The Handbook of Psycholinguistics

Edited by Eva M. Fernández and Helen Smith Cairns

WILEY Blackwell
The Handbook of Psycholinguistics
This outstanding multi-volume series covers all the major subdisciplines within linguistics today to offer a comprehensive survey of linguistics as a whole.

To see the full list of titles available in the series please visit www.wiley.com/go/linguistics-handbooks

The Handbook of Contemporary Syntactic Theory
Edited by Mark Baltin and Chris Collins

The Handbook of Historical Linguistics
Edited by Brian D. Joseph and Richard D. Janda

The Handbook of Second Language Acquisition
Edited by Catherine J. Doughty and Michael H. Long

The Handbook of Applied Linguistics
Edited by Alan Davies and Catherine Elder

The Handbook of Pragmatics
Edited by Laurence R. Horn and Gregory Ward

The Handbook of Speech Perception
Edited by David B. Pisoni and Robert E. Reeme

The Handbook of the History of English
Edited by Anes van Kemenade and Bettelou Los

The Handbook of English Linguistics
Edited by Bas Aarts and April McMahon

The Handbook of World Englishes
Edited by Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson

The Handbook of Educational Linguistics
Edited by Bernard Spolsky and Francis M. Hult

The Handbook of Clinical Linguistics
Edited by Martin J. Ball, Michael R. Perkins, Nicole Müller, and Sara Howard

The Handbook of Pidgin and Creole Studies
Edited by Silvia Kouwenberg and John Victor Singler

The Handbook of Language Teaching
Edited by Michael H. Long and Catherine J. Doughty

The Handbook of Phonetic Sciences, Second Edition
Edited by William J. Hardcastle and John Laver

The Handbook of Language and Speech Disorders
Edited by Jack S. Damico, Nicole Müller, and Martin J. Ball

The Handbook of Language Contact
Edited by Raymond Hickey

The Handbook of Computational Linguistics and Natural Language Processing
Edited by Alexander Clark, Chris Fox, and Shalom Lappin

The Handbook of Language and Globalization
Edited by Nikolas Coupland

The Handbook of Hispanic Sociolinguistics
Edited by Manuel Díaz-Campos

The Handbook of Language Socialization
Edited by Alessandro Duranti, Elinor Ochs, and Bambi B. Schieffelin

The Handbook of Phonological Theory, Second Edition
Edited by John A. Goldsmith, Jason Riggle, and Alan C. L. Yu

The Handbook of Intercultural Discourse and Communication
Edited by Christina Bratt Paulston, Scott F. Kiesling, and Elizabeth S. Rangel

The Handbook of Hispanic Linguistics
Edited by José Ignacio Hualde, Antxun Olarrea, and Erin O’Rourke

The Handbook of Historical Sociolinguistics
Edited by Juan M. Hernández-Campany and J. Camilo Conde-Silvestre

The Handbook of Conversation Analysis
Edited by Jack Sidnell and Tanya Stivers

The Handbook of English for Specific Purposes
Edited by Brian Paltridge and Sue Starfield

The Handbook of Bilingualism and Multilingualism, Second Edition
Edited by Tej K. Bhatia and William C. Ritchie

The Handbook of Language Variation and Change
Edited by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes

The Handbook of Spanish Second Language Acquisition
Edited by Kimberly L. Geeslin

The Handbook of Chinese Linguistics
Edited by C.-I. James Huang, Y.-H. Audrey Li, and Andrew Simpson

The Handbook of Language, Gender, and Sexuality, Second Edition
Edited by Susan Ehrlich, Miriam Meyerhoff, and Janet Holmes

The Handbook of Language Emergence
Edited by Brian MacWhinney and William O’Grady

The Handbook of Bilingual and Multilingual Education
Edited by Wayne E. Wright, Soviceth Boun, and Ofelia Garcia

The Handbook of Discourse Analysis, Second Edition
Edited by Deborah Tannen, Heidi E. Hamilton, and Deborah Schiffrin

The Handbook of English Pronunciation
Edited by Marnie Reed and John M. Levis

The Handbook of Classroom Discourse and Interaction
Edited by Numa Markee

The Handbook of Korean Linguistics
Edited by Lucien Brown and Jaeohoon Yeon

The Handbook of Speech Production
Edited by Melissa A. Redford

The Handbook of Narrative Analysis
Edited by Anna De Fina & Alexandra Georgakopoulou

The Handbook of Contemporary Semantic Theory, Second Edition
Edited by Shalom Lappin

The Handbook of Portuguese Linguistics
Edited by W. Leo Wetzels, Joao Costa, and Sergio Menuuzzi

The Handbook of Translation and Cognition
Edited by John W. Schwieter and Aline Ferreira

The Handbook of Linguistics, Second Edition
Edited by Mark Aronoff and Janie Rees-Miller

The Handbook of Dialectology
Edited by Charles Boberg, John Nerbonne, and Dominic Watt

The Handbook of Technology and Second Language Teaching and Learning
Edited by Carol A. Chapelle and Shannon Sauro

The Handbook of Psycholinguistics
Edited by Eva M. Fernández and Helen Smith Cairns

The Handbook of Psycholinguistics
Edited by Eva M. Fernández and Helen Smith Cairns
For
William and Maya Jagels
and
Sadie Haltom Cairns
Contents

Notes on Contributors .......................... x
Prologue ........................................... xxiii
Eva M. Fernández and Helen Smith Cairns  

Part I Production ................................ 1

1 Overview ................................. 3
   Fernanda Ferreira  

2 Syntactic Encoding: Novel Insights Into the Relationship Between Grammar and Processing 13
   Julie Franck  

3 Signal Reduction and Linguistic Encoding 38
   T. Florian Jaeger and Esteban Buz  

4 Production in Bilingual and Multilingual Speakers 82
   Daniela Paolieri, Luis Morales, and Teresa Bajo  

5 Production of Signed Utterances 111
   Ronnie B. Wilbur  

6 Parity and Disparity in Conversational Interaction 136
   Jennifer S. Pardo  

7 Models Linking Production and Comprehension 157
   Chiara Gambi and Martin J. Pickering  

Part II Comprehension ......................... 183

8 Overview ........................................ 185
   Eva M. Fernández and Helen Smith Cairns  

9 Speech Perception: Research, Theory, and Clinical Application 193
   David B. Pisoni  

10 Cross-Language and Second Language Speech Perception 213
   Ocke-Schwen Bohn  

11 Models of Lexical Access and Morphological Processing 240
   Petar Milin, Eva Smolka, and Laurie Beth Feldman
12 Orthography, Word Recognition, and Reading 269
   David Braze and Tao Gong
13 The Bilingual Lexicon 294
   Judith F. Kroll and Fengyang Ma
14 Sentence Processing and Interpretation in Monolinguals and Bilinguals: Classical and Contemporary Approaches 320
   Matthew J. Traxler, Liv J. Hoversten, and Trevor A. Brothers
15 The Comprehension of Anaphora and Verb Agreement 345
   Janet L. Nicol and Andrew Barss
16 Prosody in Sentence Processing 365
   Elizabeth Pratt
17 Semantic-Pragmatic Processing 392
   Petra B. Schumacher
18 Comprehension in Older Adult Populations: Healthy Aging, Aphasia, and Dementia 411
   Jet M. J. Vonk, Eve Higby, and Loraine K. Obler
19 Neurolinguistic Studies of Sentence Comprehension 438
   Michael A. Skeide and Angela D. Friederici

