vocabulary development, and reading comprehension. In: R.K. Wagner, A. Muse, & K. Tannenbaum (Eds), *Vocabulary development and its implications for reading comprehension*. New York: Guilford Press.
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149–180.
- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77–111.
- Chall, J. S. (1967/1983). *Learning to read: The great debate*. New York: McGraw-Hill
- Clay, M. M. (1979). *Reading: The patterning of complex behavior*. Auckland, New Zealand: Heinemann.
- Coltheart, M. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 35*, 469–495.
- Coltheart, M. (1987). Varieties of developmental dyslexia: A comment on Bryant and Impey. *Cognition, 27*, 97–101.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589–608.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.
- Crowder, R. G., & Wagner, R. K. (1992). *The psychology of reading: An introduction* (2nd ed.). London: Oxford University Press.
- Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology, 50*, 429–444.

- DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu, HI: University of Hawaii Press.
- de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, 6, 51–77.
- Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential to learning to read words in English. In: J. L. Metsala, & L. C. Ehri (Eds), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Lawrence Erlbaum.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001a). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71, 393–447.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghouh-Zadeh, Z., & Shanahan, T. (2001b). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36, 250–287.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303–306
- Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*, 54, 304–331.
- Fletcher, J. M., Francis, D. J., Rourke, B. P., Shaywitz, S. E., & Shaywitz, B. A. (1992). The validity of discrepancy-based definitions of reading disabilities. *Journal of Learning Disabilities*, 25, 573.
- Fletcher, J. M., Morris, R. D., & Lyon, G. R. (2003). Classification and definition of learning disabilities: An integrative perspective. In: H. L. Swanson, & K. R. Harris (Eds), *Handbook of learning disabilities* (pp. 30–56). New York: Guilford Press.
- Foorman, B. R., Chen, D. -T., Carlson, C., Moats, L., Francis, D. J., & Fletcher, J. M. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing: An Interdisciplinary Journal*, 16, 289–324.
- Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, 90, 37–55.
- Foorman, B. R., Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology*, 83, 456–469.
- Fox, E. (1994). Grapheme-phoneme correspondence in dyslexic and matched control readers. *British Journal of Psychology*, 85, 41–53.
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 3–17.

- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 104–115.
- Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., Braun, M., & O'Connor, R. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*, 251–267.
- Gelb, I. J. (1952). *A study of writing*. Chicago: University of Chicago Press.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10.
- Hallahan, D. P., & Mock, D. R. (2003). A brief history of the field of learning disabilities. In H. L. Swanson & K. R. Harris (Eds), *Handbook of learning disabilities* (pp. 16–29). New York: Guilford Press.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.
- Hatcher, P. J., & Hulme, C. (1999). Phonemes, rhymes, and intelligence as predictors of children's responsiveness to remedial reading instruction: Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *72*, 130–153.
- Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, *65*, 41–57.
- Helburn, S. W. (Ed.) (1995). Cost, quality, and child outcomes in child care centers (Tech. Rep.). Denver, CO: University of Colorado at Denver, Department of Economics, Center for Research in Economic and Social Policy.
- Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading*, *9*, 351–365.
- Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics*, *4*, 103–147.
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, *78*, 243–255.
- Lindamood, C. H., & Lindamood, P. C. (1984). *Auditory discrimination in depth*. Austin, TX: PRO-ED.
- Lock, A. (1978). The emergence of language. In: A. Lock (Ed.), *Action, gesture and symbol*. New York: Academic Press.

- Lonigan, C. J. (2003). Development and promotion of emergent literacy skills in preschool children at-risk of reading difficulties. In: B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale*. Timonium, MD: York Press.
- Lonigan, C. J., & Whitehurst, G. J. (1998). Examination of the relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, 17, 265–292.
- Lonigan, C. J., Anthony, J. L., Bloomfield, B., Dyer, S. M., & Samwel, C. (1999). Effects of two preschool shared reading interventions on the emergent literacy skills of children from low-income families. *Journal of Early Intervention*, 22, 306–322.
- Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology*, 36, 596–613.
- Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in two- to five-year-old children. *Journal of Educational Psychology*, 90, 294–311.
- Lovett, M. W., Steinbach, K. A., & Frijters, J. C. (2000). Remediating the core deficits of developmental reading disability: A double-deficit perspective. *Journal of Learning Disabilities*, 33, 334–358.
- Lundberg, I., Frost, J., & Petersen, O. -P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, 23, 263–284.
- Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., et al. (2001). Rethinking learning disabilities. In: C. E. Finn, A. J. Rotherham, & C. R. Hokanson Jr. (Eds), *Rethinking special education for a new century* (pp. 259–287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.
- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53, 1–14.
- MacLean, M., Bryant, P., & Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly*, 33, 255–282.
- Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Peterson, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition*, 58, 157–195.
- Manis, F. R., Seidenberg, M. S., Stallings, L., Joanisse, M., Bailey, C., Freedman, L., et al. (1999). Development of dyslexic subgroups: A one-year follow up. *Annals of Dyslexia*, 49, 105–134.
- Mason, J. M. (1980). When children do begin to read: An exploration of four year old children's letter and word reading competencies. *Reading Research Quarterly*, 15, 203–227.
- McBride-Chang, C., Cho, J. R., Lie, H., Wagner, R. K., Shu, H., Zhou, A., Cheuk, C., & Muse, A. (in press). Changing models across cultures: Associations of phonological awareness and morphological structure awareness to vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and America. *Journal of Experimental Child Psychology*.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*, 495–500.

Morais, J., Cary, L., Alegria, J. S., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323–331.

Nagy, W. (in press). Metalinguistic awareness and the vocabulary-comprehension connection. In: R. K. Wagner, A. Muse, & K. Tannenbaum (Eds), *Vocabulary development and its implications for reading comprehension*. New York: Guilford Press.

Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology, 95*, 730–742.

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development, 69*, 295–308.

National Institute of Child Health and Human Development. (2002). *Early childhood education and school readiness: Conceptual models, constructs, and measures*. Washington, DC: Author.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 ½? *Developmental Psychology, 39*, 451–469.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read* (No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.

Parilla, R., Aunola, K., Leskinen, E., Nurmi, J. E., Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97*, 299–319.

Parilla, R., Kirby, J. R., McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8*, 3–26.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A. (1986). Continuities in reading acquisition, reading skill, and reading disability. *RASE: Remedial and Special Education, 7*, 11–21.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1988). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. In: K. Stanovich (Ed.), *Children's reading and the development of phonological awareness*. Detroit, MI: Wayne State University Press.

Phillips, B. M., & Lonigan, C. J. (in press). *Social correlates of emergent literacy*. To appear In: C. Hulme, & M. Snowling (Eds), *The science of reading: A handbook*, Blackwell.

Piasta, S. B., & Wagner, R. K. (in press). Dyslexia: Identification and Classification. In: E. Grigorenko (Ed.), *Word level reading*. Mahwah, NJ: Erlbaum.

- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.
- Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, *27*, 28–53.
- Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly*, *24*, 119–134.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Upper Saddle River, NJ: Prentice-Hall.
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*, 31–74.
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2002). How should reading be taught? *Scientific American*, *286*, 84–91.
- Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology*, *8*, 209–225.
- Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, *24*, 31–44.
- Scarborough, H. S. (1984). Continuity between childhood dyslexia and adult reading. *British Journal of Psychology*, *75*, 329–348.
- Schneider, W., Ennemoser, M., Roth, E., & Kuespert, P. (1999). Kindergarten prevention of dyslexia: Does training in phonological awareness work for everybody? *Journal of Learning Disabilities*, *32*, 429–436.
- Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*, *92*, 284–295.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.
- Shaywitz, B. A., Fletcher, J. M., Holahan, J. M., & Shaywitz, S. E. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut longitudinal study. *Journal of Learning Disabilities*, *25*, 639–648.
- Shaywitz, S. E., & Shaywitz, B. A. (2003). Neurobiological indices of dyslexia. In: H. L. Swanson, & K. R. Harris (Eds), *Handbook of learning disabilities* (pp. 514–531). New York: Guilford Press.

Siegel, L. S. (2003). Basic cognitive processes and reading disabilities. In: H. L. Swanson, & K. R. Harris (Eds), *Handbook of learning disabilities* (pp. 158–181). New York: Guilford Press.

Siegel, L. S., & Faux, D. (1989). Acquisition of certain grapheme-phoneme correspondences in normally achieving and disabled readers. *Reading and Writing, 1*, 37–52.

Smith, S., Blank, S., & Collins, R. (1992). *Pathways to self-sufficiency for two generations: Designing welfare-to-work programs that benefit children and strengthen families*. New York: New York Foundation for Child Development.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snowling, M. J. (1983). The comparison of acquired and developmental disorders of reading: A discussion. *Cognition, 14*, 105–118.

Spear-Swerling, L., & Sternberg, R. J. (1996). *Off-track: When poor readers become learning disabled*. Boulder, CO: Westview Press.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: I. Word decoding. *Journal of Learning Disabilities, 15*, 485–493.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*, 590–604, 612.

Stanovich, K. E. (1994). Does dyslexia exist? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 35*, 579–595.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24–53.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1988). The developmental lag hypothesis in reading: Longitudinal and matched reading-level comparisons. *Child Development, 59*, 71–86.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89*, 114–127.

Stanovich, K. E., Siegel, L. S., Gottardo, A., Chiappe, P., & Sidhu, R. (1997). Subtypes of developmental dyslexia: Differences in phonological and orthographic coding. In: B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 115–141). Mahwah, NJ: Lawrence Erlbaum.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57*, 646–659.

Sulzby, E. (1989). Assessment of writing and of children's language while writing. In: L. Morrow, & J. Smith (Eds), *The role of assessment and measurement in early literacy instruction* (pp. 83–109). Englewood Cliffs, NJ: Prentice-Hall.



Sulzby, E., & Teale, W. (1991). Emergent literacy. In: R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds), *Handbook of Reading Research* (Vol. II, pp. 727–758). New York: Longman.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Torgesen, J. K. (1999). Phonologically based reading disabilities: Toward a coherent theory of one kind of learning disability. In: R.J. Sternberg, & L. Spear-Swerling (Eds), *Perspectives on learning disabilities*. (pp. 231–262). New Haven: Westview Press.

Torgesen, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, 15, 55–64.

Torgesen, J. K. (2002). Lessons learned from intervention research in reading: A way to go before we rest. In: R. Stainthorpe (Ed.), *Literacy: Learning and teaching* (pp. 89–104). Monograph of the British Journal of Educational Psychology.

Torgesen, J. K. (2005). Recent discoveries from research on remedial interventions for children with dyslexia. In: M. Snowling, & C. Hulme (Eds), *The science of reading* (pp. 521–537). Oxford: Blackwell.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, 34, 33–58.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, 91, 579–593.

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development through picture book reading: A systematic extension to Mexican day-care. *Developmental Psychology*, 28, 1106–1114.

Vellutino, F. R. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology*, 36, 367–397.

Wagner, R. K. (2005). Reading impairment. In: P. Strazny (Ed.), *Encyclopedia of Psycholinguistics*. New York: Routledge, Taylor, and Francis.

Wagner, R. K., & Garon, T. (1999). Learning disabilities in perspective. In: R. J. Sternberg & L. Spear-Swerling (Eds), *Perspectives on learning disabilities*. Boulder, CO: Westview.

Wagner, R. K., & Muse, A. E. (in press). Working memory deficits in developmental dyslexia. In: T. Alloway, & S. Gathercole, (Eds), *Working memory in neurodevelopmental conditions*. London: Psychology Press.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192–212.

Wagner, R. K., Muse, A. E., & Tannenbaum, K. R. (in press). Promising avenues for better understanding implications of vocabulary development for reading comprehension. In: R.K. Wagner, A. Muse, & K. Tannenbaum (Eds), *Vocabulary development and its implications for reading comprehension*. New York: Guilford Press.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., Rashotte, C. A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology*, *85*, 1–20.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, *30*, 73–87.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processes*. Austin, TX: PRO-ED Publishing Inc.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*, 468–479.

Werker, J. F., Bryson, S. E., & Wassenberg, K. (1989). Toward understanding the problem in severely disabled readers: II. Consonant errors. *Applied Psycholinguistics*, *10*, 13–30.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *68*, 848–872.

Whitehurst, G. J., & Lonigan, C. J. (2001). Emergent literacy: Development from prereaders to readers. In: S. B. Neuman, & D. K. Dickenson (Eds), *Handbook of Early Literacy Research* (pp. 11–29). New York: Guilford Press.

Whitehurst, G. J., Arnold, D. H., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in daycare and home for children from low-income families. *Developmental Psychology*, *30*, 679–689.

Whitehurst, G. J., Epstein, J. N., Angell, A. C., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology*, *86*, 542–555.

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology*, *91*, 261–272.

Wise, B. W., Ring, J., & Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology*, *72*, 271–304.

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, *70*, 65–77.

Ziegler, J. C., & Goswami, U. C. (in press). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*.

## Chapter 30

### **Cognitive and Linguistic Issues in the Study of Children with Specific Language Impairment**

*Laurence B. Leonard and Patricia Deevy*

Children with specific language impairment (SLI) exhibit a significant deficit in spoken language ability, yet show none of the symptoms associated with other types of developmental disorders. These children earn scores on non-verbal tests of intelligence that fall within the normal range for their age, they pass screening tests for hearing acuity, and they present no signs of frank neurological impairment. Furthermore, their social interactions with others seem rather typical, apart from their difficulty with language itself. This unusual developmental profile has attracted a great deal of attention from scholars interested in examining the nature of language acquisition. In this chapter, we discuss some of the issues surrounding SLI. We present some of the hallmark language symptoms of this disorder and review the types of theories that have been advanced to account for them. We then propose that a complete understanding of SLI will probably require us to soften the boundaries between language knowledge on the one hand, and language processing on the other. We begin with a brief overview of some of the basic facts about this disorder.

#### **1. SLI: AN OVERVIEW**

The term “specific language impairment” has been in use for only 25 years or so, but the disorder itself has been known since the first half of the 19th century (see reviews in Weiner, 1986, Johnston, 1988; Leonard, 1998). The labels have changed over the years, in part because particular terms placed emphasis on factors such as speech perception (e.g., “congenital auditory imprecision”) or neurological impairment (e.g., “congenital aphasia”) that proved to be playing only a minimal (or no) role in the disorder. The term “delayed language” has also been employed, though this term can give the misleading impression that all children with such a label will “outgrow” their problem without intervention. As will be seen below, the term “specific language impairment” is not without its limitations, as the disorder is not as specific to language as the term implies. However, it does serve to distinguish this kind of deficit from other types of developmental disorders that may also involve problems with language learning.

The prevalence of SLI is approximately 7% according to epidemiological studies (Tomblin et al., 1997). Earlier prevalence estimates had placed the figure at 3 to 5% (American Psychiatric Association, 1994), though these estimates were based on clinically referred samples, that is, on children who actually appeared at clinical facilities for assessment and intervention. Males outnumber females, by approximately 2.5 to 1 in clinically referred studies and 1.5 to 1 in epidemiological studies. For many children, SLI appears to be genetic. Children with SLI are two or three times more likely than typical children to have siblings with language problems or parents with a history of language problems (e.g., Tomblin et al., 2001). Twin studies show that the concordance rates for children with SLI and their identical (monozygotic) twins are substantially higher than those seen for children with SLI and their same-sex dizygotic twins (Bishop, North, & Donlan, 1995). Although the evidence for a genetic as opposed to primarily environmental basis for SLI is rather convincing, the prominent areas of linkage uncovered thus far by molecular genetic studies suggest that much more research is still needed before a clear picture emerges (Bishop, 2002). Indeed, the cause of SLI may well prove to be multifactorial.