Part III Acquisition 457
20 Overview 459
   Virginia Valian
21 Speech Perception in Infants: Propagating the Effects of Language Experience 470
   Catherine T. Best
22 Children’s Performance Abilities: Language Production 491
   Cecile McKee, Dana McDaniel, and Merrill F. Garrett
23 Language Comprehension in Monolingual and Bilingual Children 516
   Krista Byers-Heinlein and Casey Lew-Williams
24 Names for Things… and Actions and Events: Following in the Footsteps of Roger Brown 536
   Dani Levine, Kristina Strother-Garcia, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff
25 The Acquisition of Morphology 567
   Kamil Ud Deen
26 The Acquisition of Syntax 593
   Nina Hyams and Robyn Orfitelli
27 Social Interaction and Language Acquisition: Toward a Neurobiological View 615
   Sarah Roseberry Lytle and Patricia K. Kuhl
28 Bilingual Acquisition: A Morphosyntactic Perspective on Simultaneous and Early Successive Language Development 635
   Jürgen M. Meisel
<table>
<thead>
<tr>
<th>29</th>
<th>The Development of Morphosyntax in Child and Adult Second Language Acquisition</th>
<th>653</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gita Martohardjono and Elaine C. Klein</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Signed Language Acquisition: Input</td>
<td>674</td>
</tr>
<tr>
<td></td>
<td>Judy Kegl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Index</td>
<td>705</td>
</tr>
</tbody>
</table>
Notes on Contributors

Teresa Bajo is Full Professor at the Department of Experimental Psychology of Granada University and head of the Memory and Language Research Group in the Mind, Brain and Behaviour Research Center. Her research is dedicated to the study of inhibitory control in language selection and in memory, cognitive processes in language translation, and interpreting.

Andrew Barss is a faculty member in the departments of Linguistics and the Program in Cognitive Science at the University of Arizona. He has conducted research on the syntax and semantics of anaphora, WH-movement, quantifier scope, and focus constructions.

Catherine T. Best is Chair in Psycholinguistic Research at MARCS Institute, Western Sydney University, Australia. Her research and theory have focused on cross-language speech perception, and on the role of regional accent variation in native speech perception and spoken word recognition. She is widely known for her Perceptual Assimilation Model (PAM) and its more recent extension to second language (L2) learners, PAM-L2.

Ocke-Schwen Bohn is the professor and chair of English Linguistics at Aarhus University, Denmark. He received his Ph.D. from Kiel University (Germany) and spent his time as a post-doc working with James Flege at the University of Alabama in Birmingham. With funding from German and Danish research agencies, and in collaboration with American, Canadian, and Australian colleagues, Bohn’s research focuses on the causes and characteristics of foreign accented speech, speech perception (in infants, cross-language, and second language acquisition), and bilingual memory. He has published widely in journals such as Applied Psycholinguistics, Journal of the Acoustical Society of America, Journal of the International Phonetic Association, Journal of Phonetics, Memory, Speech Communication, and Studies in Second Language Acquisition.

David Braze is a linguist and senior scientist at Haskins Laboratories. He studies the cognitive structures and processes that support the human ability to fluidly assemble compositional meaning from more-or-less novel strings of words.
Dr. Braze’s investigations of reading comprehension explore how its cognitive bases (ability to comprehend speech, word knowledge, decoding skill, memory, executive function, and so on) are related to the ability to construct meaning from print, and how that may vary from one person to the next due to differences in biology or experience. Central to his research are questions of how lexical, grammatical, semantic and pragmatic processes interact with one another to yield the apprehension meaning from language, whether perceived by ear or by eye.

**Trevor A. Brothers** is a doctoral candidate in the Department of Psychology at UC Davis. His research focuses on anticipatory processes in sentence and discourse comprehension.

**Esteban Buz** is a post-doctoral researcher and lecturer at Princeton University. He received his Ph.D. in the Brain and Cognitive Sciences at the University of Rochester in 2016. His graduate work was partly funded by a pre-doctoral National Research Service Award from the National Institutes of Health (F31HD083020), a Provost Fellowship from the University of Rochester, and a National Institutes of Health training grant awarded to the Center for Language Sciences at the University of Rochester (T32DC000035). His research focuses on language production and specifically how and why speakers’ vary their speech in and across different situations. Of special focus in this researcher is whether speech variability is partly for achieving and maintaining robust communication.

**Krista Byers-Heinlein** (B.A., McGill University; M.A., Ph.D., University of British Columbia) is Associate Professor in the Department of Psychology at Concordia University, in Montreal, Canada. She holds the Concordia University Research Chair in Bilingualism and directs the Concordia Infant Research Lab. Her research investigates monolingual and bilingual language development in the early years.

**Helen Smith Cairns** is Professor Emerita, City University of New York, in the Department of Linguistics and Communication Disorders, Queens College, and the doctoral programs in Linguistics and in Speech, Hearing, and Language Sciences, at the CUNY Graduate Center. After receiving her Ph.D. in Psycholinguistics and Cognitive Processes at the University of Texas at Austin in 1970, she joined the Queens College faculty in 1971, where she taught until her retirement in 2003, having served as Department Chair and Dean of Graduate Studies and Research. She has pursued research in sentence processing and in first language acquisition, writing or editing six books (one co-authored with her husband, Chuck) and numerous articles and chapters. She is most proud of the students she has mentored over the years.

**Laurie Beth Feldman** is Professor of Psychology, at The University at Albany, State University of New York and Senior Scientist at Haskins Laboratories in New Haven, CT. She completed her undergraduate degree at Wellesley College and graduate degrees at the University of Connecticut. The unifying theme to her program of research is the question of how a language user concurrently manages
two linguistic codes. These interactions include two languages, two writing systems for one language, native and accented speech, speech and text, and emoticon and text. Much of this work has depended on collaboration with people in China, Germany, Israel, The Netherlands, Serbia, and Turkey and was supported by funds from NÁS, NSF, NICHD, IARPA, and Fulbright. She is a Fellow of APS, and a member of the AAAS and the Psychonomic Society.

**Fernanda Ferreira** is Professor of Psychology and Member of the Graduate Group in Linguistics at the University of California, Davis. She obtained her Ph.D. in Cognitive Psychology in 1988 from the University of Massachusetts, Amherst, and prior to moving to UC Davis in 2015, she held faculty positions at Michigan State University and the University of Edinburgh. She has published over 100 papers and her research has been funded by the National Science Foundation and the National Institutes of Health in the United States, and the Economic and Social Research Council in the United Kingdom. She served as Editor in Chief of the *Journal of Experimental Psychology: General*, and she is currently an Associate Editor of *Cognitive Psychology* and of *Collabra*, an Open Access journal recently launched by University of California Press. She is a Fellow of the American Psychological Society and the Royal Society of Edinburgh, and is currently an elected member of the Psychonomic Society’s Governing Board.