Seemingly universal characteristics of SLI include the late appearance of first words, a protracted period of lexical acquisition from the first word until the first 50 or so words, and the late appearance of word combinations. Often, these children's comprehension of language is more advanced than their language production abilities; however, comprehension problems are also seen and in some children can be rather severe.

Most children with SLI whose language problems are moderate to severe receive some type of treatment during the preschool years. Intervention studies show that the children's gains in treatment exceed those that are seen through maturation alone (Leonard, 1998). Treatment gains are often hard won; dramatic changes in a child's language abilities across a short time frame are the exception rather than the rule. In fact, when children are randomly assigned to treatment and control groups, the evidence concerning the effectiveness of intervention for expressive syntax is rather mixed, though it is somewhat stronger for children with better comprehension (Law, Garrett, & Nye, 2004). Evidence for the effectiveness of intervention for vocabulary is considerably stronger.

For children showing symptoms of SLI at 5 years of age, there is substantial likelihood that these children will perform significantly below the level of same-age peers on language tasks in later years (e.g., Tallal, Curtiss, & Kaplan, 1988; Bishop & Adams, 1990; Beitchman, Wilson, Brownlie, Walters, & Lancee, 1996). Even by adulthood, individuals with a history of SLI perform below the level of other adults of the same age on measures of spoken language ability (Tomblin, Freese, & Records, 1992).

There is a great deal of evidence indicating that when children with SLI reach school age, they frequently experience reading difficulties (e.g., Haynes & Naidoo, 1991; Snowling, Bishop, & Stothard, 2000; Conti-Ramsden, Botting, Simkin, & Knox, 2001). Furthermore, the nature of the reading difficulty does not seem to be confined to one particular area of reading. For example, Catts, Fey, Tomblin, and Zhang (2002) found that children with persisting SLI exhibited difficulties not only in reading comprehension but also in decoding print.

Given findings such as those above, it is natural to ask whether SLI and developmental dyslexia are overlapping conditions, or even if the same children would receive one or the other label depending on their age and enrollment in an academic setting. Such a possibility is reinforced by prospective studies of children born into families with a history of dyslexia. In these studies, such children at risk for reading problems were found to have significantly weaker oral language skills during the preschool years than children not at risk for reading difficulties (e.g., Scarborough, 1990, 1991). However, the fact that the SLI and developmental dyslexia populations have limitations in both oral language and reading abilities should not be taken to mean that the key symptoms and causes of the impairments are necessarily the same in the two groups. Bishop and Snowling (2004) have identified some important (though easily overlooked) differences. There are many children with dyslexia whose principal difficulties rest in the phonologically related aspects of language, as reflected in poor scores on phonological awareness measures, for example. Many children with SLI, on the other hand, have weaknesses in both phonologically related and non-phonologically related (morphosyntactic, semantic) areas of language. Both types of weaknesses can cause difficulties in reading. A third pattern of weakness is a limitation in non-phonologically related areas of language without a corresponding problem in phonologically related areas. These children show poor reading comprehension. More complete testing of these children often reveals oral language deficiencies along with reading problems. However, perhaps because these children do not exhibit significant problems in areas related to phonology, they often go undiagnosed until the reading difficulties are detected.

## 2. WHY STUDY SLI?

Much research on SLI has a strictly clinical focus, devoted to the discovery of factors that may significantly increase the accuracy of identifying children with SLI or significantly increase the effectiveness of treatment. This type of research is greatly needed and by itself provides justification for the study of this disorder. However, there are other important reasons to study SLI. For example, the study of SLI can prove helpful because this disorder serves as a type of baseline against which the prevalence of language impairment in other populations can be compared. For example, it may not be accurate to assume that all (otherwise typical) deaf children exposed to a natural sign language from birth will acquire sign language in an efficient manner. Judging from the existence of SLI, some small percentage of deaf children might also experience difficulties learning language even when it is a language of signs.

SLI can also serve as an important testing ground for current notions of typical language development. Most obviously, the age-appropriate non-verbal IQ scores of children with SLI suggest that there is some degree of autonomy between language and other cognitive abilities. Indeed, some studies have reported rather serious language impairments in spite of seemingly intact abilities in non-linguistic areas (e.g., van der Lely, 1997).

However, there is also ample evidence indicating that many children with SLI have limitations in processing speed and capacity. These limitations are seen in non-linguistic as well as linguistic processing tasks (a point we shall return to later). Researchers have used

such evidence to argue that limitations in cognitive resources or reduced speed of processing may constitute part of the problem that these children have in acquiring language. Importantly, such a position requires the assumption that these processing factors—including those that affect performance on non-linguistic tasks—are important for language. If so, it would seem important to determine whether current models of language learning in typical children can accommodate them.

SLI can also provide information regarding the relative ease with which particular details of language can be acquired when the children's development does not conform to the usual timetable. For example, although lexical learning continues throughout life, it is not clear if any non-linear patterns in lexical acquisition (e.g., a vocabulary spurt), are driven primarily by age, the number of words already acquired, or some combination of these and other factors. Because children with SLI are older when they accumulate the same number of words ordinarily associated with a break from linear growth, this type of question can be pursued.

The study of SLI can also be useful to the extent that this disorder constitutes language development "in slow motion." Some language attainments usually occur so rapidly in typically developing children that achieving a complete understanding of their nature might be difficult without the benefit of studying less typical learning rates such as that of SLI.

A greater understanding of crosslinguistic differences in language development and processing might also arise from the study of children with SLI. Comparisons across languages are valuable because they permit researchers to separate factors (e.g., frequency of occurrence, transparency of function, regularity of morphological paradigm) that may be confounded in any one language. Children with SLI appear to show profiles that represent exaggerated versions of the profile that is most characteristic of the language they are acquiring (Leonard, 2000). For example, English-speaking children with SLI have unusually telegraphic sentences, and Swedish-speaking children with SLI have special difficulties mastering the verb-second characteristic of the grammar. In cases where studies of typical development reveal only subtle crosslinguistic differences, children with SLI might serve as an important addition. The crosslinguistic comparisons among the children with SLI might indicate whether the differences are real, and if so, which characteristics of the languages are most responsible for them.

Because children with SLI often participate in language intervention programs, their progress in such activities might provide us with a better understanding of the potential contribution of the language environment to children's language learning. Children with SLI do not appear to be raised in atypical language environments. For example, studies of their linguistic input reveal few differences from the input provided to typically developing children. Therefore, it is clear that much of the language difficulty facing these children can be traced to factors inherent in the children (which may, of course, have been inherited, as noted above). Treatment can be characterized as providing children with an optimal language environment, in which problematic details of language are presented frequently, and in a manner that makes their function and form as transparent as possible.

Although the environmental manipulations employed in treatment are clearly not necessary for typically developing children, they may provide insight into the kind of cues that all children benefit from, even (in the case of typically developing children) when they are not emphasized in the input.

### 3. MORPHOSYNTACTIC VARIABILITY AS A PHENOTYPE OF LANGUAGE IMPAIRMENT

Major advances toward explaining SLI will come from a clearer understanding of its symptoms. By understanding the behavioral manifestation of SLI—its phenotype—researchers will be in a better position to identify children with SLI for purposes of understanding the genetic bases of the disorder. Furthermore, with a clear picture of the symptoms of SLI, investigators can propose and test hypotheses about this disorder with increasing precision.

Unfortunately, the goal of specifying the phenotype is complicated by the fact that children with SLI appear to constitute a heterogeneous population, at least as the disorder is currently defined. One complication, as described below, is that there are crosslinguistic differences among children with SLI. In addition, there are individual differences within each language. However, there are common profiles. The profile that appears most frequently in English-speaking children with SLI is a mild to moderate deficit in semantic and phonological areas and a more serious deficit in the area of morphosyntax. Within the area of morphosyntax, the expression of tense and agreement seems to be especially affected. Children with SLI have difficulties in their use of grammatical morphemes such as past tense *-ed* (e.g., *jumped*), present third person singular *-s* (e.g., *jumps*), and both copula and auxiliary forms of *is*, *are*, *am*, *was*, and *were* (as in *The bus is here*, *The bus is coming*). Measures based on children's use of these morphemes show good sensitivity and specificity in distinguishing children with SLI from typically developing same-age peers (e.g., Bedore & Leonard, 1998). Here, "sensitivity" refers to the degree to which children who are truly language impaired (as determined by independent means) are in fact identified as language impaired by the tense/agreement morpheme measure. "Specificity" refers to the degree to which children who are truly typically developing are identified as not language impaired by this measure.

The search for measures with potentially high sensitivity and specificity often begins with comparisons between children with SLI and two comparison groups of typically developing children. The first comparison group is the familiar age-matched group. The second comparison group is a group of younger typically developing children. These children are matched with the SLI group on a measure that might assist the interpretation of any group differences that are seen on the dependent measure. For example, a common basis for matching is the children's mean length of utterance (MLU). During the toddler and preschool years, children's MLUs increase as they become more proficient in the use of language. Children with SLI usually produce utterances with shorter MLUs than are expected for their chronological age. Of course, utterances containing tense/agreement morphemes are longer than those that lack them (compare *Chris is running home* with



*Chris running home*). Therefore, if children with SLI are found to make less use of these morphemes than typically developing children of the same age, it is not clear if the difference is due to problems with tense/agreement morphemes, or to an inability to produce utterances of the length necessary to accommodate these morphemes. If children with SLI are also found to make less use of tense/agreement morphemes than younger children matched according to MLU, sentence length is not likely to be the reason.

MLU matching has another value, in addition to removing a potentially confounding variable. During the preschool years, typically developing children selected as MLU matches are approximately two years younger than the SLI group (e.g., Rice, Wexler, & Cleave, 1995). Therefore, any difference favoring the MLU group is likely to constitute an area of serious weakness on the part of children with SLI. Given that numerous studies have found group differences of this type for tense/agreement morphemes, this area of morphosyntax certainly qualifies as one area of concern. Examples of studies reporting such differences include Bliss (1989), Leonard, Bortolini, Caselli, McGregor, and Sabbadini (1992), Rice et al. (1995), Cleave and Rice (1997), Leonard, Eyer, Bedore, and Grela (1997), Oetting and Horohov (1997) among others.

An accurate portrayal of this potential phenotype requires a closer inspection of the tense/agreement data. Children with SLI may be quite limited in their use of tense/agreement morphemes, but at the same time, they provide evidence of understanding several important details about how these morphemes work. For example, these children only rarely produce these morphemes in inappropriate contexts; productions such as *They jumps* are seldom seen. By far the greatest proportion of difficulty involves the children's failure to supply the morpheme where it should appear in the adult grammar, as in *Mommy drive to work everyday, Daddy play basketball yesterday, and Jack tall but Josh not tall* (see Bishop, 1994).

Of course, children could restrict their use of tense/agreement morphology to appropriate contexts without understanding a great deal about these morphemes. For example, children might learn the inflected forms of specific verbs and the contexts in which these should be used without having learned the inflected forms for other verbs. As a result, the children might use, say, *runs* and *likes* consistently in appropriate contexts but never use *walks* and *wants* when these inflected forms are required. Yet the data do not bear this out. Miller and Leonard (1998) found that children with SLI showed a great deal of variability with the same verbs; thus, the same child might say *likes* in one instance and *like* in another.

Even if children with SLI were variable in using tense/agreement morphemes with the same verb, there is no assurance that they could apply their knowledge of tense/agreement beyond highly attested examples in the input. However, evidence of such ability is available, taking the form of over-regularizations. For example, in spite of their limited use of past tense *-ed* in obligatory contexts, children with SLI also produce forms such as *throwed* and *runned* (e.g., Bishop, 1994; Leonard et al., 1997; Oetting & Horohov, 1997; Marchman, Wulfeck, & Ellis Weismer, 1999; Redmond & Rice, 2001). Because

such productions do not occur in the input, it would appear that children's use of tense/agreement morphology is not limited to highly practiced forms. As in the literature on normal language development, investigators of SLI disagree on the mechanisms of learning or acquisition that are responsible for over-regularizations. Some researchers assume that over-regularizations occur when the child fails to retrieve a previously learned irregular past form of a verb and a productive (*-ed*) rule is then applied to that verb (van der Lely & Ullman, 2001). Others assume that all past forms are learned through a single mechanism and over-regularizations are cases where the strength of the *-ed* form is greater than its irregular past counterpart for that particular verb stem at the time of production (Marchman et al., 1999). Despite these differences, all agree on the phenomenon itself; children with SLI sometimes produce past tense forms that go beyond the information available in the ambient language.

It would appear that the most salient detail about these children's difficulty with tense and agreement morphemes is their inconsistency in producing them. Table 1 provides a summary across several studies involving children with SLI during the preschool years. As can be seen from the table, the mean percentages of use for each morpheme are well above 0% for the SLI groups. Furthermore, although there is variability within each SLI group, it would appear that most if not all children used each morpheme at least to a limited degree. Any account of the tense/agreement difficulties of children with SLI must be capable of explaining how these children know enough about these morphemes to use them creatively and avoid their misapplication while at the same time produce them inconsistently.

Table 1

Mean percentages of use (with standard deviations) of tense/agreement morpheme use by children with SLI, younger typically developing children matched according to MLU (TD-MLU) and typically developing children matched according to age (TD-A).

	SLI	TD-MLU	TD-A
Past tense <i>-ed</i>			
Leonard et al. (1992)	32 (27)	65 (23)	98 (6)
Rice et al. (1995)	18 (20)	56 (32)	90 (20)
Present third person singular <i>-s</i>			
Leonard et al. (1992)	34 (20)	59 (30)	91 (7)
Rice et al. (1995)	34 (26)	51 (40)	85 (20)
Leonard et al. (1997)	34 (24)	51 (30)	98 (4)
Copula <i>be</i> forms			
Leonard et al. (1992)	41 (32)	71 (21)	97 (6)
Rice et al. (1995) <sup>a</sup>	46 (24)	68 (24)	97 (5)
Leonard et al. (1997)	64 (21)	80 (16)	96 (4)

<sup>a</sup>Copula *be* and auxiliary *be* forms were combined in the Rice et al. (1995) study.

#### 4. EXPLAINING VARIABILITY IN TENSE/AGREEMENT USE: TWO APPROACHES

The variability in tense/agreement use by children with SLI has been studied from two different perspectives. One perspective relies heavily on linguistic theory, especially the linguistic framework of Chomsky (1995). Accounts of tense/agreement difficulty of this type assume that children with SLI lack some grammatical insight that other children acquire quite readily. The other perspective relies on an assumption that children with SLI have processing limitations. Accounts of this type share the assumption that the children have the potential to grasp all relevant details of language, but because of limitations in processing, they are inconsistent in storing, retrieving, or applying them. Here, the type of information is less essential than how it is mentally manipulated.

#### 5. PROPOSALS OF AN EXTENDED OPTIONAL PERIOD

Several approaches employing linguistic frameworks have been advanced to explain the tense/agreement morpheme weaknesses seen in SLI (e.g., Clahsen, 1989; Gopnik & Crago, 1991; van der Lely, 1997, 1998). We focus here on one type of account that has received considerable attention in the literature on SLI. Rice, Wexler, and their colleagues (e.g., Rice et al., 1995; Rice & Wexler, 1996; Rice, Wexler, & Hershberger, 1998) proposed that all children learning languages such as English proceed through a period of grammatical development during which they treat tense morphemes as optional in main clauses. According to this view, the children understand the function of tense, and where it is to be applied; however, they fail to grasp that it must be applied in all cases where tense is used in the adult grammar. When tense is not selected in such contexts, the children employ instead a non-finite form. Thus, a production such as *Chris play basketball yesterday* is comparable to the non-finite phrase in the adult sentence *We saw Chris play basketball yesterday*. In contexts requiring a copula or auxiliary *be* form in the adult grammar, children in this optional stage produce the utterance without the copula or auxiliary form, as in *Gemma eating ice cream*. In the adult grammar, too, non-finite clauses lack such tense and agreement morphemes (e.g., *I saw Gemma eating ice cream*).

The period of optional use is a natural phenomenon, seen in typically developing children and children with SLI alike. However, typically developing children proceed through this optional period rather quickly. Children with SLI are assumed to have a protracted stay in the optional period. For this reason, the proposal as it applies to children with SLI is referred to as the extended optional infinitive (EOI) account. This protracted stay in an optional period is assumed to reflect a maturational principle that has not yet taken hold (Wexler, 2003). Thus, children with SLI are viewed as differing from their typical peers not in their learning ability but rather in possessing a grammar that in at least one key respect has remained in an immature state.

An inspection of the assumptions of the EOI account reveals that it possesses many of the essential properties that any successful account of the tense/agreement difficulties in SLI must possess. The well-attested variability in using tense/agreement morphemes is, of

course, the central feature of the EOI account, which renders it highly relevant. Because children with SLI are assumed to understand tense and the contexts in which it applies, the low frequency of commission errors (as in *They runs*) seen in SLI is consistent with this account. Furthermore, the children's understanding of tense allows for the possibility that over-regularization errors (as in *runned*) could also be found in these children's speech. As we noted earlier, such errors have been reported in many studies on SLI.

Schütze and Wexler (1996) (for typically developing children), and Wexler, Schütze, and Rice (1998) (for children with SLI) recognized that the EOI account failed to account for one important detail. Children occasionally alternate between utterances such as *Him played basketball yesterday* and *He play basketball yesterday*. The latter can be characterized as the selection of a non-finite verb form in a context requiring tense in the adult grammar. It can be seen that the pronoun used in subject position (*he*) is in the appropriate nominative case. Within the linguistic framework adopted by Rice, Wexler, and their colleagues, nominative case is licensed by agreement, not tense. Thus, the absence of tense from this utterance has no adverse affect on the subject pronoun. In contrast, the utterance *Him played basketball yesterday* includes a subject pronoun that fails to show nominative case. (The pronoun is in the default form for English. For example, the answer to the question "Who wants ice cream?" can be "Me!" in English, though it is "I" in many other languages.) The fact that the utterance *Him played basketball yesterday* contains an appropriate past tense *-ed* inflection seems to confirm the assumption that nominative case is licensed by agreement, not tense. However, at the same time, the use of *him* as well as *he* in subject position suggests that agreement, like tense, might also be optional. Wexler, Schütze, and their colleagues therefore proposed that either tense or agreement can be optional in the grammars of children. This proposal was termed the agreement-tense omission model (ATOM). As in the case of the EOI account, it is assumed that children with SLI remain in this optional period for a prolonged period of time.

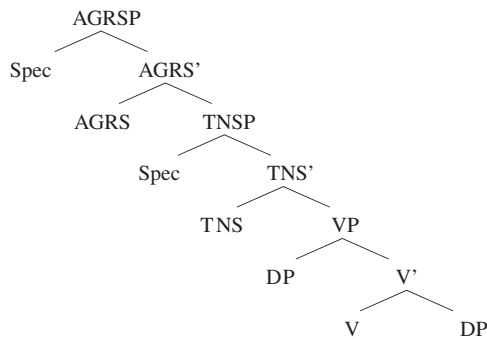
Both the EOI and ATOM accounts receive considerable support from investigations of English-, Swedish-, German-, Dutch-, and French-speaking children with SLI (e.g., Wexler et al., 1998; de Jong, 1999; Hansson, Nettelblatt, & Leonard, 2000; Paradis & Crago, 2000). It can be recalled that when English-speaking children fail to use a lexical verb marked for tense and agreement, they presumably use a non-finite form. However, in English, this is a bare stem. Thus, it is not clear that a production such as *Chris play basketball yesterday* includes a non-finite form or simply reflects a failed attempt to include the past tense inflection *-ed*. In languages such as Swedish, these two interpretations can be distinguished. Although Swedish permits bare stems (e.g., the bare stem *spring* "run" is used as an imperative), the infinitive form carries an overt inflection (*springa* "to run"). In contexts requiring present and past tense verb forms, Swedish-speaking children with SLI alternate between appropriate tense forms (e.g., present tense *springer*) and the infinitive forms of the same verb (e.g., *springa*).

Proponents of the EOI and ATOM accounts acknowledged that the period of optional use of lexical verbs requiring tense and agreement is not seen in children learning "null-subject" languages. These are languages such as Italian and Spanish that do not require

the inclusion of the subject of the sentence if the non-linguistic or linguistic context makes the referent of the subject clear. The inflections used with lexical verbs in these languages mark both tense and agreement, and bare stems are not permitted. Thus, in Italian, *dormo*, *dorme*, and *dormono* translate as “(I) sleep,” “(he/she/it) sleeps,” and “(they) sleep,” respectively, and the stem *dorm* is never used without an inflection.

To accommodate the pattern of use seen in these languages without losing the notion of optionality, Wexler (1998, 2003) proposed a modification of the EOI and ATOM accounts. He assumes a structure containing the functional categories tense or TNS and subject agreement or AGRS. The addition of these functional phrases above the verb phrase (VP) reflects an elaboration of the syntactic tree; the idea that agreement and tense features project separate phrases in the tree has been supported most directly by word order facts (see Radford, 1997). For languages such as English and German, both TNS and AGRS have a non-interpretable determiner (D) feature, and sentence subjects, represented as determiner phrases or DPs, contain an interpretable D feature. This structure is illustrated in (1).

1.

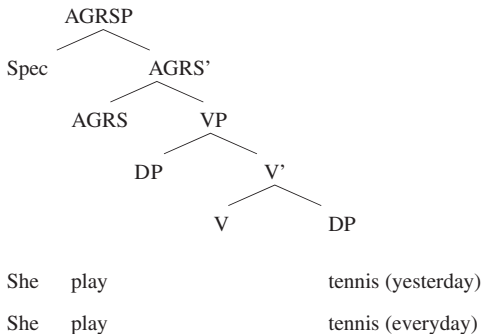


In the linguistic framework adopted by Wexler (1998, 2003), the checking of non-interpretable features is obligatory. As a result, the D feature in TNS attracts the subject DP, leading it to raise to the specifier or Spec position of TNSP for checking. In addition, the D feature of AGRS attracts the subject DP to the Spec position of AGRSP, permitting checking of the D in AGRS. As can be seen, the D feature of DP checks against two functional categories, namely, TNS and AGRS. Whether or not the mechanism of feature-checking in AGRSP and TP is a psychologically real account of speakers' knowledge of verbal inflection, it has been argued that children's acquisition patterns can be described quite accurately with these assumptions. In particular, the ATOM model of Schütze and Wexler (1996) demonstrates that the error patterns of young children reflect the separability of AGRS and TNS features, allowing co-existing sequences such as *Him played yesterday* and *He play yesterday*.

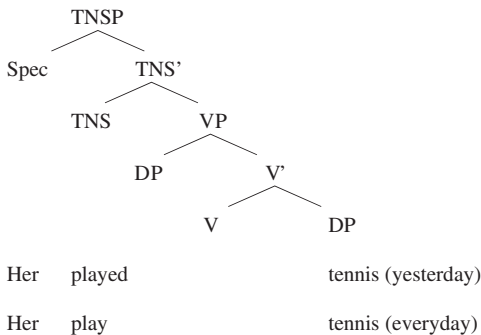
In Wexler's (1998, 2003) new account, optional use is viewed as the result of a limitation on the number of functional categories at which features can be checked. Checking can occur at only one functional category. This constraint, the “unique checking constraint”

(UCC), is operative for only a short time in typical development. However, for children with SLI, this period is extended in time, and can therefore be referred to as an “extended unique checking constraint” (EUCC). As noted above, for languages such as English and German, two checking operations are required for the expression of tense/agreement morphemes. When the EUCC applies, then, there are two possibilities, illustrated in (2) and (3).

2.



3.



In (2), only AGRS is projected and checking occurs there. Because AGRS is responsible for nominative case, the pronoun *she* can occur. The past tense form *played* is not possible because TNS was not projected. Thus, a non-finite form, *play* is used. Note that *plays* is also not possible. The third person singular function of *plays* might be permitted thanks to checking at AGRS. However, the morpheme *-s* expresses tense as well as agreement. Because TNS was not projected, the expression of this morpheme is blocked. In (3), the converse applies. With no projection of AGRS, nominative case is not permitted. Past tense *-ed* is possible because this morpheme marks tense only, and checking has taken place at TNS. Again, *plays* is not possible, in this instance because the absence of AGRS provides no basis for the (third person singular) agreement function of *-s*.

For languages such as Italian and Spanish, Wexler assumes that the null-subject property of the language permits the expression of agreement without the need for checking. However, as in English and German, the non-interpretable D feature at TNS must be

checked. Because checking only occurs at one functional category, the EUCC is not violated. Consequently, Italian- and Spanish-speaking children with SLI should not experience special difficulties with verb inflections (e.g., present tense first person singular *-o* in Italian, preterite first person singular *-í* in Spanish). Findings from Leonard et al. (1992) and Bedore and Leonard (2005) are consistent with this prediction. In those studies, children with SLI were highly similar to MLU-matched children in their use of these inflections.

## 6. DEGREES OF OPTIONALITY?

Although proposals of an optional or single-checking stage of development have clear merit, they appear to be insensitive to important differences in tense/agreement morpheme use that can be found between children, between languages, and even within the same child. We discuss these in turn.

### 6.1. Children with SLI and Younger Typically Developing Children in the Same Language

To illustrate one type of insensitivity, we can refer again to the data shown in Table 1. Here, the children with SLI used tense/agreement morphemes with significantly lower percentages of use than did the younger typically developing children matched for MLU. However, an inspection of the percentages of use in Table 1 reveals that the MLU-matched children, too, were well below levels of mastery in their use of these morphemes. Put differently, they, too, were in an optional period. Group differences when both groups are showing optional use may not seem like a serious obstacle to optionality accounts of SLI, until one recalls that this type of finding is omnipresent in the literature. It is routine to find that typically developing three-year-olds average from 50% to 80% use of tense/agreement morphemes in obligatory contexts, whereas five-year-old children with SLI average from 20% to 60%. The distinction between “optional use” and “obligatory/adult use” does not capture this reliable difference found in the literature.

### 6.2. Children with SLI in Different Germanic Languages

As noted earlier, children with SLI who are acquiring languages such as English, German, Swedish, and Dutch show variable use that is quite compatible with the predictions of the optional period accounts. In particular, children with SLI in each of these languages seem to alternate between lexical verbs appropriately inflected for tense and agreement, and lexical verbs in non-finite form. However, there is also an important crosslinguistic difference. Children with SLI-acquiring English appear to use the tense and agreement inflections with significantly lower percentages than children with SLI in these other languages. An example can be seen in Table 2. The data for past tense inflections come from a study by Leonard, Hansson, Nettelblatt, and Deevy (2004) that compared English- and Swedish-speaking children with SLI. The children were carefully matched on the severity of their language disorder (as determined by standardized tests), their MLU measured in words, and their chronological age. As can be seen, the Swedish-speaking children with SLI used past tense inflections with higher percentages than did

Table 2

Percentages of use of grammatical inflections by children with SLI in four languages.

	English <sup>a,b</sup>	Swedish <sup>c</sup>	German <sup>d</sup>	Dutch <sup>e</sup>
Past tense				
M	53	86		
SD	32	18		
Third person singular present tense				
M	34		50	61
SD	24		21	34

<sup>a</sup>English past tense data from Leonard et al. (2004).<sup>b</sup>English third person singular present tense data from Leonard et al. (1997).<sup>c</sup>Swedish past tense data from Leonard et al. (2004).<sup>d</sup>German third person singular present tense data from Roberts and Leonard (1997).<sup>e</sup>Dutch third person singular present tense data from de Jong (1999).

their English-speaking counterparts. This is true despite the fact that the past tense inflections in each language mark tense only (not agreement), and, importantly, even the Swedish-speaking children with SLI were found to have more difficulty with past tense inflections than younger MLU-matched compatriots (Hansson et al., 2000). Although the Leonard et al. (2004) study seems to be one of the very few that employed direct crosslinguistic comparisons, an inspection of the percentages of use reported for languages such as German (Roberts & Leonard, 1997) and Dutch (de Jong, 1999) suggest that the crosslinguistic advantage over English is not limited to Swedish. As can be seen from Table 2, English, German, and Dutch make use of third person singular inflections in present tense. Although the children from these three language groups came from separate studies, they were similar in age and severity of language impairment. As with past tense inflections, the lowest percentages of use of third person singular present tense inflections are seen for English. Later in this chapter, we suggest some possible reasons why English-speaking children might be disadvantaged with regard to tense/agreement morphology, and how these reasons might be factored into an account of SLI.

### 6.3. Differences in Degree of Use Following Intervention

A third source of evidence relating to problems with optionality comes from the language intervention literature. An investigation by Leonard, Camarata, Brown, and Camarata (2004) provides a suitable example. Thirty-one children with SLI received treatment focusing on either third person singular *-s* (the “3S” children) or auxiliary *be* forms (*is, are, was*) (the “AUX” children). At the outset, the children showed no use of these morphemes. Following treatment, the children were again assessed for their use of both the target and non-target morphemes. (Leonard et al. also selected other morphemes absent from the children’s speech to serve as control forms. These will not be discussed here.) The children participated in 48 treatment sessions, covering a span of approximately three months. The two groups of children made greater gains on the target



morphemes than on the non-target morphemes. For the 3S children, gains were greater for third person singular *-s* than for auxiliary *be* forms, whereas the reverse was true for the AUX children. Importantly, these systematic differences occurred in spite of the fact that the percentages of use for both the target and non-target morphemes were well above zero and thus reflected optional use.

#### 6.4. Changes in Degree of Use within the Same Task.

Another reason why current notions of optional use seem incomplete is that predictable changes can occur in children's degree of use of tense/agreement morphemes within the same speaking task. Previous studies have suggested that these morphemes might be vulnerable when the child produces a sentence that is relatively long or complex (Chiat & Hirson, 1987; Fletcher, 1992; Bishop, 1994). Leonard et al., (2000) examined this issue from the standpoint of sentence formulation demands. These investigators employed a syntactic priming task, adapted from adult work by Bock and her colleagues (e.g., Bock & Loebell, 1990). Five-year-old children with SLI and three-year-old typically developing children participated. All children were using tense/agreement morphemes on an inconsistent basis. The children were shown drawings on a computer screen that could be readily described with a present progressive construction (e.g., *The Grinch is reading a book*). Of special interest was the children's tendency to include (or exclude) the auxiliary *is* in these target sentences. Each target sentence was preceded by a prime sentence, where the child saw a drawing and was asked to repeat a descriptive sentence that either employed the syntactic frame needed for the following target sentence (e.g., *The cats are drinking the milk*) or employed a different type of syntactic frame (e.g., *The monkey jumped*).