**Eva M. Fernández** received her Ph.D. in Linguistics in 2000. She is Professor at Queens College, City University of New York (CUNY) in the Department of Linguistics and Communication Disorders, and doctoral faculty at the CUNY Graduate Center in the Programs in: Linguistics; Hispanic and Luso-Brazilian Literatures and Languages; and Speech-Language-Hearing Sciences. Her research focuses on cross-linguistic language processing in monolingual and bilingual populations; she has additional research interests in measures of student success in higher education. Her work has been supported by grants from the National Science Foundation, the United States Department of Education, and the Association of American Colleges & Universities. She currently serves as Assistant Provost at Queens College.

**Julie Franck** is Maître d’Enseignement et de Recherche at the University of Geneva. Her work combines theoretical insights from linguistic theory and psycholinguistic experimental methods with the aim of identifying general principles of syntactic production and comprehension in adults and children. Her interest in exploring the relations between grammar and processing has developed through several topics including agreement, the acquisition of word order and subordination, the learning of hierarchical structure in artificial grammars, and the processing of islands.

**Angela D. Friederici** studied German, Romance languages, linguistics, and psychology in Bonn, Germany and Lausanne, Switzerland. She obtained her doctorate from Bonn University, Germany and conducted postdoctoral research at the
Massachusetts Institute of Technology. After being appointed full professor of Cognitive Psychology at the Free University Berlin, Germany, she was founding director of the Max Planck Institute for Human Cognitive and Brain Sciences, where she is head of the Department of Neuropsychology. Her research is focused on the neuroscience of language. Angela D. Friederici is Vice President of the Max Planck Society, and honorary professor at the Universities of Leipzig, Potsdam and at the Charité Berlin.

Chiara Gambi is a post-doctoral researcher at the University of Edinburgh. She received her Ph.D. in Psychology from the same university in 2013 and has subsequently been a post-doc with the Psycholinguistics Group at Saarland University before joining the Developmental Lab at the University of Edinburgh in 2014. She has published in Cortex, Journal of Experimental Psychology: Learning, Memory, and Cognition, and in Language, Cognition, and Neuroscience. Her research focuses on the relationship between comprehension and production, and particularly between the latter and prediction processes.


Roberta Michnick Golinkoff, Unidel H. Rodney Sharp Chair in the School of Education at the University of Delaware, and also in the Departments of Psychological and Brain Science and Linguistics and Cognitive Science, directs the Child’s Play, Learning, and Development Lab whose focus is language acquisition, spatial development, and the benefits of playful learning. Her research has resulted in numerous professional articles and 14 books, including books for parents and practitioners. Her latest book is entitled, Becoming Brilliant: What Science Tells Us About Raising Successful Children (APA Press). Dr. Golinkoff has received various awards including the John Simon Guggenheim Fellowship, the James McKeen Cattell Sabbatical award, and with Dr. Hirsh-Pasek, the James McKeen Cattell Award from the Association for Psychological Science and the Society for Research in Child Development Distinguished Scientific Contributions Award. She is also a Francis Alison Scholar, the highest honor given to faculty at her University. Dr. Golinkoff has appeared on many television and radio shows and speaks at conferences around the world about children’s development.
Tao Gong is a computational modeler and research scientist at Haskins Laboratories. His research investigates the origins and evolution of language, the unique ways in which biology and culture interact during language learning, and the roles of language and literacy related skills on reading behaviors. By means of agent-based and mathematical modeling, as well as behavioral, artificial language learning, and eye-movement experiments, Dr. Gong’s studies of language evolution and learning explore how domain-general cognitive capacities (sequential learning, pattern extraction, memory, and so on) give rise to specific language processing mechanisms underlying particular language structures. His work addresses questions concerning how individual learning, biological evolution, and socio-cultural transmission interact with one another to yield culturally specific language and linguistic behaviors.

Eve Higby is a post-doctoral researcher at the University of California, Riverside. She received her Ph.D. in Speech-Language-Hearing Sciences from the City University of New York Graduate Center, with a focus on neurolinguistics and cognitive science. Her research focuses on brain and language changes in bilingualism and aging and the interaction between language and other cognitive faculties. Her publications include topics such as multilingualism and the brain, cross-linguistic influence in bilingualism, brain and language in aging, and second language acquisition. She is currently working on the second edition of the Language and the Brain book by Cambridge University Press with Loraine K. Obler, Kris Gjerlow, and Ioulia Kovelman.

Kathy Hirsh-Pasek is the Stanley and Debra Lefkowitz Faculty Fellow in the Department of Psychology at Temple University and a Senior Fellow at the Brookings Institution. Her research examines the development of early language and literacy as well as the role of play in learning. With her long-term collaborator, Roberta Golinkoff, she is author of 14 books and hundreds of publications. She is the recipient of three American Psychological Association Awards: the Bronfenbrenner Award, the Distinguished Service to Psychological Science Award, and the Distinguished Lecturer Award. She is the President Elect of the International Society for Infant Studies and served as an Associate Editor of Child Development. She is on the Steering Committee of the Latin American School for Education, Cognitive, and Neural Science and Vroom, the Boston Children’s Museum, and Jumpstart. Her 2016 book, Becoming Brilliant (with Golinkoff) was on The New York Times best sellers list in education and parenting. Dr. Hirsh-Pasek is also routinely quoted on national media such as The New York Times and National Public Radio.

Liv J. Hoversten is a doctoral candidate in the Department of Psychology at UC Davis. Her research focuses on language competition and language control in bilingual speakers.

Nina Hyams is Professor of Linguistics at the University of California, Los Angeles. Her primary area of research is childhood language development, especially the
acquisition of syntax and semantics. She is author of the book *Language Acquisition and the Theory of Parameters* (D. Reidel Publishers), a milestone in language acquisition research and she co-authors (with Robert Rodman) a popular introductory textbook, *An Introduction to Language* (Cengage, Wadsworth Learning). She has published numerous articles on the development of syntax, morphology and semantics in children and she frequently lectures on these topics throughout the United States, Europe and Japan. She received her Ph.D from the Graduate Center of the City University of New York in 1983.

**T. Florian Jaeger** holds an M.A. in Linguistics and Computer Science from Humboldt University to Berlin (2001) and a Ph.D. in Linguistics with a designation in the Cognitive Sciences from Stanford University (2006). Other education includes stints at MIT, UC Berkeley, Edinburgh, and at post-doc in psychology at UCSD. In 2007, he joined the University of Rochester, where he now is a Professor in the Departments of Brain and Cognitive Sciences, Computer Science, and Linguistics. He is the Director of the Center for Language Sciences, an Alfred P. Sloan Fellow in the Neurosciences, and NSF CAREER recipient in Robust Intelligence. Research in his Human Language Processing lab focuses on the development of computational frameworks for language production and understanding. This includes research on the consequences of noise and variability (e.g., between talkers) for communication and the study of inference under uncertainty and learning in production and comprehension. This research is funded by the National Science Foundation and the National Institutes for Health.

**Judy Kegl** is Professor of Linguistics and Director of the Signed Language Research Laboratory at the University of Southern Maine, where she teaches both linguistics and interpreting. Her primary research areas are the linguistics of American Sign Language (focusing on syntax), neurolinguistics of both signed and spoken languages, and predicate argument structure in English and ASL. She is best known for her work documenting the emergence of a signed language in Nicaragua, which has been ongoing for the past 25 years. She argues that it is a creation of language from language-like, but non-language input that, via children’s acquisition process, has yielded a full-fledged human language.

**Elaine C. Klein** is Professor of Linguistics Emerita at the CUNY Graduate Center. Professor Klein specializes in second language development and, recently, in studies of and interventions for immigrant students with special language and literacy needs. With Dr. Gita Martohardjono, Dr. Klein studied a population of adolescent emergent bilinguals who had been under-served and under-researched: Students with interrupted or inconsistent formal education (SIFE). Currently, as PI for Bridges for Academic Success, Dr. Klein and her team, funded by the New York State Education Department, have developed a set of research-based curricula for SIFE, and professional development for their teachers, now being used in school districts around New York State.
Judith F. Kroll is Distinguished Professor of Psychology at the University of California, Riverside and the former director of the Center for Language Science at Pennsylvania State University. The research that she and her students conduct concerns the way that bilinguals juggle the presence of two languages in one mind and brain. Their work, supported by grants from NSF and NIH, shows that bilingualism provides a tool for revealing the interplay between language and cognition that is otherwise obscure in speakers of one language alone. She was one of the founding editors of the journal *Bilingualism: Language and Cognition* (Cambridge University Press).

Patricia K. Kuhl is the Bezos Family Foundation Endowed Chair for Early Childhood Learning, Co-Director of the UW Institute for Learning & Brain Sciences, Director of the NSF-funded Science of Learning Center, and Professor of Speech and Hearing Sciences. She is internationally recognized for her research on early language and brain development, and studies that show how young children learn. Dr. Kuhl’s work has played a major role in demonstrating how early exposure to language alters the brain. It has implications for critical periods in development, for bilingual education and reading readiness, for developmental disabilities involving language, and for research on computer understanding of speech.

Dani Levine is a Ph.D. candidate in the Department of Psychology at Temple University. She received her B.A. in Psychology and Biology with a concentration in Neuroscience from Williams College. Her research interests include language acquisition, cognitive development, and early childhood education. Dani is currently studying the processes young children use to parse events into meaningful units and is examining the link between event segmentation and word learning. She also worked on a project, funded by the Institute of Education Sciences, in which she helped to develop a new way to measure language growth in preschoolers.

Casey Lew-Williams (B.A., University of California, Berkeley; Ph.D., Stanford University) is an Assistant Professor in the Department of Psychology at Princeton University. He co-directs the Princeton Baby Lab, and his research focuses on understanding how experience shapes language learning in infants, children, and adults.

Sarah Roseberry Lytle is the Director of the Outreach and Education division at the Institute for Learning & Brain Sciences (I-LABS) at the University of Washington. The I-LABS Outreach and Education team communicates the latest science of child development to those who can act on it, including parents, educators, policymakers, and opinion leaders. Dr. Lytle was previously a Postdoctoral Fellow at I-LABS. Before coming to the Institute, she earned a B.A. in Psychology and Spanish from the University of Notre Dame and a Ph.D. in Developmental Psychology at Temple University. Her research focused on the role of social cues in infants’ and toddlers’
language learning and how social interactions might help toddlers learn from
screen media. Dr. Lytle was 2014–2016 Zero To Three Fellow.

Fengyang Ma is Visiting Assistant Professor in the School of Education at the
University of Cincinnati. Her research interests center on the investigation of
bilingualism from psycholinguistic and neurolinguistic perspectives. She received
her Ph.D. from Tsinghua University in 2012 and then from 2012 to 2015 was a
postdoctoral visiting scholar at Pennsylvania State University, working in the
Center for Language Science and the Social, Life, and Engineering Sciences
Imaging Center.

Gita Martohardjono is Professor of Linguistics and the Executive Officer of the
Ph.D. Program in Linguistics at the Graduate Center, where she also co-directs
RISLUS (Research Institute for the Study of Language in Urban Society). Her
research interests center on psycholinguistic investigations of syntactic development
in second language learning; the development and use of the first-learned language
in heritage language speakers; and the development of academic literacy in child
and adolescent bilinguals. Her most recent work examines cross-linguistic influence
in first- and second-generation bilinguals through the use of ERP and eye-tracking
methodologies. She is currently leading a team of linguists in developing native
language tasks in 15 languages for the New York State Education Department
(NYSED).

Dana McDaniel is a Professor of Linguistics and the Chair of the Department of
Linguistics at the University of Southern Maine. Together with Professor Wayne
Cowart, she developed the USM undergraduate Linguistics program from the
ground up. She directs the Child Language Lab, which emphasizes children’s
syntax and language production. Her current focus is the potential role of the lan-
guage production system in the evolution of syntax, cross-linguistic variation, and
children’s acquisition of syntax. Her work in these areas has been supported by
the National Science Foundation.

Cecile McKee is a Professor of Linguistics at the University of Arizona. She has
affiliate appointments in Psychology, Cognitive Science, and Second Language
Acquisition and Teaching. She directs the Developmental Psycholinguistics Lab,
where current research emphasizes children’s production system and children’s
informal science learning. With collaborator Sam Supalla, McKee has also studied
deaf children’s literacy development in research that was supported by the James
S. McDonnell Foundation. McKee is also active in outreach, especially via festivals
and museums. She also served from 2006 to 2015 on University of Arizona institu-
tional review boards. She worked as a program director at the National Science

Jürgen M. Meisel is Emeritus Professor of Romance Linguistics at the University
of Hamburg (Germany) and Adjunct Professor in the Department of Linguistics
Notes on Contributors

and Distinguished Fellow in the Language Research Centre at the University of Calgary. In 2004, he was awarded an honorary doctorate by the Faculty of Humanities, Lund University (Sweden). He directed the Research Center on Multilingualism at the University of Hamburg from its foundation in 1999 through 2006, and he is founding editor of the journal Bilingualism: Language and Cognition. His main research areas are bilingual first language acquisition, naturalistic second language acquisition by children and adults, Romance syntax, and historical linguistics.

Petar Milin is Senior Lecturer in Data Science at The University of Sheffield, United Kingdom. In the late 2000s he was among most active researchers in developing what is known as information-theoretic approach to language processing. He continued his work as one of the principal co-developers of the Naive Discrimination Learning model, which is simultaneously a computational model and research framework to investigate language acquisition and processing. He collaborates widely and on various topics with colleagues from the United States, Canada, UK, Belgium, Germany, and Italy. He is a member of the Psychonomic Society, and a co-founder of the International Quantitative Morphology Association.

Luis Morales is currently Professor at Universidad Loyola Andalucía (Spain), although the chapter was written during his academic affiliation at the University of Granada (Spain). His research is about lexical access in monolingual and bilingual speakers and the cognitive control processes involved in language selection. He focuses at the level of grammatical gender through production and comprehension tasks.