Following current language production models (e.g., Bock & Levelt, 1994), Leonard et al. (2000) assumed that when a syntactic frame has been retrieved for use in a sentence it maintains a higher activation level than other frames and is thus more readily retrieved for use in a subsequent sentence. This facilitated retrieval, Leonard et al. further assumed, might allow children to expend fewer resources in the subsequent stages of sentence production, including the retrieval and insertion of the auxiliary *is*. Accordingly, these investigators predicted greater use of auxiliary *is* in target sentences following primes such as *The cats are drinking the milk* than following primes such as *The monkey jumped*. The findings were consistent with these predictions. Both groups of children showed the expected priming effect. However, the children with SLI showed a significantly larger priming effect than did the younger typically developing children. These priming effects indicate that even when children with SLI are showing levels of use that can be described as optional, their degree of use within this variable range can be altered in systematic ways. We shall return to this type of finding in a later section when discussing processing limitations in children with SLI.

#### 6.5. Differences when the Adult Grammar Permits Optionality

Another instance in which degree of variability proves relevant can be found in a language such as Cantonese. Although Cantonese does not employ tense and agreement,

grammatical aspect is used to convey notions such as the continuous nature of an action or the completion of an action. These morphemes are free-standing syllables that usually appear after the verb. Examples include the continuous aspect morpheme *gan2* and the perfective aspect morpheme *zo2* (numbers refer to the type of tone required for the morpheme). Importantly, these morphemes are not obligatory, even in the adult grammar. For every sentence containing an aspect morpheme, an otherwise identical sentence that does not contain the morpheme would be fully grammatical.

Fletcher, Leonard, Stokes, and Wong (2005) constructed a set of probes that constituted contexts that encouraged (though could not obligate) the use of aspect morphemes in the speech of adults and typically developing Cantonese-speaking children. They then compared the probe descriptions of five-year-old children with SLI, a group of same-age peers, and a group of younger typically developing children with comparable MLUs. Although no group showed consistent use of the aspect morphemes, the children with SLI made significantly less use of these morphemes (36–52%) than both the younger typically developing group (62–77%) and the age-matched group (73–85%). As can be seen, the pattern of results resembles the pattern seen for languages in which the comparison morphemes are obligatory in the adult grammar.

How should such findings be characterized? None of the groups can reasonably be described as being in a pre-mastery, optional period of development, given that the adult grammar allows optional use. The especially limited use by the children with SLI, then, seems to be even less interpretable from an optional period point of view. These children may significantly underutilize aspect morphemes, but not because they make an assumption about optionality that is qualitatively different from that of the other children.

In summary, optional use and single checking constraint accounts have several strengths. They provide a basis for expecting variability in tense/agreement use, and, in Wexler's (1998, 2003) recent proposals, a rationale for differences between Germanic languages and null-subject languages is also offered. All of these accounts predict correctly that children with SLI will rarely if ever commit commission errors but will occasionally show over-generalizations. On the other hand, these accounts do not provide a full explanation of variability. Reliable differences between groups of children, between languages, and even between speaking tasks can be found, and in some cases even predicted, even though all children involved appear to be in an optional (or single checking constraint) period of development. Mechanisms or principles that might explain such differences do not seem to be part of these accounts.

## 7. PROPOSALS OF PROCESSING LIMITATIONS

A second type of approach seeks to explain the language difficulties of children with SLI in terms of a limitation in processing. In this type of approach, mental functions are seen as performing within a resource-limited system; processing resources determine how much work can be done within a given period of time. In characterizing children

with SLI, limitations on processing resources have been cast in terms of capacity and in terms of speed. Limited capacity (understood as a restricted or inefficient working memory system) or reduced speed of processing could both lead to consequences for language due to lost or inaccurately encoded information. Accounts of this type differ in whether they assume that this processing limitation is general—affecting non-linguistic as well as linguistic domains—or whether they assume that it is confined to a particular domain or type of processing. We will consider each of these possibilities in turn.

General processing approaches are motivated by research on children with SLI, which has consistently found subtle weaknesses in performance on a range of non-linguistic tasks (e.g., Johnston & Ellis Weismer, 1983; Nelson, Kamhi, & Apel, 1987; Montgomery, 1993). For example, Johnston and Ellis Weismer asked children with SLI and age-matched controls to judge whether two geometric forms were the same or different. Identical geometric forms could differ from each other in degree of rotation. Whereas the form on the left was vertical, the one on the right was either vertical or rotated about its center 45, 90, or 135°. Children with SLI exhibited slower response times than age controls but resembled the control children in showing slower response times with increasing degree of rotation.

Findings such as those above led Kail (1994) to propose a generalized slowing hypothesis for SLI. This account assumes that rather than using separate resources (e.g., for retrieving a word and for executing a motor response to a stimulus), some single mechanism is responsible for the rate of processing in different domains. Because this slowing is general in nature, it is predicted to affect not only tasks from different domains, but also the individual steps in a complex task. Thus, as complexity (quantified simply in number of steps) increases, slowing should increase at a constant proportion.

Kail tested this hypothesis by comparing the reaction times of children with SLI to those of typically developing children matched for chronological age. Plotting group mean response time (RT) data for each task and condition from five previous studies using linguistic and non-linguistic tasks, he found that the mean RTs of children with SLI increased linearly as a function of the mean RTs of the age-matched children and that the increase was proportional (about 33% slower for the SLI group). That is, regardless of the type of process involved in performing a task, the children with SLI were slower than their peers to the same degree. If the slower RTs in children with SLI were due to limitations on only certain independent processes, a linear relationship would not have been found.

These findings have been replicated in another meta-analysis of RT data from previous studies which found that children with SLI were 18% slower than age-matched children (Windsor & Hwang, 1999) and again in a study designed to directly test the hypothesis (Miller, Kail, Leonard, & Tomblin, 2001). Miller et al. compared the performance of a single group of children with SLI to a group of typically developing age-matched peers across 10 different linguistic and non-linguistic tasks involving 42 different conditions. They too found that children with SLI were slower than their peers by a constant proportion (14%) and that the data best fit a proportional regression model.

While group results show differences in processing speed across tasks, there is evidence of heterogeneity within the SLI population in terms of the degree of slowing (Windsor, Milbrath, Carney, & Rakowski, 2001). One interesting possibility is that a higher degree of slowing may be associated with children who show both receptive and expressive language deficits as compared to children who show only expressive language deficits (Lahey, Edwards, & Munson, 1999; Windsor & Hwang, 1999; Miller et al., 2001). There may also be differences in the degree of slowing incurred by different tasks. Windsor et al. (2001) argue that the particular regression analysis commonly used may obscure task-specific differences in speed of processing. In particular, they noted a greater degree of slowing (14%) in a phonological task (rhyme judgment) compared to a perceptual motor task (tapping a key with one finger, 4%).

A more fundamental challenge that researchers who propose the generalized slowing hypothesis have to address is why language is affected more consistently and severely than other cognitive tasks. Miller et al. (2001) suggest that the processing of verbal information is more time-dependent and thus more vulnerable to slow processing than other kinds of information. Presumably, the encoding and storage of verbal representations may be affected to a greater degree by general slowing than other tasks. At this point, such suggestions are quite speculative and in need of empirical testing. Furthermore, it is not clear why grammatical morphology and tense/agreement morphology in particular would be most vulnerable to general speed limitations. We shall return to this point in a later section.

An alternative conception of children's processing abilities uses the more general notion of capacity. Evidence from a variety of studies has suggested that there are restrictions on the quantity of information that children with SLI can retain and use during tasks (e.g., Kirchner & Klatsky, 1985; Johnston, Smith, & Box, 1988; Bishop, 1992). These limitations in functioning could be due to slower, less efficient, or less automatic processing, but they have most commonly been studied from the perspective of inadequate working memory capacity. Although working memory may be considered a processing resource available for cognitive tasks in general, most research in this area has focused more exclusively on its relationship to language, using constructs and tests adapted from the literature on adult working memory and language.

In the model developed by Carpenter and associates (Danneman & Carpenter, 1980; Just & Carpenter, 1992), working memory is a limited capacity system within which storage and processing draw on the same pool of resources. When the demands of a task exceed this capacity, trade-offs between the two functions are seen. A number of tasks have been developed to measure storage and processing capacity in children. In one (Gaulin & Campbell, 1994), children listen to simple sentences and judge the truth of each; after increasingly longer sets of sentences, they must recall the last word of each sentence. Span is measured in number of words recalled.

Using these kinds of tasks, children with SLI have been shown to maintain accuracy in sentence comprehension while retaining fewer words in recall than age-matched controls

(Ellis Weismer, Evans, & Hesketh, 1999; Montgomery, 2000; Marton & Schwartz, 2003). These studies have also attempted to demonstrate direct relationships between working memory span and language performance. In the Ellis Weismer et al. study, the children with SLI showed significantly lower scores than the typical group on a test of comprehension, but only the typical group showed a correlation between these scores and memory span. Montgomery tested children's comprehension of shorter sentences compared to sentences of similar structure and meaning but that had extra (redundant) material, e.g., "the (dirty) little boy climbed the (big) fat tree". He found that children with SLI were significantly less accurate in comprehending the redundant sentences than younger controls. Although this difference would seem to implicate memory limitations, there again was no significant correlation between memory span and task performance for any of the groups. A number of reasons for this null effect can be considered. Two candidates are limitations in the nature of the language tasks chosen for comparison, and the possibility that subgroups within the SLI group are not homogeneous with respect to the deficits underlying their language disorder (Ellis Weismer et al., 1999).

Another model of memory which has been influential in research on SLI is that of Baddeley and his colleagues (e.g., Baddeley, 1986; Gathercole & Baddeley, 1993). This model consists of two modality-specific short-term storage systems (for verbal and visual information) and a "central executive" component which in part coordinates flow of information within working memory. The nature of the short-term verbal store, or "phonological loop" has been studied in children with SLI using the non-word repetition task. Children repeat non-words of increasing length, and accuracy is measured. Gathercole and Baddeley (1990) found that although children with SLI could do as well as peers in repeating words of one or two syllables, words of three or four syllables caused them greater difficulty. These results have been replicated numerous times (Montgomery, 1995; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Botting & Conti-Ramsden, 2001; Ellis Weismer et al., 2000). Their poorer non-word repetition performance has generally been taken to show that children with SLI have a reduced capacity for storage in phonological memory, rather than impaired auditory perception or slow articulation rate (see Gathercole & Baddeley, 1990). The exact nature of the storage deficit continues to be explored. Possibilities include smaller storage capacity, more rapid decay of information, and incomplete initial encoding or recoding of representations.

Gathercole and Baddeley (1990, 1993) assert that this capacity limitation is at least partially responsible for the below age-level language attainment of children with SLI. This account of SLI differs from others in that it appeals to a deficit in a language-specific process, phonological working memory, rather than a general processing capacity. There is evidence for associations between phonological working memory abilities and vocabulary development in typically developing preschoolers (Gathercole, Willis, Emslie, & Baddeley, 1992). In addition to consistently poor non-word repetition abilities, children with SLI show below age-level vocabulary abilities and are less able to learn novel words in experimental contexts (e.g., Ellis Weismer & Hesketh, 1996). It is further asserted that a deficit in phonological working memory could also adversely affect language processing at higher levels, namely sentence comprehension (Gattercole & Baddeley, 1993;

Joanisse & Seidenberg, 1998). Montgomery (1995) showed a strong positive correlation between non-word repetition abilities and performance on the sentence comprehension task described above.

Researchers who have investigated other aspects of working memory in children with SLI, including the central executive and visual storage components, question the idea that a deficit specific to phonological working memory is responsible for language impairment. For example, Hoffman and Gillam (2004) tested recall of visual and verbal information under no load and secondary load conditions. Sequences of digits or “X”s were presented visually; recall modality was verbal (for digits) or pointing (for Xs). A secondary task performed before recall required identifying the color of the stimuli either verbally or by pointing. Children with SLI recalled fewer items than age-matched controls in both visual and verbal recall conditions, in both no load and secondary load conditions, indicating a deficit in storing and/or retrieving information that is not restricted to the verbal domain. In addition, the typical group showed better verbal recall when the secondary task required pointing rather than a verbal response; the SLI group did not show such a benefit. The authors interpret this as showing that children with SLI have less efficient central executive functioning in that they were not able to benefit from the possibility of dispersing processing effort across the two domains.

Although working memory limitations have been found in children with SLI, statistically significant relationships between various measures of working memory and language ability have not always been found. Some researchers have suggested that the observed limitations in working memory are confined to phonological working memory, while others have found evidence of broader deficits in processing capacity, encompassing visual processing and coordination of processing resources.

## **8. WHY TENSE/AGREEMENT MORPHOLOGY?**

In principle, processing approaches provide a reasonable explanation for the variable use of tense/agreement morphology. When the language task at hand demands more resources than the child can muster, some required element will be adversely affected, through omission or perhaps through substitution by an element more firmly established in the child’s language system. Variability can be explained in two ways. In some circumstances, the language task might require few resources, permitting the child to include all required elements; in other circumstances, the demands are greater, and the element will be omitted or replaced. Alternatively, there could be trade-off effects, such that in one instance, one element might be retained and another omitted, and in another instance, the element retained and the element omitted will be reversed.

However, this type of explanation has two major shortcomings. First, it has not yet been put to a test. Although there have been many studies that have manipulated the amount of information to be processed, few if any have sought to determine whether tense/agreement morphemes are more vulnerable than other details of information when available resources have reached their limit. Second, even if there was strong evidence

that tense/agreement morphemes are more likely to be omitted than other elements when processing demands increase, there is no theory-internal reason within the limited processing approaches to expect this specific area to be most in peril. Why should tense/agreement morphology be weaker than other areas and hence more likely to be adversely affected by processing factors?

## **9. TOWARD A CLOSER RELATIONSHIP BETWEEN KNOWLEDGE AND PROCESSING**

Thus far, neither the most promising grammatical knowledge approaches nor the available processing approaches seem to provide a fully adequate explanation for tense/agreement morpheme variability. Because such variability constitutes the most dependable means of identifying language impairment at present, the incomplete picture painted by each of these alternative approaches represents a significant problem. One profitable direction for future research would be to carefully examine theoretically coherent ways in which the assumptions of both knowledge and processing approaches might be compatible and incorporated into the same account. There are two ways in which such a merger might take place. First, limitations in grammatical knowledge and processing deficits might be independent, co-occurring difficulties in children with SLI. Second, the input might have a more incremental influence on children's grammars than previously assumed, and processing limitations may undermine the degree and timing of input effects in children with SLI.

### **9.1. Optional Use and Processing Limitations as Independent, Co-Occurring Difficulties.**

It might be the case that the grammars of children with SLI have a constraint that limits the number of functional categories at which features can be checked. However, these children may also have processing limitations such that the constraint is most likely to take effect when demands on available resources are high. This is not an unnatural marriage between these two forces; an inspection of the normal language development literature suggests that young, typically developing children do not remain at a particular level of optional use (e.g., 50%) and then suddenly show 90% or greater use within a few days. Instead, there seems to be a progression from relatively little to greater use before mastery levels are attained. As the constraint is being shed, use or non-use of the tense/agreement morpheme is probabilistic. Processing demands in the moment may well influence these probabilities.

This type of relationship between knowledge and processing might also provide a plausible explanation for the protracted nature of the optional period seen in SLI. If the processing limitations of children with SLI are significant, these children may persist in producing less mature forms even after the constraint is no longer in effect. That is, non-finite syntactic frames with a long history of being retrieved may have activation levels that allow them to compete with finite frames even when the original basis for their retrieval is no longer present.

Although the co-occurrence of knowledge and processing limitations remains somewhat speculative, there is evidence that supports it. In the syntactic priming study of Leonard et al. (2000) described earlier, children's use of auxiliary *is* in target sentences was facilitated by prime sentences that employed the syntactic frame needed for the target. However, Leonard et al. observed instances in which the children failed to repeat the prime sentence accurately, producing utterances such as *The girls flying the kite* instead of *The girls are flying the kite*. On these occasions, the children usually failed to include the auxiliary *is* in their production of the subsequent target sentence. Leonard et al. reasoned that these particular prime sentence productions may have been instances in which the children selected the non-finite option, which was compatible with their grammars. Because a non-finite sentence had been produced, a non-finite syntactic frame (cf. *We saw the girls flying the kite*) rather than a finite syntactic frame was activated and thus more readily retrieved.

To test for this possibility, Leonard et al., (2002) conducted a new study that included a non-finite as well as a finite priming condition. In the non-finite condition, the child and his/her adult assistant were asked "what do you see?" and the assistant provided primes to be repeated of the type [*We see*] *the mouse eating the cheese*. As in the earlier Leonard et al. (2000) study, the target sentences obligated use of auxiliary *is*. The question was whether after producing a non-finite form such as *The mouse eating the cheese* the children would be more likely to exclude auxiliary *is* from the target sentence (*The goose chasing the cat* rather than *The goose is chasing the cat*). Leonard et al. (2002) found that indeed this was the case. Furthermore, the degree to which they excluded the auxiliary *is* from the target following a non-finite prime was virtually identical to the degree to which they excluded *is* from the target following a finite prime that they failed to repeat with the auxiliary verb (e.g., repeating *The dogs are chewing the sock* as *The dogs chewing the sock*).

The findings emerging from this study support not only the notion of priming effects that might be related to processing demands, but also gives further credence to a key assumption of the optional use approaches. Note that if the children had been attempting to repeat the finite prime sentences accurately and simply committed the performance error of leaving out the auxiliary form, their prime sentence productions would have involved the same finite syntactic frame needed for the target sentence. With the syntactic frame activated, the retrieval of the same frame for the target sentence production should have been facilitated, leaving more resources for the final step of retrieving the necessary auxiliary verb. However, such a facilitative effect was only seen when the finite prime sentence was produced with the auxiliary form. This suggests that prime sentence repetitions lacking the auxiliary form may have been attempts at non-finite sentences. Across these two studies, then, evidence consistent with optional use or single checking accounts was found in the form of apparent use of non-finite sentence frames, and evidence consistent with processing accounts was seen in the form of non-finite frames increasing the likelihood of non-finite descriptions of the target and finite frames enhancing the likelihood of finite descriptions of the target.



## 9.2. Processing Limitations Affecting the Degree and Timing of Incremental Influences of Input on Children's Grammars

A second possible way that knowledge and processing factors may operate together is to assume that the types of insights credited to children in linguistic knowledge accounts actually require more evidence from the input than is usually acknowledged. Here, knowledge is not defined as principles dependent only on biological maturation and minimal input, but rather as the accumulation of evidence from the input. Because processing limitations influence the consistency with which children with SLI interpret exemplars in the input, their development will be slower than that of typically developing peers.

Are there findings in the literature that are consistent with this possibility? Certainly one possible indication might be the differences seen among the Germanic languages in "degree of optionality" that we reported earlier. Assuming that biological maturation is comparable across all of the children, it would seem that the "minimal" input that is required must differ somewhat from language to language. Incomplete processing of exemplars in the input due to processing limitations would be the functional equivalent of reductions in input frequency. Thus, children with SLI should show the same crosslinguistic differences seen in typical development but these children should lag behind their typically developing peers in each language.

The increased importance placed on the input in this scenario is certainly reminiscent of proposals by Bates and MacWhinney (1987) and others (see MacWhinney, 1999) who have discussed multiple cues that are used by the child to build up distributional "packages" that, functionally, operate like grammatical categories. Bishop (1994) has noted that the same may be the case for children with SLI. We, too, acknowledge such a possibility; if it proves true that typically developing children arrive at grammatical categories in this manner, we can see how processing limitations could adversely affect this process in children with SLI.

However, assuming that input effects on grammars are more incremental than immediate does not require a departure from all assumptions made in linguistic knowledge accounts. For example, such an assumption is quite compatible with approaches that assume symbol-based representations. Pinker (1984, p. 191) discusses how typically developing children must hypothesize the function of morphemes appearing in the input, and place these morphemes in a paradigm. Importantly, these entries change in strength as a function of the number of instances in which the appearance of the morpheme in the input is consistent with the hypothesis of its function. This process might well proceed more slowly in languages with many bare stems as in English. For example, consider the case of the tense/agreement morpheme *-s* as in *likes*. Bare verb stems appear in third person present tense contexts (e.g., *They like*) and singular present tense contexts (*I like, you like*) and therefore, as potential competitors for the same cell in the paradigm, could influence the number of encounters with *-s* that are required for the morpheme to achieve sufficient strength to be used consistently.

This might be especially true when processing limitations are in play. For example, it seems possible that inflected words might decay partially and then be mistaken as the bare stem due to the similarity between the two forms. Alternatively, due to slower processing, the processing of an inflected word might be abandoned prior to the inflection in favor of the next word appearing in the speech stream.

It can be seen that the relationship is one of processing limitations affecting knowledge by altering the reliability with which the input arrives in a form that can influence the children's grammars. This relationship is unidirectional; limitations in linguistic knowledge are not assumed except for those created by inconsistent processing of the input.

## 10. SUMMARY

In this chapter, we have introduced a group of children whose difficulties with language cannot be attributed to problems with hearing, neurological impairment, or serious cognitive limitations. This apparent discrepancy between language and other abilities offers an interesting test case for theories of language development, as does the most common profile seen in SLI, a deficit in tense/agreement morphology that is more serious than problems in other areas of language. We have discussed the strengths and weaknesses of the two prominent types of accounts that have been advanced to explain these tense/agreement difficulties. One of these relies heavily on assumptions about linguistic knowledge; the other focuses on processing ability. We believe that close examination of the core assumptions of these two types of accounts will reveal ways that they can be considered together without losing their theoretical coherence. Such integration should lead to a more comprehensive explanation of this disorder.

## REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders IV*: Washington, DC: American Psychiatric Association.
- Baddeley, A. (1986). Working memory and comprehension. In: D. Broadbent, J. McGaugh, M. Kosslyn, N. Mackintosh, E. Tulving, & L. Weiskrantz (Eds), *Working memory* (pp. 33–74). Oxford: Oxford University Press.
- Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In: B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–194). Hillsdale, NJ: Lawrence Erlbaum.
- Bedore, L., & Leonard, L. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 42, 678–689.
- Bedore, L., & Leonard, L. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Psycholinguistics*, 26, 195–225.

- Beitchman, J., Wilson, B., Brownlie, E., Walters, H., & Lancee, W. (1996). Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 804–814.
- Bishop, D. V. M. (1992). The underlying nature of specific language impairment. *Journal of Child Psychology and Psychiatry*, 33, 3–66.
- Bishop, D. V. M. (1994). Grammatical errors in specific language impairment: Competence or performance limitations? *Applied Psycholinguistics*, 15, 507–550.
- Bishop, D. V. M. (2002). Putting language genes into perspective. *Trends in Genetics*, 18, 57–59.
- Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment and phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, 31, 1027–1050.
- Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: Evidence from a twin study. *Developmental Medicine and Child Neurology*, 37, 56–71.
- Bishop, D. V. M., & Snowling, M. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, 130, 858–886.
- Bliss, L. (1989). Selected syntactic usage of language-impaired children. *Journal of Communication Disorders*, 22, 277–289.
- Bock, K., & Levelt, W. (1994). Grammatical encoding. In: M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.
- Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.
- Botting N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment. *International Journal of Language and Communication Disorders*, 36, 421–432.
- Catts, H., Fey, M., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, 45, 1142–1157.
- Chiat, S., & Hirson, A. (1987). From conceptual intention to utterance: A study of impaired language output in a child with developmental dysphasia. *British Journal of Disorders of Communication*, 22, 37–64.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Clahsen, H. (1989). The grammatical characterization of developmental dysphasia. *Linguistics*, 27, 897–920.
- Cleave, P., & Rice, M. (1997). An examination of the morpheme BE in children with specific language impairment: The role of contractibility and grammatical form class. *Journal of Speech, Language, and Hearing Research*, 40, 480–492.

Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001). Follow-up of children attending infant language units: Outcomes at 11 years of age. *International Journal of Language and Communication Disorders, 36*, 207–219.

Danneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

de Jong, J. (1999). Specific language impairment in Dutch: Inflectional morphology and argument structure. *Groningen Dissertations in Linguistics, 28*, 1–203.

Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1136–1146.

Edwards J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19*, 279–309.

Ellis Weismer, S., & Hesketh, L. (1996). Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *Journal of Speech, Language, and Hearing Research, 39*, 177–190.

Ellis Weismer, S., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1249–1260.

Ellis Weismer, S., Tomblin, B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43*, 865–878.

Fletcher, P. (1992). Subgroups in school-age language-impaired children. In: P. Fletcher, & D. Hall (Eds), *Specific speech and language disorders in children* (pp. 152–163). London: Whurr.

Fletcher, P., Leonard, L., Stokes, S., & Wong, A. M. -Y. (2005). The expression of aspect in Cantonese-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 621–634.

Gathercole, S., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336–360.

Gathercole, S., & Baddeley, A. (1993). *Working memory in language processing*. Hove, East Sussex, England: Lawrence Erlbaum.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887–898.

Gaulin, C., & Campbell, T. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills, 79*, 55–64.

Gopnik, M., & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition, 39*, 1–50.

- Hansson, K., Nettelblatt, U., & Leonard, L. (2000). Specific language impairment in Swedish: The status of verb morphology and word order. *Journal of Speech, Language, and Hearing Research, 43*, 848–864.
- Haynes, C., & Naidoo, S. (1991). *Children with specific speech and language impairment*. London: Mac Keith.
- Hoffman, L., & Gillam, R. (2004). Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 114–125.
- Joanisse, M., & Seidenberg, M. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences, 2*, 240–247.
- Johnston, J. (1988). Specific language disorders in the child. In: N. Lass, L. McReynolds, J. Northern, & D. Yoder (Eds), *Handbook of speech-language pathology and audiology* (pp. 685–715). Toronto: Decker.
- Johnston, J., & Ellis Weismer, S. (1983). Mental rotation abilities in language disordered children. *Journal of Speech and Hearing Research, 26*, 397–403.
- Johnston, J., Smith, L., & Box, P. (1988). Six ways to skin a cat: Communication strategies used by language impaired preschoolers. Paper presented at the symposium on research in child language disorders, University of Wisconsin, Madison.
- Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.
- Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech and Hearing Research, 37*, 418–421.
- Kirchner, D., & Klatsky, R. (1985). Verbal rehearsal and memory in language disordered children. *Journal of Speech and Hearing Research, 28*, 556–564.
- Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 47*, 924–943.
- Lahey, M., Edwards, J., & Munson, B. (1999). Is processing speed related to severity of language impairment? *Journal of Speech, Language, and Hearing Research, 44*, 1354–1361.
- Leonard, L. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
- Leonard, L. (2000). Specific language impairment across languages. In: D. V. M. Bishop, & L. Leonard (Eds), *Speech and language impairments in children: Causes, characteristics, intervention, and outcome* (pp. 115–129). Hove, UK: Psychology Press.
- Leonard, L., Bortolini, U., Caselli, M. C., McGregor, K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition, 2*, 151–179.

Leonard, L., Camarata, S., Brown, B., & Camarata, M. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research, 47*, 1363–1379.

Leonard, L., Eyer, J., Bedore, L., & Grela, B. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*, 741–753.

Leonard, L., Hansson, K., Nettelbladt, U., & Deevy, P. (2004). Specific language impairment in children: A comparison of English and Swedish. *Language Acquisition, 12*, 219–246.

Leonard, L., Miller, C., Deevy, P., Rauf, L., Gerber, E., & Charest, M. (2002). Production operations and the use of nonfinite verbs by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45*, 744–758.

Leonard, L., Miller, C., Grela, B., Holland, A., Gerber, E., & Petucci, M. (2000). Production operations contribute to the grammatical morpheme limitations of children with specific language impairment. *Journal of Memory and Language, 43*, 362–378.

MacWhinney, B. (Ed.) (1999). *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum.

Marchman, V., Wulfeck, B., & Ellis Weismer, S. (1999). Morphological productivity in children with normal language and SLI: A study of the English past tense. *Journal of Speech, Language, and Hearing Research, 42*, 206–219.

Marton, K., & Schwartz, R. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1138–1153.

Miller, C., Kail, R., Leonard, L., & Tomblin, B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*, 416–433.

Miller, C., & Leonard, L. (1998). Deficits in finite verb morphology: Some assumptions in recent accounts of specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 701–707.

Montgomery, J. (1993). Haptic recognition of children with specific language impairment: Effects of response modality. *Journal of Speech and Hearing Research, 36*, 98–104.

Montgomery, J. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research, 38*, 187–199.

Montgomery, J. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43*, 293–308.

Nelson, L., Kahmi, A., & Apel, K. (1987). Cognitive strengths and weaknesses in language impaired children: One more look. *Journal of Speech and Hearing Disorders, 52*, 36–43.

Oetting, J., & Horohov, J. (1997). Past tense marking by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*, 62–74.

- Paradis, J., & Crago, M. (2000). Tense and temporality: A comparison between children learning a second language and children with SLI. *Journal of Speech, Language, and Hearing Research, 43*, 834–847.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Radford, A. (1997). *Syntactic theory and the structure of English: A minimalist approach*. Cambridge, UK: Cambridge University Press.
- Redmond, S., & Rice, M. (2001). Detection of irregular verb violations by children with and without SLI. *Journal of Speech, Language, and Hearing Research, 44*, 655–669.
- Rice, M., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39*, 1239–1257.
- Rice, M., Wexler, K., & Cleave, P. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research, 38*, 850–863.
- Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1412–1431.
- Roberts, S., & Leonard, L. (1997). Grammatical deficits in German and English: A crosslinguistic study of children with specific language impairment. *First Language, 17*, 131–150.
- Scarborough, H. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728–1743.
- Scarborough, H. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia, 41*, 207–220.
- Schütze, C., & Wexler, K. (1996). Subject case licensing and English root infinitives. In: A. Stringfellow, D. Cahana-Amitay, E. Hughes, & A. Zukowski (Eds), *Proceedings of the 20th Annual Boston University Conference on Language Development*, Volume 2 (pp. 670–681). Somerville, MA: Cascadilla Press.
- Snowling, M., Bishop, D. V. M., & Stothard, S. (2000). Is pre-school language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry, 41*, 587–600.
- Tallal, P., Curtiss, S., & Kaplan, R. (1988). The San Diego longitudinal study: Evaluating the outcomes of preschool impairments in language development. In: S. Gerber, & G. Mencher (Eds), *International perspectives on communication disorders* (pp. 86–126). Washington, DC: Gallaudet University Press.
- Tallal, P., Hirsch, L., Realpe-Bonilla, T., Miller, S., Brzustowicz, L., Bartlett, C., & Flax, J. (2001). Familial aggregation in specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*, 1172–1182.
- Tomblin, J. B. (1989). Familial concentration of developmental language impairment. *Journal of Speech and Hearing Disorders, 54*, 287–295.