Janet L. Nicol is a Professor of Linguistics, Psychology, and Cognitive Science at the University of Arizona. She is an active faculty member in the Graduate Interdisciplinary Programs in Cognitive Science and Second Language Acquisition and Teaching. She has conducted extensive research on the comprehension of agreeing elements and discontinuous dependencies, and the production of agreement, testing various populations (including children, people with aphasia, and monolingual and bilingual young adults), with a range of methodologies (including self-paced reading, eye-tracking, ERP, and elicited production). Recent research has also focused on the learning of vocabulary and grammar in a second language and the factors that optimize such learning, including the manner of presentation of materials and learner-inherent capacities, such as working memory and executive function. Other research interests include how to optimize the learning of any kind of material, particularly in the context of online teaching. Her edited volume, One Mind, Two Languages: Bilingual Language Processing, was published in 2001.

Loraine K. Obler is Distinguished Professor of Speech-Language-Hearing Sciences, with a joint appointment in Linguistics, at the City University of New York Graduate Center. She also heads the Steering Committee of the MA
Program in Multilingual and Multicultural Speech-Language Pathology at Hadassah Academic College Jerusalem. Her publications in neurolinguistics focus on aphasiology (e.g., cross-language studies of agrammatism) and changes in comprehension and lexical retrieval associated with advancing age. Her bilingualism publications also include The Bilingual Brain: Neuropsychological and Neurolinguistic Aspects of Bilingualism (co-authored with Martin Albert) and Bilingualism Across the Lifespan (co-edited with Kenneth Hyltenstam), as well as studies on poor and excellent second-language learners. As well she is currently in the process of revising Language and the Brain (with Kris Gjerlow) with co-authors Eve Higby and Ioulia Kovelman.

Robyn Orfitelli is a Lecturer of Linguistics at the University of Sheffield. Her primary research areas are syntax and the childhood language development of syntax. She has published articles in a variety of areas, including the development of syntax in first and second language acquisition, experimental methodology, and prosody, and has a forthcoming monograph, Argument Intervention in the Acquisition of A-movement (Springer). She has presented conference papers and lectures on these topics in North America, Europe, and Australia. She received her Ph.D. from the University of California, Los Angeles in 2012.

Daniela Paolieri is Professor of Memory and Augmentative Communication Systems at Granada University. In her research, she has focused on grammatical gender processing in language production and comprehension, addressing this issue both in monolingual and bilingual contexts.

Jennifer S. Pardo is Professor of Psychology and Director of the Speech Communication Laboratory at Montclair State University. She received her Ph.D. in Cognitive Psychology from Yale University in 2000, and has held academic positions at Barnard College, Wesleyan University, and The New School for Social Research. Her research centers on the production and perception of spoken language with an emphasis on understanding variation and convergence in phonetic form, and has been supported by grants from the National Science Foundation and the National Institutes of Health. Some recent publications of this work have appeared in Journal of Memory & Language, Frontiers in Psychology, and Attention, Perception, & Psychophysics.

Martin J. Pickering is Professor of the Psychology of Language and Communication at the University of Edinburgh. He is the author of more than 130 journal articles and numerous other publications in language production, comprehension, dialogue, reading, and bilingualism. His articles have appeared in Psychological Bulletin; Trends in Cognitive Sciences; Journal of Neuroscience; Cerebral Cortex; Psychological Science; Cognitive Psychology; Journal of Memory and Language; Cognition; and Journal of Experimental Psychology: Learning, Memory, and Cognition. In particular, he published “Toward a mechanistic psychology of dialogue” (Behavioral and Brain Sciences, 2004).
and “An integrated theory of language production and comprehension” (*Behavioral and Brain Sciences*, 2013), both with Simon Garrod. He is a Fellow of the Royal Society of Edinburgh and a former editor of the *Journal of Memory and Language*.

**David B. Pisoni** holds the position of Distinguished Professor of Psychological and Brain Sciences and Chancellor’s Professor of Cognitive Science at Indiana University in Bloomington. He has carried out seminal research on human speech perception, spoken word recognition, language processing, and perceptual development in infants and children. For the past 24 years, Professor Pisoni has worked with researchers at the IU School of Medicine on clinical problems associated with hearing impairment in deaf children who use cochlear implants. His work has had important clinical applications for understanding the enormous variability in speech and language outcomes following implantation. Throughout his career, Professor Pisoni has made significant contributions in basic, applied and clinical research in areas of speech perception, production, synthesis, and spoken language processing.

**Elizabeth Pratt** received her Ph.D. in Linguistics from the City University of New York, where she is currently engaged with linguistic and educational research through the Center for Advanced Study in Education (CASE), and the Research Institute for the Study of Language in Urban Society (RISLUS). Her primary interests are in the unique integration of linguistic theory and research within a general framework of cognitive processing. Her recent research has concentrated on structural and prosodic effects on agreement processing and comprehension, with the goal of incorporating these linguistic insights into a comprehensive model of memory, executive control, and attention. Aligned with this research program is the parallel development and testing of pedagogical strategies and interventions for speakers and learners from diverse language backgrounds and profiles.

**Petra B. Schumacher** is Professor of Empirical Linguistics at the Department of German Language and Literature I at the University of Cologne, Germany. She received her Ph.D. in Linguistics from Yale University. Her research focuses on processes at the interface of syntax, semantics and pragmatics, including reference resolution, information structure, and experimental pragmatics. She is primarily interested in the time-course of language processing and conducts research with different populations and languages.

**Michael A. Skeide** studied Linguistics at Heidelberg University, Germany, and Neurobiology at Harvard University. For his work on the functional and structural plasticity underlying the developmental emergence of higher-order language processing, he received a Ph.D. in Psychology from the University of Leipzig, Germany. He is now a postdoctoral researcher at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, where he studies the neural basis of normal and impaired language and literacy development.
Eva Smolka is a psycho- and neurolinguist. She studied cognitive psychology and English linguistics at Bar-Ilan University and the University of Haifa, Israel, and at the Philipps-University Marburg, Germany. In her research, she explores the storage and processing of complex word formations in lexical memory. She applies psycho- and neurolinguistic methods to study word formation in healthy adults, language acquisition in children and bilinguals, as well as language loss in patients with language impairments such as aphasia. Dr. Smolka is currently at the University of Konstanz, Germany, and is collaborating with many researchers from Europe, Israel, Canada, and the United States.

Kristina Strother-Garcia is a student in the Department of Linguistics & Cognitive Science at the University of Delaware. Her research interests include computational/mathematical phonology, syllable theory, abstract representations of linguistic structures, and grammatical inference.

Matthew J. Traxler is a professor of psychology and administrator at the University of California, Davis. His research focuses on sentence and discourse comprehension in healthy adults, deaf adults, and patients with schizophrenia.

Kamil Ud Deen was born and raised in Nairobi, Kenya, and came to the United States as a young man. He received his Ph.D. from UCLA, and is currently as associate professor in Linguistics at the University of Hawaii at Manoa. He has conducted research on his native Swahili, as well as on a wide variety of languages, including Korean, Japanese, Thai, Tagalog, Serbian, and Vietnamese. His interests lie in cross-linguistic research on the acquisition of morphology and syntax.

Virginia Valian is Distinguished Professor of Psychology at Hunter College and is a member of the doctoral faculties of Psychology, Linguistics, and Speech-Language-Hearing Sciences at the CUNY Graduate Center. She directs the Language Acquisition Research Center. Valian works in two areas—the psychology of language and gender equity. In first language acquisition, Valian performs research with the aim of developing a model of acquisition that specifies what is innate, how input is used by the child, and how the child’s syntactic knowledge interacts with knowledge in other linguistic and extra-linguistic domains. To approach those questions she uses a variety of methods, including computer-assisted corpus analysis, comprehension experiments, elicited imitation experiments, and elicited production experiments. Valian’s second language area is the relation between bilingualism and higher cognitive functions in adults; she thinks there is one though she has yet to find it. To learn more about her work: http://maxweber.hunter.cuny.edu/psych/faculty/valian/valian.htm, http://littlelinguist.hunter.cuny.edu/, and http://www.hunter.cuny.edu/genderequity/.

Jet M. J. Vonk is a Ph.D. Candidate in the Neurolinguistics Laboratory of Loraine K. Obler, Ph.D., within the Speech-Language-Hearing Sciences department at the
Graduate Center, City University of New York (CUNY). Additionally, she is a Visiting Scholar at the Memory and Aging Center of the University of California, San Francisco (UCSF) working with Maria Luisa Gorno-Tempini, M.D., Ph.D., and Nina Dronkers, Ph.D. Her research projects focus on lexical-semantic aspects in healthy aging, Alzheimer’s disease, and primary progressive aphasia and her broader interests include changes in language and cognition due to age and neurobiological factors in healthy aging and dementia. Multiple of her projects are in collaboration with the Taub Institute for Research on Alzheimer’s Disease and the Aging Brain at Columbia University and the Language in the Aging Brain Project at Boston University/Boston VA Healthcare System.

Ronnie B. Wilbur, Ph.D, is a theoretical and experimental linguist who holds the rank of full professor in Linguistics and in Speech, Language, and Hearing Sciences at Purdue University. She has conducted research on sign languages for more than three decades, pioneering new methods for analyzing sign languages that parallel those available for spoken languages and developing educational applications for her research findings. She was the founding editor of the field’s flagship journal, *Sign Language and Linguistics*. In addition to her funded research (NIH, NSF, Office of Education), guest lectureships and visiting professorships in the United States and Europe, Wilbur served a dozen years as Director of the Purdue Linguistics Program. A member of the Purdue University Book of Great Teachers, she received the 2014 Schleman Gold Medallion Award for accomplishments related to women’s issues on campus, as well as a Focus Award for diversity and disability service, and a Seed for Success Award for research funding in excess of $1 million. In 2015 she was awarded the Purdue University Research and Scholarship Distinction Award, Purdue's highest research honor in Humanities and Social Sciences.
Psycholinguistics is the field of study that addresses how people process and acquire a central aspect of what it means to be human: language. The body of research surveyed in the chapters that follow addresses this essential faculty of the human species from a number of perspectives, drawing predominantly from the disciplines of linguistics and psychology, from cognitive science, and from neuroscience. Our goal in assembling this collection of contemporary research is to present the state of the field of psycholinguistics early in the twenty-first century, a field with vibrant research trends that, combined, provide a rich picture of how language works in the human mind and how it is acquired.

Psycholinguistics is a relatively new field, with origins in a seminar in 1953 at Indiana University held in conjunction with the Linguistics Institute, resulting in a book edited by Charles Osgood and Thomas Sebeok titled *Psycholinguistics: A Survey of Theory and Research Problems* (Osgood & Sebeok, 1954). Their approach to the study of language focused on three disciplines: linguistics, learning theory, and information theory. It laid out foundational questions regarding the mechanisms and units that underpin hearing and speaking. A reviewer of Osgood and Sebeok’s book made the prescient observation that “the joint exploration which it describes is something more than just another interdisciplinary venture” (Newman, 1955: 1097). Indeed, in the mid-twentieth century the then-new field of generative linguistics collided with behaviorist psychology, resulting in a scientific revolution with many of the characteristics observed by Thomas Kuhn in his book *The Structure of Scientific Revolutions* (2012). The history of psycholinguistics has been a story of the influence of linguistic theory on theoretical psychology, and the emergence of psycholinguistics as a dramatically altered but ultimately autonomous and prolific science.

In the early days of psycholinguistics, linguistic theory was actually taken to be a theory of linguistic performance, but the falsification of the Derivational Theory of Complexity (Fodor, Bever, & Garrett, 1974) demonstrated the fallacy of that approach. However, the profound insight (Chomsky, 1959) that language is a mental construct—specifically, that knowledge of language is represented in the individual’s mind/brain—changed psychology forever and replaced behaviorist models of language with a cognitive view that relies on mental representations
and processes that underlie the linguistic life of humans. The depth and breadth of the chapters in this Handbook attest to the depth and breadth of the content of contemporary psycholinguistics, a field that has expanded well beyond its early conceptualization as providing a model of how linguistic competence is deployed in the production and comprehension of sentences. A glance at the table of contents of this Handbook reveals how far we have come from that original conceptualization: the research is informed by cognitive and neurocognitive frameworks as well as by information theory, explores sociocultural parameters, incorporates concepts from evolutionary biology, and has direct relevance to education, speech/language pathology, and medicine.

Parts 1 and 2 of the Handbook focus respectively on language production and language comprehension, but it is clear that in some ways this is an artificial distinction. There are representations and processes that are common to both speaking and comprehending, and they overlap as they unfold. The field is not only concerned with the production and comprehension of spoken language, but also of signed languages. Sign language has profound ramifications for cognition and offers insights and research avenues unavailable if we restrict the domain of study to spoken languages. Likewise, research concerned with how speakers of more than one language produce and comprehend their languages provides unique ways to explore the architecture of the mechanisms that underpin linguistic performance.

Both comprehension and production at the sentence level and beyond rely on the activation of lexical information, prosodic analysis, and internal parsing principles, as well as principles of linguistic organization subsumed under the formal grammar. Higher-level processes are invoked when speakers and hearers engage in conversations. Linguistic theory, while characterizing individual linguistic competences, describes universal characteristics of human languages that constrain representations at every level of both production and comprehension. Psychological processes involved in the production and comprehension of language (and in some cases multiple languages) go far beyond representations constrained by linguistic theory to encompass powerful processes of linguistic organization and parsing.

The acquisition of language, the subject of Part 3, has similarly undergone extensive revision and expansion since the early days of psycholinguistics. A child’s development progresses from an initial state sensitive to universal properties of languages to a state consisting of fully formed representations of the native language. This development takes place in a remarkably brief period of time: by the time a child begins school (typically around 5 or 6 years old), a marvelously sophisticated mental system is in place. That trajectory is informed and constrained by basic principles of linguistic organization, as well as by the child’s developing perceptual system, lexical store, and additional cognitive abilities. Powerful internal capacities of pattern recognition, statistical monitoring, and memory contribute to the acquisition of a child’s native language. An explosion of research on the acquisition of two or more languages and also of signed languages has enriched what we know about language development. We have always known
that a child must be exposed to a language to acquire it, but recent advances in contemporary research have augmented how we understand and describe the characteristics of linguistic input, the feedback available to the child, and the quality of interactions with the child’s linguistic environment.