- Tomblin, J. B., Freese, P., & Records, N. (1992). Diagnosing specific language impairment in adults for the purpose of pedigree analysis. *Journal of Speech and Hearing Research, 35*, 832–843.
- Tomblin, J. B., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*, 1245–1260.
- van der Lely, H. (1997). Language and cognitive development in a grammatical SLI boy: Modularity and innateness. *Journal of Neurolinguistics, 10*, 75–107.
- van der Lely, H. (1998). SLI in children: Movement, economy, and deficits in the computational-syntactic system. *Language Acquisition, 7*, 161–192.
- van der Lely, H., & Ullman, M. (2001). Past tense morphology in specifically language impaired children and normally developing children. *Language and Cognitive Processes, 16*, 177–217.
- Weiner, P. (1986). The study of childhood language disorders: Nineteenth century perspectives. *Journal of Communication Disorders, 19*, 1–47.
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106*, 23–79.
- Wexler, K. (2003). Lenneberg's dream: Learning, normal language development, and specific language impairment. In: Y. Levy, & J. Schaeffer (Eds), *Language competence across populations: Toward a definition of specific language impairment* (pp. 11–61). Mahwah, NJ: Lawrence Erlbaum.
- Wexler, K., Schütze, C., & Rice, M. (1998). Subject case in children with SLI and unaffected controls: Evidence for the Agr/Tns Omission Model. *Language Acquisition, 7*, 317–344.
- Windsor J., & Hwang, M. (1999). Testing the generalized slowing hypothesis in specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1205–1218.
- Windsor, J, Milbrath, R., Carney, E., & Rakowski, S. (2001). General slowing in language impairment: Methodological considerations in testing the hypothesis. *Journal of Speech, Language, and Hearing Research, 44*, 446–461.



This page intentionally left blank

## Subject Index

- abstract words, 674–675  
 Abstractionist, 222  
 Accommodation, 222, 237  
 acoustic noise contamination, 129  
 Action-contingent analyses, 871–872  
 activation, 250–251, 253–254, 256–257, 263–272  
     Sphere, 407, 412  
     Spreading activation, 406–407, 409–414, 417, 434, 436, 440  
 adaptations, generalized vs. particularized, 921  
 adaptations, self-prompted vs. other-prompted, 908, 923, 926  
 Adjectives  
     Depictive, 539  
     Resultative, 539  
     Scalar, 550  
 adult readers, 801, 803, 808, 824  
 age-of-acquisition, 36, 287, 316, 972, 985  
 agrammatism, 107  
 agreement, 66, 69–70, 78, 83, 684, 690, 692, 694, 700–701, 705, 1147–1157, 1159, 1161–1162, 1164–1165  
 agreement-tense omission model (ATOM), 1151  
 AGRS, 1152–1153  
 algorithm, 942, 944  
 Alignment (in dialogue), 10  
 Alignment and dialogue, 887  
 Alleviation hypothesis, 80  
 Alzheimer's disease (AD), 403, 421, 430, 433  
 Ambiguity, 378, 381, 383, 386, 390–393, 507–508, 512–517, 519–524, 529–532, 564, 906–907, 910–913  
 ambiguity, case, 696–698, 705–706  
 ambiguity resolution, 465, 486, 582, 584, 586, 590–591, 594–595, 597–598, 602  
 American Sign Language (ASL), 1001–1002  
 anaphora, 692  
 anaphoric references, 773, 808  
 angular gyrus, 766–767, 775  
 animacy, 698–699  
 antecedent, 692, 777  
 anterior temporal lobe, 775, 785  
 Anticipatory eye movements, 867, 873–874, 886  
 aphasia, 1013, 1015  
     Broca's, 568  
     Wernicke's, 568  
 aphasics, 767  
 Argument  
     vs. Adjunct, 547, 548  
     Advantage, 548, 550  
     Saturation, 544  
     Structure, 543, 544  
 arithmetic, 700  
 Articulation, 21–22, 36, 38–39, 43–44, 202, 205, 210, 212, 214–221, 233, 235, 237  
 articulator motion, 130  
 Articulatory phonology, 216–218  
 Artificial languages, 881  
 Artificial lexicons, 881  
 Attachment preferences, 984  
 Attention, 203, 204, 206–208, 211, 214, 216, 218, 220, 224, 232, 238  
 audience design, 902, 905, 907–908, 911–914  
 auditory discrimination task, 128  
 autism, 768  
 automatic process, 691, 695  
 Automatic versus controlled processes in semantic memory, 413  
 Automaticity, 83, 972, 981  
 autonomous model, 261, 263  
 auxiliary, 1147, 1149–1150, 1155–1156, 1163  
 Background Knowledge, 736, 802, 804, 811, 813, 815–816, 819, 821–822, 824  
 basal ganglia, 693, 700  
 baseline, 766, 769–771  
 baseline prosody, 514, 516, 528  
 Basic level objects and semantic memory, 416  
 Beginning reading, 1111  
 behavior interleaved gradients (BIG) technique, 134  
 Behaviorism, 3, 5  
 BIA + Model, 988–989  
 Bigram Frequency Effects, 292–293, 308, 346

- Bilingual Interactive Activation Model (BIA), 977  
     lexicon, 968–969, 979  
     mode, 973  
     sentence processing, 985–987  
     word recognition, 975, 978, 983–984, 988  
 Bilinguals, 967–976, 978–990  
 Binding theory, 872  
 blend, 24–25  
 blended space, 782  
 blood oxygenation-level dependent (BOLD)  
     contrast, 126  
 Bock-Levelt (BL) model, 62  
 Bottom-up processing, 978  
 boundary tone, 511–512  
 brain, 765–772, 774, 779, 784, 790–792  
 brain imaging, 767–768, 772, 779  
 break index, 510, 512  
 bridging inference, 772, 777  
 Broca, 955–958, 960  
 Broca's area, 767, 1016  
 Broca's aphasia, 673, 691, 958  
  
 3CAPS Model, 790  
 canonical, 703  
 capacity, 1145, 1158–1161  
 capacity theory of comprehension, 809–811  
 caregivers, 1028, 1042  
 case (features/markings/relations?), 691–692,  
     697–698, 702, 706–707  
 categorical grammar, 71  
 categorical perception, 171–173, 175–177, 181,  
     184, 187, 1001  
 Category-specific impairment (CSI), 420  
 causal relatedness, 771, 777–778  
 central executive, 1160–1161  
 characters' intentions, 786  
 child-directed speech, 1029, 1032  
 Chimera, 205–206  
 Choppers, 236  
 closed class, 672–673  
 closure interval, 109–116  
 cloze probability, 560–561, 667, 679, 681–682,  
     684, 686, 688–689, 704  
 clustered volume acquisition, 134  
 coarse coding theory, 766, 779, 785, 790–791  
 coarse semantic processing, 765, 785  
 coarticulation, 161–164, 215, 219  
 Coarticulatory information in word recognition, 881  
 Cocktail party problem, 203  
 codability, 29  
 Code switching, 986  
 Coercion  
     Aspectual, 564  
     Complement, 555  
     and the Effect of context, 567  
     as VP-insertion, 556  
 Cognates, 975–976, 980–981  
 cognitive, 768–773, 775, 778, 780, 783–784, 786,  
     788, 790, 792  
     consequences of bilingualism, 968  
     development, 1034, 1037  
     effect, 836, 848–849, 853, 855  
     effort, 836, 838, 848–849, 855  
     effort and effects trade-off, 849  
     insertion, 128  
     revolution, 6  
     workload, 771–772  
 Coherence, 730–731, 742, 765, 770, 776, 779, 782,  
     784, 788, 790–791  
 coherence monitoring network, 765  
 cohesion, 770  
 cohort, 31, 1002, 1004  
 Collaboration, 2, 901, 906  
 Common ground, 864, 867, 887–888, 902–908,  
     914–915  
 Common ground vs privileged ground, 891  
 communication, 125–127, 130, 143, 901–902,  
     905–906, 914  
 Comparative method, 10  
 Compensation, 234–236  
 competition, 250, 254–255, 258, 263–270, 272,  
     383–385, 388, 391, 474–475  
 Competition Model, 589–590, 983  
 complexity, structural, 666, 693, 701, 703, 705–706  
 component, 661–667, 670–671, 673–674, 677–678,  
     692–693, 697  
 Compositionality  
     Strong vs. weak, 541, 542  
 Comprehension, 479, 488, 728–729, 731, 734,  
     736–737, 743, 745, 765–768, 770–772, 774,  
     776, 779–780, 782, 784–792, 801–803, 805,  
     807–825,  
     comprehension ability, 801–803, 805, 807, 809,  
     811, 813, 815, 817, 819–821, 823  
 Comprehension strategies, 987  
 computational modeling, 583, 595  
 Concept Mediation Model, 969–972  
 Concepts, 1–2, 7  
 conceptual blending theory, 781  
     integration, 781  
     processing, 969, 972  
     structure theory, 424–426  
     topography theory, 426–427  
 conceptualization, 21–22  
 concrete words, 675  
 concreteness, 29–30  
 concreteness effect, 671, 674–675  
 Conjunctive neurons, 426–427

- connectionism, 587, 812–814, 816–817  
 Connectionism and semantic memory  
   Distributed representation models, 418–419, 426  
   Localist models, 418, 439  
 connectionist model, 286–287, 291, 306, 346,  
   588–590, 597, 600, 1032  
 connectionist-based account, 812–814, 816–817  
 Consciousness  
   Anoetic, 409–410  
   Autoanoetic, 409–410  
   Noetic, 409–410  
 conservative learning, 1084  
 Consistency Effects, 300, 302–305, 309, 319  
 consonant harmony, 102–103  
 constraint satisfaction, 581, 583, 585, 587, 589,  
   591, 593, 595, 597, 599, 601–602  
 Constraint satisfaction model, 844–845  
 constraint-based theories, 470, 474  
 constraints, 98–105, 1095  
 Construction grammars as semantic memory  
   models, 415, 436  
 Construction-Integration model, 766, 776  
 content word, 380–381, 390, 672  
 context, 1080, 1081, 1091–1093, 1096, 1099, 1093,  
   1094, 1105, 582, 584–585, 588–589, 590, 592,  
   593, 595  
 context effects, 377, 379, 381, 383, 385–391,  
   393–394  
 context updating, 700  
 contextual constraint, 672, 674, 679, 682, 685–686  
 contingent negative variation (CNV), 673  
 continuity, 1036, 1055, 1076, 1100, 1102, 1105  
 contraction, 1082–1083, 1086  
 Contrast, 864, 866, 875–876, 885, 890–891  
 Contrastive gesture, 217, 219  
 Contrastive phonemes, 974  
 controlled attention, 809, 819, 825  
 controlled process, 695, 701  
 conventional metaphors, 781  
 Convergence zones, 426–427  
 Conversation, 864–865, 887–888, 891–892  
 Conversational implicature, 838  
 conversational precedents, 923–924, 928  
 Coordination, 202, 237  
 corpus, 583, 587, 596–597, 599  
 Corpus studies, 549, 556, 561  
 cortical activity, 765  
 covert responses, 132  
 cross linguistic generalizations, 1086, 1095, 1105  
 Cross-language competition, 976, 988  
   comprehension, 987  
   neighbors, 976  
   parsing, 983  
 cross-modal lexical priming (CMLP), 954  
 cross-modal naming, 515–516, 518  
 Cycle, 958  
 Data-driven processing, 988  
 dative, 65–68, 78–79  
 deaf, 1001–1002, 1006–1008, 1013–1014, 1016,  
   1018  
 Decoding, 1112, 1118–1120, 1127–1128  
 Deep Dyslexia, 298  
 Degradation  
   Graceful degradation, 418–419  
   Partial degradation, 428–429, 435, 438  
 Demonstratives, 887  
 dependency, filler-gap, 692, 702–703  
 dependency, long-distance, 692, 701–702  
 Depictive, 539, 549  
 Derivational theory of complexity, 6  
 Development of  
   Lexical processing, 871, 874, 890  
   Pronoun resolution, 864, 886–887  
   Reference, 864, 866, 877, 880, 883, 886  
   Sentence processing, 874, 880, 883  
   Word recognition, 866, 870, 877, 881, 890  
 Development of comprehension abilities, 889–891  
 Developmental dyslexia, 1125–1126, 1128,  
   1130–1131  
 Developmental dyslexia subtypes, 1125–1126,  
   1128, 1131  
 Dialect, 219, 221, 223–224, 226, 230, 232  
 Dialogue, 8–11, 13, 887–889  
 Dialogue corpora, 10  
 Diary studies, 4  
 Different script languages, 988  
 Direct realism, 216, 218  
 disambiguation, 696, 706  
 disambiguation (prosodic, syntactic), 911–912  
 Discourse, 384, 387, 389–394, 539–540, 550, 564,  
   667, 678, 680–684, 696–697, 727, 731, 733,  
   738  
 discourse context, 486  
   model, 801–802, 807–808, 822–825  
   processing, 765–769, 772–773, 776, 779,  
   781–783, 785–790, 792–793  
   referent, 696  
   representation, 818  
   status (given/new), 886  
   structure, 513  
 discourse-level processes, 803–804, 811, 813,  
   815–816, 819, 821  
 disfluencies, 72, 80, 864  
 disjunction, 1087–1100, 1102–1105  
 dissociation, 704–706, 948, 958  
 distributed representations, 254  
 Distributivity, 572

- domain expertise, 815–816, 821  
   general, 700  
   knowledge, 816, 820, 822, 824  
   specificity, 672, 684, 687  
   specific hypothesis, 423–424  
 dorsolateral prefrontal cortex, 767, 778–779, 781, 788  
 double-object, 66, 78–79  
 duration, 505, 508, 510–511, 515, 518, 521  
 Dutch, 681, 683–684, 699, 704  
 dysarthria, 114, 116–118  
 dyslexia, 184–185, 1145
- echo-planar imaging, 127, 129, 143  
 effects, 621, 639, 641  
 efficiency, 1036, 1039, 1047, 1052–1054  
 egocentrism, 918  
 ELAN or early left anterior negativity, 691–693, 695–697, 700, 704–705  
 ELIZA (Weizenbaum), 7  
 Embedded words, 881  
 Embodied Cognition, 747  
 Embodiment in semantic memory theories, 415  
 Emergent literacy, 1123, 1130  
 emotional knowledge, 773  
 empathy, 773  
 empty categories, 62, 70  
 encapsulation, 456, 469, 690, 693, 704, 707  
 encephalogram (EEG), 659  
 English, 676–677, 681, 684, 700–705  
 enhancement effect, 165–166  
 Entailment, 545, 1093, 1097, 1104  
 environmental sounds, 669, 699  
 epilepsy, 1016  
 Episodic vs semantic memory, 410, 437  
 ERPs, 659, 661–663, 665–667, 669–671, 673–674, 676–680, 682, 684–686, 689–693, 696–697, 700–701, 704–708, 984, 986  
 Event, 571  
   vs. Adjunct, 547, 548  
   Advantage, 548, 550  
   Argument, 543–544  
   Structure, 543, 544  
 event-related brain potential (ERP), 659  
   designs, 134–135, 143  
   fMRI, 339, 770  
 Executive function, 989–990  
 exemplar theory, 272  
 exhaustive access, 382–385  
 Expectations, 975, 979, 982, 986, 988  
 exposure, 75  
 extended optional infinitive (EOI), 1150  
 external argument, 945
- eye movements, 616, 631, 635  
   control, 634, 640, 642  
   Data analysis, 865  
   during reading, 872  
   in language production, 867  
   in natural tasks, 864  
   in reading, 5  
   Linking assumptions, 865  
   measures of processing time, 616  
   monitoring, 517, 526  
   Task variables, 867  
   word recognition, 625  
 Eye-tracking, 471, 553, 560–561, 567–568, 572, 979
- facial expression, 1007  
 factor analysis, 950–951  
 familiarity, 378–380, 822–824  
 Featural Analyses, 287  
 Features  
   vs Attributes, 417, 423  
   Correlated, 425, 438  
   Functional, 429, 440  
   Visual, 429, 440  
 feedback, 261–263  
 Feedbackward Consistency Effects, 303  
 Feedforward Consistency Effects, 303, 309  
 field potential, 660  
 Figurative interpretation, 541  
 figurative language, 686, 689, 835–856  
 figurative language processing, 779, 782  
 Figurative Meaning, 835–843, 845–846, 849  
 filler retrieval, 669, 671, 687, 702  
 Finnish, 886–887  
 First pass reading times, 876  
 Fixations, 5, 628  
 Flexible word-order languages, 886  
 Fø (average phonatory frequency), 227, 229, 232, 235  
 focus, 506–507, 509–510, 524–525, 527  
 focus operators, 1096, 1105  
 Formant, 205, 207, 214, 227, 229–230, 235  
 frequency, 258–260, 263–266, 268, 462, 581, 586, 591, 593, 598–599, 602  
   Structural, 548  
 Frequency effects, 864  
 Freudian Slip, 4  
 fricative, 96, 100, 103  
 frozen metaphor, 779–781, 786, 789  
 function words, 380–381, 672  
 Functional application, 542, 544–545, 547–548, 550, 555, 570  
   magnetic resonance imaging (fMRI), 126, 339, 341–343, 769–773, 777–778, 1017–1018  
   misarticulators, 95–96