We are extremely fortunate to have recruited 52 leading scientists in contemporary psycholinguistics from 32 institutions in 9 different countries in North America, Australia, and Europe. Their contributions to this *Handbook* describe both the results of contemporary psycholinguistic research and the puzzles that remain for scholars to tackle in the future. To better frame the presentations in each of the chapters, each of the three sections begins with a chapter providing an overview of the contributions in that section, how they connect to one another, and how they relate to psycholinguistics in general. Our contributors have also strived to make the content accessible to readers who may not necessarily be experts in the sub-disciplines featured in each chapter. Our hope is that this volume will be of value to students and senior scholars alike and will make a contribution to the exciting, robust field of psycholinguistics.

The development of this volume has been very positively informed by the rich stimulation in the area of the study of language provided by the City University of New York, and we are grateful to all of our colleagues and students there for their support and insights, particularly Diane Bradley, Janet Fodor, and Irina Sekerina. We also owe a special thanks to the support team at Wiley: Danielle Descoteaux (who helped us envision the volume in its earliest phases), Mark Calley, and Tanya McMullin, as well as Manish Luthra, Vimali Joseph, and the editorial/production group. But we owe the most to our contributors who, individually and collectively, have made this volume an extensive and authoritative review of the state of the field.

REFERENCES


Part I  Production
1 Overview

FERNANDA FERREIRA
University of California, Davis

As a person speaks, a great deal of processing activity is taking place behind the scenes. In the situation of some sort of dialogue or conversation, a speaker will grab one of the many ideas activated in working memory and commit it to the processes of linguistic formulation. This thought must then be translated into specific concepts, the words that express those concepts must be retrieved, the words must be organized into a structure that communicates the thought effectively, and the entire plan must be converted to a phonological representation (either in speech or sign) that will allow the utterance eventually to be articulated. At the same time, speakers must take into account the needs of their conversational partners. They also need to keep track of how the communication is unfolding: That is, they must consider the knowledge their interlocutors bring to the conversation, how their common ground is incrementally being built up as the interaction proceeds, and how effective their contributions are given the goals of the exchange. The chapters in this section touch on many of these important ideas, spelling out in detail how syntactic structures are generated, how redundancy and givenness are conveyed, how multiple language systems are coordinated, and how conversations are managed. These chapters make clear the enormous progress that has been made over the last 50 years in uncovering the architecture of the production system as well as the systems with which it interacts.

Generating a syntactic structure

The syntactic level of representation takes center stage in research on language production. As Bock pointed out decades ago (Bock, 1982), it is at least coherent to ask whether people build a syntactic level of representation to mediate between form and meaning during comprehension, in part because, for many semantically
constrained sentences, a “bag of words” approach will be sufficient to establish the underlying event structure specifying who did what to whom. But when it comes to production, you can’t fake syntax. If you attempt to speak a language with grammatical agreement, it will be painfully obvious when you err. If you do not understand the rules of word order, the result may be awkward at best and misleading at worst. For these and other reasons, theoretical models of production have always taken the problems of syntactic planning seriously. In speech error models of production like the ones proposed by Garrett and Fromkin (Fromkin, 1971; Garrett, 1975), the first stage of grammatical encoding (in Garrett’s model, the functional level) is the entry point into the language system proper, as distinct from the general-purpose conceptual system that supports any type of perceptual-motor encoding. Syntax cannot be ignored in theories of language production, and research has proceeded accordingly.

Franck thus opens up this production section with a careful, analytical discussion of the syntactic system in general, and then moves quickly to what she treats as the model process for syntactic planning, which is agreement. As Franck notes, agreement is genuinely syntactic: It is a formal mechanism for linking words across often widely separated sentence positions. Franck begins with the studies by Bock and colleagues that initiated this research program, which used agreement errors as a source of information about the nature of syntactic representations. As Franck also points out, this early work seemed to reinforce Garrett’s fundamental assumptions about the architecture of production, which mapped the modularity of representations onto strictly serial processing. Speakers were assumed to first plan the meaning of what they would say, then generate the appropriate syntactic forms, and then engage in phonological planning.

This early work was critical for fostering a discussion about the consequences of this architecture for people’s ability to make syntactic decisions online, and as one hopes to see in a field of inquiry in which concepts are specified in enough detail to be falsifiable, some of the key assumptions did not survive tough empirical scrutiny. Franck summarizes the problematic results, many of which came from her own research on agreement, done in collaboration with her colleagues. These challenges to what Bock and colleagues referred to as the Marking and Morphing model motivated the development of Franck and colleagues’ alternative Feature and Controller Selection model, which differs from the Marking and Morphing model in a number of key respects. Perhaps the most central is the separation of the stage responsible for selecting features relevant to agreement from the stage that identifies the relevant controller of agreement. Semantic, syntactic, and morphological features influence both stages, but in different ways. This model also tightly links syntactic structures and memory representations, since the selection of both features and controllers is strongly influenced by the availability of information in working memory. In addition, the model assumes that the more prominent the syntactic position of a word, the greater its accessibility in memory. Franck ends her chapter with a useful roadmap for future research on syntax in production.

But what does this model of agreement imply for the concept of incremental production, for example? In traditional models of language production, a key
question concerned the planning units for any level of representation, including for syntax. Do speakers plan an entire clause before beginning to speak, or are planning and execution processes cascaded? In the early days of production research, many papers were published on this subject (e.g., Ford & Holmes, 1978), with evidence suggesting clausal planning units for grammatical encoding. Late twentieth-century models such as Levelt’s (1993) moved away from this idea and toward incrementality with the suggestion that planning domains should be as small as possible, and perhaps no larger than a single word. But how do we reconcile this incremental approach with the facts concerning agreement, where, as Frank notes, the controller and the form with which it agrees could in principle be indefinitely separated (and in practice are often separated by several words)? In fact, in recent years, the pendulum has begun to swing back toward the view that planning units for syntax are probably at least phrasal (Allum & Wheeldon, 2007; Bock & Cutting, 1992; Ferreira, 2000; Martin, Crowther, Knight, Tamborello, & Yang, 2010), and that the size of those units are likely not architecturally determined, but instead vary depending on the goals of the speaker (Ferreira & Swets, 2002; Wagner, Jescheniak, & Schriefers, 2010). It would be interesting to know, then, how findings concerning agreement speak to this question of planning units for grammatical encoding in more detail.

Distributing information

Typically, the same idea can be linguistically conveyed in more than one way, which presents the production system with both an opportunity and a set of processing decisions. This issue of flexibility in production connects to the previous discussion concerning how syntactic structures are generated, because one of the tasks of the production system is to make syntactic choices such as whether to produce an active or passive sentence, or whether to include an optional element such as the complementizer that in a sentence. In addition, speakers vary the way they pronounce the same word depending on features such as familiarity as well as predictability. The chapter by Jaeger and Buz focuses on the phenomenon of reduction, and probabilistic reduction in particular, which they link to contextual predictability. The general idea is that the more expected something is, the more reduced will be its pronunciation.