- neuroimaging, 125, 127, 129, 131–135, 137, 139,  
 141, 143  
 processing, 63  
 fundamental frequency, 505, 508, 510–511
- gamma band, 681  
 gap-filling, 702–703  
 gapping, 704  
 gaps, 68–70, 83  
 garden path, 692  
 garden-path effects, 476  
   sentences, 810  
   theory, 461, 466, 470  
 gating, 677–678  
 gender, 683–685, 687, 692, 694  
 general cognitive ability, 820  
 generalization, 100–104  
 Generalized implicature, 885  
 generative grammar, 62–63, 68–70  
   linguistics, 9  
   semantics, 7  
 Generic semantic memory, 408–411  
 Genre, 729  
 German, 673, 697–698, 702–704, 706–707  
 germanic languages, 1154, 1157, 1164  
 glides, 101  
 Good old-fashioned AI (GOF AI), 7  
 Graded Saliency Hypothesis, 766, 787, 792  
 Graded saliency model, 844  
 gradient noise, 127–129, 133–134, 136, 143  
 grammar, 2, 1028–1031, 1033, 1043, 1047, 1051,  
 1054–1056  
 grammar development, 1055  
 grammatical encoding, 62–64, 66–70, 72–73, 75,  
 78, 82–83  
 grammatical gender, 683–684  
 grammaticality, 478, 674, 691, 705  
 grammaticality judgment, 947–948, 950, 956–957
- head motion, 129–131, 134, 137, 143  
 hemispheric asymmetry, 686  
 hemodynamic response, 134–135, 143, 769–770  
 heuristic, 940–942, 945, 947  
 Hidden competitors, 879  
 high-amplitude sucking, 1037  
 higher order effects, 626–627  
 Hyperbole, 846  
 hyponymy, 700
- Ideas, 1–4, 6, 8  
 Idiolect, 219, 221, 223, 229–230, 232, 234, 238  
 Idioms, 838, 846–847, 850  
 imageability, 30, 33  
 immediacy, 685
- Implicit naming, 878–879  
 incrementality, 76–78, 82–83, 455, 685  
 Indeterminacy, 836, 847  
 individual differences, 802, 808–810, 812–813,  
 817, 819–820, 822, 824–825, 1033,  
 1045–1047, 1056  
 infants, 659, 671  
 inference generation, 776, 779, 788, 791  
 Inferences, 726, 729, 735–736, 771–772, 776–779,  
 783, 785, 790, 804–805, 811, 813, 815–816,  
 819, 821, 823  
 inferior frontal gyrus, 766–768, 775, 778–781, 788  
 inflectional morphology, 588, 590, 1033  
 information structure, 506, 524–525  
 Information theory, 5–6, 158  
 Inhibition, 413–414, 418, 435–436  
 innate, 1079, 1089  
 Instructions, 975, 978, 982, 988  
 integration, 677–678, 680, 685, 687, 689, 693,  
 701–703  
 integration cost, 481–482  
 integration of information across saccades, 616  
 interaction, 674, 690, 693–696, 707  
 Interactions between language and vision, 864  
 Interactive accounts, 461  
   Activation Model (IA), 292–293, 297, 977  
   alignment, 929  
   model, 263  
 interleaved technique, 135–137, 143  
 Interlingual homographs, 976  
 intermediate phrase, 511–512, 514, 517–518, 520,  
 529–530  
 internal argument, 945  
 Interpretation, 986–987  
 Interpretation-contingent analyses, 871  
 Intersensory  
   Combination, 208–210  
   Competition, 210  
 intonation phrase, 511–512, 514, 516–518, 520,  
 523–524, 529–530  
 intonation, 81, 505, 507, 511–512, 514, 516–518,  
 520, 523–526, 529–530  
 intracranial recording, 670  
 Intraparietal Sulcus Spatial network, 789  
 inverse problem, 155–158  
 IQ, 1145  
 Irony, 837–839, 841–842, 845–846, 849–850,  
 852–856  
 Isomorphism, 218–221  
 item-priming-in-recognition, 805  
 IViE, 511
- Japanese, 691, 703  
 jaw movement, 130, 132, 137, 143

- joke processing, 687  
 Junggrammatiker, 3
- kernel sentence, 69  
 ko-tätaste, 45
- L2 proficiency, 971–972  
 lack of invariance, 154  
 LAN or left anterior negativity, 661  
 Landscape architecture, 766  
 language, 125–130, 132–133, 136–137, 740  
   Acquisition Device, 1027  
   acquisition, 4, 6  
   as action, 856  
   as product, 856  
   comprehension, 1013  
   nonselectivity, 968, 978, 988  
   processing, 487  
   production, 19, 21–23, 25, 27, 29, 31, 33, 35, 37,  
     39, 41, 43–45, 525, 1007–1008, 1013,  
     1016–1017  
   selectivity, 981  
   switching, 986  
   use, 901, 904–906  
 Language-as-action, 865, 881  
 Language-as-product, 865, 881  
 laryngeal devoicing gesture (LDG), 110  
 Late closure, 548  
 Latent Semantic Analysis, 561  
 lateral asymmetry, 669  
 lateral inhibition, 28  
 lateralization, 768, 782, 784–785  
 left hemisphere, 666, 669–670, 683, 687, 694, 766,  
   773, 777–779, 781–785, 788–789, 792  
 left hemisphere lesions, 670  
 lemma, 22, 33, 64–65, 67, 82, 84  
 lesion, 959–961  
 Letter Recognition, 285–291, 293, 298  
 lexical access, 251, 256, 263, 272–273, 772, 788,  
   792  
 lexical ambiguity, 581–582, 584–586, 590, 631,  
   676, 875–876  
 Lexical Decision, 293, 297–298, 303–304, 307,  
   310–312, 314–322, 324–329, 331–332,  
   334–336, 339, 342–346, 412, 414, 425,  
   433–434, 548, 566, 974–976, 979–981  
   entry, 22  
   factors, 671  
   neighborhoods, 878  
   priming, 65–66  
   priming, 985  
   processing, 377, 379–383, 385, 387, 389, 391,  
     393–394, 976, 982–983  
   processing negativity (LPN), 667  
     representation, 22, 25, 28–29, 31–32, 37, 40  
     functional grammar, 63  
 lexicon, 582, 584–585  
 lexico-pragmatic route, 943, 961  
 linear additivity, 128, 136  
 linguistic competence, 1030–1032, 1056  
   constraints, 1042, 1049–1051  
   Convergence, 983  
   division of labor, 8  
   indeterminacy, 1050  
   universals, 1074, 1095  
 Linguistics, 1–5, 7, 9  
 Linking assumptions  
 listening, 487  
 Literal language, 836–837, 855  
 Literal meaning, 836–842, 851, 855  
 Logical problem of language acquisition, 6  
 logic, 2, 1087, 1094–1095  
 Lombard sign, 234  
 long-term memory, 773, 778  
 long-term working memory, 812, 815  
 Luce choice rule, 870
- M350, 563  
 Magnetoencephalography (MEG), 563, 663  
 Map Task, 237  
 mapping, 773–776, 781  
 markedness constraints, 99–100, 102, 104  
 Masked Priming Studies, 323  
 Masked priming, 980  
 mature readers, 801–803, 805, 807, 809, 811, 813,  
   815, 817, 819, 821, 823  
 mean length of utterance (MLU), 1147  
 Meaning, 835–847, 851, 855–856, 1094  
 meaning dominance, 382–385  
 Meaning-Knowledge-Units, 785, 788  
 medial frontal cortex, 775–776, 781, 789  
 Medial frontal protagonist/agent interpreter  
   network, 789  
 Megastudies, 345  
 memory load, 480  
 memory resources, 480–481  
 Mental model theory, 8  
 mental models, 782  
 Mental Representation, 726–727, 730, 735–738,  
   747–748  
 Message Planning, 22, 41–42  
 Metaphor, 687–689, 837–839, 841, 846, 848–852,  
   854, 856  
 metaphor comprehension, 768, 779–780  
 Metonymy, 838, 850, 852, 856  
 micro-structure, 772  
 Minimal attachment, 548  
 Minimal inferences, 11

- misanalysis, 474, 477–479, 485, 488  
 mixed error, 41  
 modeling eye movements., 634  
 Modification, 116, 127, 267, 517, 544–549, 558, 883, 1152  
     Intersective vs. nonintersective, 545, 549  
 modularity, 456, 690, 704  
 module, 1030  
 monitoring, 902, 906, 909–910, 913  
 Monologue, 8–11  
 morpheme, 153, 378, 1147–1150, 1153–1154, 1157, 1162, 1164  
 morphological, 617, 621, 625–627  
     Analyses, 310–311  
     Effect, 311  
 morphology, 377–378, 1008–1009  
 morphosyntactic variability, 1147  
 morphosyntax, morphosyntactic, 694, 696, 705  
 morphosyntax, 1147–1148  
 motor development, 95  
 Motor theory, 216–217  
 movement, 62, 66, 68–70, 83  
 multilevel modeling, 807, 825  
 muscular control, 110  
 music, 700  
  
 N400, 541, 563, 664, 666–690, 692–699, 705–707, 986  
 N400-700, 664, 666–690, 692–700, 705–707  
 Naming, 289, 293, 297–298, 300, 302–303, 309, 311–312, 314–322, 324–326, 331–332, 334–337, 339, 342–345  
 narrative text, 775, 784, 787, 789  
 nativism, 1075  
 nativist approach, 1027  
 nature versus nurture, 1073–1074  
 nearest neighbor interpolation, 131  
 negation, 679, 682  
 negative evidence, 1028–1030  
     polarity items, 1085–1086, 1089, 1093  
     polarity, 692, 700  
 neighborhood, 1002–1003, 1005  
 networks, 587, 590, 595–596, 765–767, 769, 771, 775, 778–779, 781, 787–790, 792  
 Neurocognitive, 787  
 neuroimaging, 125, 127, 129, 131–135, 137, 139, 141, 143, 765–769, 771, 773, 775, 777, 779, 781–787, 789, 791–792, 1001, 1013, 1017  
 neuronal activity, 125, 134–135  
 Neurons, specialized  
     Invariant visual representation, 438  
     Mirror, 427  
 neuropsychological, 767–769, 777, 782, 787  
 Neuroscience, 10, 13  
  
 (non-)preferred structure, 693  
 non-linguistic violations, 664, 690, 699, 707  
 nonliteral language, 670, 686  
 nonwords, 664–666, 671  
 novel metaphors, 779–781, 786, 789  
 Null context, 874  
 number, 660–662, 664, 667, 677, 690–694, 698–699, 703, 707  
  
 Object affordances, 880, 885  
 Odawa, 77  
 oddball task, 665  
 on-line processing, 954, 956–957  
 Ontology, 543, 570  
 open class, 667, 672–673, 680, 685  
 operation span, 808–809, 812  
 optimality theory (OT), 96  
 optional period, 1150–1151, 1154, 1157, 1162  
 optionality, 1152, 1154–1157, 1164  
 Organization  
     Perceptual, 203–205, 207–211, 224  
     Unimodal, 211  
 Orthographic Neighborhood Effects, 293, 317  
     neighborhood, 675  
     Priming, 323–324, 326  
     processing, 664–665, 672, 693  
 orthography, 377  
 overextensions, 1045  
 overt responses, 126, 132–133  
  
 P300, 661, 700  
 P3b, 676, 693, 699–700, 707  
 P600, 674, 684, 691–707, 984  
 parafoveal preview, 379  
 Parallel and Serial Processing, 289  
 Parallel Distributed Processing Models, 299  
 Parallel Networks of Discourse, 765–767  
 paraphasia, 1014–1015  
 parenchymal brain motion, 130  
 Parity, 218–221, 232  
 parser, 582, 591–592  
 parsing and syntactic ambiguity, 628  
 parsing, multi-stage, 693  
 parsing, serial, 692, 704, 706  
 passages, 768, 771–772, 774–775, 778–779, 781, 784  
 passives, 62, 64–65, 68–69, 74–77, 79, 944, 946–949  
 Pathfinder networks (PFnets), 430, 432–433  
 Perception  
     Audiovisual, 208, 210, 221  
     Bistable, 214  
     Multimodal, 208, 210  
 Perceptual integration, 210  
     learning, 1038  
     span, 616, 627



- Perspective, 744–745  
 perspective monitoring, 766  
 perspective-taking, 902  
 Pet Scanning, 339  
 phoneme, 175, 184, 211–218, 221–223, 230, 231, 236  
 Phoneme monitoring, 978  
 phoneme-discrimination, 96  
 phonemic errors, 107, 957  
 Phonesthesia, 221  
 phonetics, 505, 507–508, 513  
   convergence, 238  
   encoding, 22, 38  
   segment, 153, 183–184  
   similarity, 974  
   transcription, 512  
   unit, 154, 183  
 phonological activation, 31  
   awareness, 1006, 1114–1123, 1127, 1130  
   coding, 622, 624  
   competition, 40  
   Dyslexia, 298, 304  
   encoding, 63, 78  
   loop, 1160  
   mismatch negativity or PMMN, 677  
   Priming, 325–326  
   processes, 97–100, 102  
   processing, 1117–1118, 1128, 1131  
   training, 1118–1119  
 phonology, 377, 1015  
 phonotactic regularities, 1040  
 phrase accent, 511–512, 514  
 Phrase structure grammar, 7  
 Picture noun phrases, 872  
 picture-word interference, 24, 26, 38, 43  
 pitch accent, 511–513, 524–528  
 Plausibility, 486, 548, 562, 586, 593, 597, 621, 626, 683, 694, 699, 707  
 Point-of disambiguation, 866, 886, 889  
 Polysemy, 844  
 Port Royal Grammar, 2  
 positional processing, 63  
 positive evidence, 1029  
 positron emission tomography (PET), 126, 768  
 poverty of the stimulus, 1029–1030  
 PP attachment, 547  
 PP-attachment ambiguity, 884  
 Pragmatics, 11, 540–541, 550, 564, 772, 848  
 preaspiration, 114  
 Predicate abstraction, 546–547  
 Predicate modification, 544–545  
 predictability effects, 623, 627  
 prediction, 684–685, 696  
 predictive inferences, 771  
 Pre-nominal adjectives, 866, 886  
 prevalence, 1144–1145  
 primary auditory cortex (AI), 169, 186  
 priming, 377, 379, 382, 385, 388–390, 805–806, 1001  
 Priming and alignment, 888  
 Principal Components Analysis, 173  
 print exposure, 802–804, 807–808, 811, 813–816, 819–821, 824  
 PRO, 945, 952  
 probabilistic constraints, 583, 586–587, 589–591, 593, 595, 597, 600  
 probability, 666–667, 679, 681–682, 684, 686, 688–689, 693, 700, 704  
 Processing, 836–837, 839–850, 852–855  
   difficulty, 474  
   load measures, 874  
   resources, 1157–1158, 1161  
 Production-Distribution-Comprehension account, 602  
 prominence, 506, 508, 510–511, 524, 526, 532  
 Pronoun resolution, 864  
 Pronouns, 864, 872, 886–887  
 Proportion of fixation curves, 868–869  
 propositional content, 939–940  
 propositions, 776–777, 786, 790  
 Proprioception, 202, 216  
   Prosodic bootstrapping, 1039  
   boundary, 507, 514, 516, 518, 523–524, 527, 529, 531  
   phrasing, 506–509, 513–520, 522, 524, 527–528, 530–531  
   word, 43  
 prosody, 62, 80–82, 84, 469, 505–509, 511–519, 521–525, 527–532, 881–883  
 Prosody and word recognition, 881–883  
 protagonist, 765–766, 775–776, 781–782, 785, 787, 789  
 Prototypes, 416  
 Proverbs, 838, 841, 845, 847, 850, 855  
 pseudowords, 664–667, 671–672, 693  
 Psychoacoustic(s), 201, 213  
   Benchmarks, 223  
   Effects, 215, 229  
   Models, 225  
 Psycholinguistics, 476, 480, 1002  
 psychological predicate, 706  
  
 Quantifier  
   Ambiguity, 552–553  
   Interpretation, 550  
   Raising, 551  
   Scope, 572

- Reaction Time Distribution Analyses, 345  
 reactivation, filler, 703  
 Reading, 5, 8–9, 11–13, 477, 613–616, 621, 626,  
 630, 633–635, 640–642, 729, 771, 777–778,  
 780–782, 784, 787, 789, 793, 801–803,  
 ability, 789, 816, 825, 807–810, 812–814,  
 816–817, 819–822, 825,  
 comprehension, 801–802, 808–810, 817,  
 819–820, 1120–1122, 1127, 1130  
 development, 1124  
 disability, 1125, 1129  
 impairment, 1125–1127  
 span, 701–702, 808, 810, 812, 819  
 Reafference, 234–237  
 reanalysis, 475, 477, 486, 687, 693, 698  
 recency, 463  
 recollection, 822–824  
 recurrent network model, 1033  
 reduced relative, 592–595, 598  
 reference, 903, 905  
 Referential domains, 864, 880–881, 884–885, 889  
 principle, 883  
 theory, 469  
 Referring expressions, 869, 882, 884, 886–887, 889  
 Reflexive pronouns, 872, 887, 947  
 regional cerebral blood flow (rCBF), 126  
 regions of interest (ROI), 959  
 registration algorithms, 131  
 regressions, 615, 620, 628–631, 636  
 Regularity Effects, 292  
 related anomaly, 683  
 relative clauses, 696, 701, 705–706, 944, 947, 955  
 Relevance theory, 848  
 repair, 677, 693, 705  
 repeated reference, 903  
 repetition effect, 667, 671–672, 674  
 repetition time (TR), 135  
 Representational measures, 874  
 residual errors, 95  
 Resonance model, 766, 791  
 resource reduction, 950, 952  
 Resultative, 539, 549  
 Revised Hierarchical Model (RHM), 971  
 revision, 685, 705  
 rhythm, 505, 511  
 right hemisphere, 766–768, 770, 773, 775–785,  
 788–792  
 right hemisphere lesions, 669–670, 686–687, 700  
 rigid-body motion, 130–131  
 rotational motion, 130  
  
 saccades, 5, 616, 631, 638  
 Sapir-Whorf hypothesis, 4–5  
 Scalar adjectives, 866, 886  
 Scalar implicatures, 885–886  
 schema, 788  
 scrambling, 703–704  
 Scripts, 733–734  
 Second language acquisition, 983  
 second-language perception, 181  
 segmentation, 1040, 1042, 1047  
 selective access, 382–386  
 Self-paced reading, 548–549, 560, 567–568  
 Self-regulation, 233–234, 236, 238  
 Semantic anomaly, 693–694  
   Automatic processes in, 413, 430, 433–435  
   Binding problem, 419, 438  
   categorization, 980  
   composition vs. World-knowledge integration, 541  
   congruity, 693  
   Controlled processes in, 413–414, 430, 433–436  
   Deficits/impairments, 403, 420–422, 429  
   Dementia (SD), 311, 420, 422, 670  
   development, 1100, 1102, 1105  
   Entries and bindings, 405, 406, 410, 418  
   features, 26, 32  
   Generic, 403, 408–411  
   Hyperpriming in AD, 433–435  
   Information Processing, 7  
   Memory Bindings and entries, 405–406, 410  
   Mismatch, 550, 563  
   Networks in, 413, 430–432, 437  
   networks, 7  
   Organization, in AD and aging, 403, 430  
   Priming, 322–323, 325, 327–334, 336–337,  
   345–346, 979–980  
   processing, 659, 666–667, 675–677, 685, 690,  
   Quillian's model, 403, 405–407, 409, 41, 436  
   Shifting, 551, 555, 569  
   Stimulus onset asynchrony (SOA), 414, 433–435  
   Type, 695, 697, 699  
   violation, 680–681, 690, 693–697, 700, 704  
   semantics, 461, 464, 765, 768, 775, 779, 782–783,  
   785, 787–788, 790–792  
 Semaphore, 215  
 sensory adaptation, 165  
   contrast, 161  
   functional theory, 421–423  
   motor theory, 423  
   constraint, 981–982  
   parsing, 866, 890, 984–985  
   processing, 465–466, 469, 473, 486–487, 507,  
   808, 814  
   production,  
   level processes, 804, 807, 811, 813, 815–816,  
   819, 821  
 separate-sentence-interpretation-resources theory,  
 809

- Shallow interpretation, 540  
 Shrdlu (Winograd), 7  
 Sign language, 1002, 1013  
 silent responses, 133  
 similarity, 252–254, 256–258, 263, 266  
 Similarity judgments, 403, 417, 430, 433, 436  
 Similarity-in-topography (SIT) principle, 427, 439  
 sinc interpolation, 131  
 Sinewave, 205, 207–208, 229–230  
 single photon emission computed tomography (SPECT), 126  
 single resource model (of working memory), 701  
 Situation model, 11, 744, 772–773, 775, 780, 782, 784–785, 788–789  
 slowing hypothesis, 1158–1159  
 smallworld semantic networks, 322  
 social cognition, 1030, 1034, 1048  
 Sortal shift, 71  
 span, listening, 701  
 span, reading, 680, 702  
 Spanish, 683–684  
 sparse temporal sampling, 134  
 spatial imagery network, 765  
 speaking, 21–22, 32, 38, 42–44  
 speech perception, 1028, 1036, 1038–1039, 1041  
 specific language impairment (SLI), 177, 1143  
 spectral contrast speech segmentation, 166  
 Spectrum, 203–205, 208  
   Long-term, 229, 232, 235  
   Short-term, 205, 208, 229  
 speech, 125–143  
 speech delay, 94–101, 103–105  
 speech disorders, 93–95, 97, 99, 101, 103, 105–107, 109, 111–115, 117–118  
 Speech Errors, 4, 23–25, 28, 32, 34–36, 40–41, 63–64, 66, 93, 105  
 speech responses, 127, 130, 132–133, 138, 143  
 Speed-accuracy tradeoff (SAT), 562  
 spelling, 700  
 spelling to sound translation, 295  
 spillover of processing, 766, 782, 785, 790  
 spoken language comprehension, 508, 519  
 spoken word recognition, 667, 669  
   Cohort effects, 870  
   Frequency, 864  
 spontaneous speech, 519–520, 522–523  
 Sprachpsychologie, 3  
 statistical learning, 1074–1075, 1085–1086  
 stimulus masking, 128–129  
 stop consonants, 93, 95, 97–98, 106–110, 116–117  
 Stop making sense task, 566  
 stops, 97, 101, 103–106, 108–118  
 storage, 701–703  
 Storage vs access deficits, 428–430, 440  
 stress, 505–506  
   complexity, 480, 486  
   priming, 985  
   structural alignment in semantic memory, 416–417, 432–433, 436, 440  
 Structuralism, 2  
 structure building framework, 766, 774, 817  
 structure building, phrase, 691–692  
 structure dependence, 1078  
 subcategorization, 694, 696  
 Sublexical processing, 973  
 subordinate bias effect (SBE), 383  
 Subordinate objects in semantic memory, 416, 429  
 Sub-phonetic information, 881  
 summation, 661, 694  
 superior temporal sulcus (STS), 186  
 Superordinate objects in semantic memory, 416, 429  
 Superset binding, 406, 408, 411  
 suppression, 165  
 suppression ability, 802, 804, 809, 811, 813, 815–817, 819, 821  
 Surface developmental dyslexia, 1128  
 surface dyslexia, 298, 304  
 susceptibility artifacts, 130, 132, 136–137  
 syllables, 1034, 1038, 1040  
 Symbol grounding problem, 12  
   Affordances, 885  
   complexity, 73–74, 82, 959–960  
   integration, 693, 701–703  
   predictions, 481  
   priming, 65, 68, 70–71, 74, 83, 331–332, 985–986  
   processing, 772, 780, 784, 788  
   Referential context, 876  
   syntactic ambiguity, 507–508, 514, 516, 523–524  
 “Syntactic Structures” (Chomsky), 6  
   Verb preferences, 876  
 syntax, 780, 1009  
 Tachistoscope, 5  
 Talker identification, 224–227, 229, 230  
 Task schema system, 988  
 task-specific effects, 324  
 taxonomic assumption, 1049  
 temporal gyrus, 766–767, 780, 789  
 Temporally ambiguous sentences, 984  
 tense, 1147–1157, 1159, 1161–1162, 1164–1165  
 text comprehension, 815–816, 821–822, 824  
 text integration, 765, 772–773, 775–776, 779, 782, 785, 788  
 Thematic relations, 543  
 Thematic role assignment, 873

- thematic roles, 594, 596, 692, 698–699, 703, 939, 941–945, 949, 951, 956  
 theme-experiencer verbs, 79  
 theoretical framework, 766, 787, 790  
 theory of mind, 1035  
 Theory of Mind, 766, 781, 785–786, 789–790  
 Theory of word meaning, 7  
 Thesaurus method, 405  
 time criterion models, 337  
 tip-of-the-tongue, 33  
 ToBI, 510–512, 514–515, 517–518, 523  
 Top-down processing, 978, 988  
 topicalization, 702, 705  
 topic-related inferences, 804–805  
 TR, 770  
 trace deletion hypothesis, 944, 946, 957  
 TRACE model, 870  
 traces, 62, 68–70, 83, 944–946, 955, 957  
 trading relation, 170, 183  
 Translation, 969–973, 975, 978, 980, 986–987  
   asymmetry, 971  
   direction, 971–972  
   equivalents, 969, 971, 980  
 translational motion, 130  
 Translators, 987  
 treatment effects, 100, 103  
 tree-adjointing grammar, 71  
 Truth Value Judgment task, 1104  
 tune, 509, 511  
 two-stage accounts, 457  
   approach, 582–583  
   model, 64, 66–67  
  
 unbounded dependencies, 458  
 Underspecification model, 843–844  
 Understatement, 846–847, 850  
 ungrammatical, 674, 691, 705–707  
 Unification in post-Quillian semantic  
   memory, 436  
 Universal grammar, 2, 1089  
 universal quantification, 1085, 1086, 1088, 1091–1095, 1098  
  
 Ventromedial prefrontal cortex, 563, 569  
 Verb argument preferences, 883–884  
 “Verbal Behavior” (Skinner), 4  
 verbal efficiency theory, 803–804, 807–808, 814, 821  
   illusion, 540  
   learning, 6  
 violation, 684, 692, 694–697, 699, 704–707  
   combined, 694–696, 705  
   double, 694, 696–697, 705  
   morphosyntactic, 690, 692–694, 699, 701, 704  
   syntactic, 670, 687, 693, 695, 700, 707  
   word category, 691–692, 695–696, 704–705  
 visual features, 377  
   Action-based studies, 873–874  
   Data analysis, 868  
   half field, 686  
   hemifield, 687  
   imagery, 773, 781  
   Linking assumptions, 868  
   “Passive listening”, 867, 873  
   Task variables, 872  
   word recognition, 659, 664  
   word recognition, 968, 975, 977–978  
   world paradigm, 550  
 visual-world method., 473  
 Vocabulary, 1116–1118, 1120–1122, 1124, 1127, 1130  
 vocabulary spurt, 1042  
 Voice, 201, 205, 209, 220, 224, 227, 234, 237  
   Quality, 224, 225, 227, 229–231  
 voice-onset time (VOT), 93, 871  
   Recognition, 224  
 Voiceprint, 224, 226  
 VOTs, 979  
 Vowel, 215, 227, 230, 232, 235  
   Advancement, 215  
   Height, 215  
 voxel, 959–960  
  
 weaver++, 26–28, 31, 33, 38, 40–41  
 Wernicke, 941, 957–958, 960  
 Wernicke’s area, 767, 782, 785, 1014–1015  
 wh-phrase, 703  
 wh-question, 700–703  
 Word Association Model, 969–970  
   class, 673  
   familiarity, 315  
   frequency effects, 619, 621–622, 627, 637, 640  
   frequency effects, 312, 314–315  
   frequency, 666, 674  
   identification, 801–804, 807–808, 820  
   length effects, 311  
 word order, 459, 472–473, 697–698, 703, 706  
   canonical, 703  
   sov, 706  
   verb-final, 698  
 word position, 673–674  
   production, 21, 23, 25–26, 28–29, 31–32, 34, 36–37, 39–41, 44–45  
   recognition in bilinguals, 881  
   recognition, 1002, 1005, 1127  
   recognition, 659, 664, 671–672  
   retrieval, 33  
   superiority effect, 290, 292, 319

- word-decoding ability, 807, 820
- word-identification skill, 801–803, 808
- word-level ability, 802–804, 808, 812, 814, 821, 825
- word-level processes, 803–804, 807–808, 811, 813, 815–816, 819, 821
- working memory, 680, 686–687, 689–690, 692, 700–704, 707, 773, 778, 784–785, 790, 802, 809, 812, 815, 819, 825, 943, 1011, 1158–1161
  - limitations, 810, 812
  - resources, 982, 984, 987
- working-memory capacity, 802–803, 808–812, 814, 819
- working-memory span, 807
- world knowledge, 681, 768, 772, 775, 788, 790–791
- Wundt's lab (Leipzig), 2
- Yes/No questions, 1075–1077