Jaeger and Buz then link the phenomenon of reduction to three general accounts of production: one that emphasizes ease of production, another that emphasizes the facilitation of communication, and one that emphasizes representational issues. The first account they consider assumes that reduction occurs because it facilitates the job of the speaker. The second account links reductions to speakers’ attempts to make the task of the listener easier. And the third account attempts to connect phenomena of language change to online language production. Jaeger and Buz conclude by arguing that all three influences must play a role in explaining the robust, cross-linguistically attested tendency on the part of speakers to reduce predictable forms. As they point out at the end of their chapter,
an important question that remains to be answered is precisely how these three approaches mesh with one another. Another open question concerns omission of linguistic forms, which is also related to predictability. For example, in a null subject language, the likelihood that a speaker will produce an overt pronoun or leave the position null depends on the predictability of the corresponding referent. Is omission simply the extreme case of reduction, or does the speaker face a binary choice in cases such as these? This would seem to be an important question for future work, as Jaeger and Buz note.

**Bilingualism, multilingualism, and signing**

The majority of people live with more than one language system, and so a critical question for theories of language production concerns how these different databases of information are coordinated and managed. Paolieri, Morales, and Bajo’s chapter examines this issue in detail. They present the problem as follows: How do speakers choose between two forms from different languages that express the same idea? One class of models of bilingual production assumes selective access, so that the decision to speak in a given language effectively shuts off any other languages the speaker might know. In contrast, in nonselective models, forms across languages interact, potentially leading to competition and interference. Linking back to Frank’s interesting chapter, Paolieri *et al.* (this volume) go beyond standard evidence for lexical interference effects to highlight findings concerning grammatical gender interactions across languages as well. The so-called gender congruency effect is observed when words in different languages happen to belong to the same gender class. Negative transfer is even observed during production of a non-gendered language (e.g., English) when spoken by people who also know a gendered language (e.g., Spanish). Syntactic information is also thought to be explicitly marked according to whether forms are shared or not, a conclusion that emerges from research on cross-linguistic syntactic priming.

Of course, no discussion of bilingual language production would be complete without some consideration of the so-called bilingual advantage. Paolieri *et al.* present a balanced and up-to-date analysis of the evidence for and against the theory of bilingual production which postulates the need for inhibitory control, and which further assumes that the frequent exercise of cognitive control sharpens the non-language cognitive system overall. These ideas have recently received a fair bit of pushback in the literature, with some investigators highlighting concerns related to publication bias (de Bruin, Treccani, & Della Sala, 2015), and others claiming not to find any evidence that bilinguals indeed reliably show any cognitive advantages (e.g., Paap & Greenberg, 2013). Paolieri *et al.* do the field a great service by providing a nuanced perspective, suggesting that immersion and language experience play a role in determining how selection operates in an individual, which in turn has implications for the extent to which any bilingual advantage will be observed.
Signing is another domain in which issues relating to production involving multiple languages arise. This is because, as Wilbur points out in his chapter, most people who communicate in sign know sign as a second language. Many of the challenges for people producing sign languages are similar to those that have been identified for spoken languages, in part because the two kinds of languages have many similarities. Wilbur notes that the prosody of sign is based on prosodic constituents that are ordered hierarchically, starting with the smallest unit, the mora, and topping out with the intonational phrase and the phonological utterance. This organization is just like what is observed for spoken languages. One prosodic domain that has been extensively studied is the production of syllables in sign, with research suggesting that although both spoken and signed languages have syllables as prosodic constituents, their internal structures differ due to the differing modes of transmission in the two modalities. In addition, whereas English permits sentence stress to be marked on any constituent within a sentence, American Sign Language (ASL) is similar to spoken languages like Italian in that sign permits only sentence-final stress. Thus, one consideration for ASL speakers generating a syntactic form is to decide how to organize the sentence so this prosodic constraint can be respected while at the same time conveying the intended semantic focus within a grammatical form. Signers also generate speech errors similar to those found in spoken languages, including word substitutions and errors involving phonetic features. Wilbur’s chapter ends with a discussion of how speech and sign are coordinated in individuals who attempt to communicate in both modalities simultaneously. Contrary to what might seem intuitive, it appears that the simultaneous production of a sign and a spoken expression is interfering (similar to what is observed for multiple spoken languages, as Paolieri et al. argue), leading to disruptions in the production of both types of linguistic forms, as would be expected from any attempt to communicate two spoken languages at the same time. One interesting advantage of sign is that signers have the ability to communicate more than one concept simultaneously—for example, two referents can be conveyed, one with each hand. A fascinating question for psycholinguistic investigation is to determine how this information is represented and executed in sign compared with speech, and to conduct experiments to discover how comprehenders efficiently process such information.

**Linking production and comprehension**

Of course, the production system does not operate in a psychological vacuum: It works with other cognitive systems, including those responsible for perception, attention, and memory. Production processes also interact with those responsible for comprehension, and vice versa. The two systems influence each other. The two chapters in this section, one by Pardo and the other by Gambi and Pickering, discuss ideas for capturing these relationships, as well as the empirical evidence concerning the details of these mutual effects. The fundamental conclusion that emerges from both chapters is that the demands of communication helped to
shape the structure of language, which in turn influence the online processes that allow speakers to efficiently generate utterances that are communicatively effective.

Pardo focuses on speaker-addressee interactions, noting the large body of research showing how speakers tailor linguistic forms to suit the addressee. She summarizes studies demonstrating coordination, entrainment, alignment, and accommodation between interlocutors. At the same time, divergence is also observed, particularly when the conversational participants are of different status or differ from each other on other traits tightly bound up with social identity. Moreover, individuals differ in their tendency to adapt in this way to their interlocutors. Pardo further makes the case that these effects challenge traditional approaches to psycholinguistics that distinguish competence from performance, and those that treat language as a system that is primarily for the transmission of information. This argument is not new; it appears that many researchers investigating these kinds of topics believe their conclusions and even the entire approach is incompatible with, say, a formal analysis of grammatical encoding. But this claim seems to me to be somewhat exaggerated. Consider, for example, Frank’s chapter on the computation of agreement during production. Is any mechanism or process proposed in that chapter inconsistent with the notion that speakers would tailor their utterances so they’re appropriate given their addressees? The answer, it seems to me, is no; it’s more a matter of whether an important topic—the tailoring of utterances to addressees—receives attention or is neglected. Pardo is certainly right to emphasize the importance of processes promoting alignment between interlocutors, and it is also clearly true that the field had for too long ignored the kinds of questions her chapter brings to the fore. Both kinds of inquiries can co-exist, and indeed must co-exist, if we are to emerge with a complete theory of the language production system.

These ideas are further delineated in the chapter by Gambi and Pickering, which focuses on models linking production and comprehension. One of their original suggestions is for the field to redefine what it means for something to be a production or a comprehension process. As they note, the traditional approach is to assume that whatever happens during production is a production process, and whatever happens during comprehension is a comprehension process. On this view, production permits feedback to the extent that we observe “lower-level” processes influencing those that originate from higher representational levels. For example, if phonological information affects choice of syntactic form during production, that is an example of feedback, and the existence of such effects motivates non-modular models. The same logic holds for comprehension, except that the interactive effects are ones in which higher levels influence lower ones (e.g., a semantic effect on syntactic parsing decisions). Gambi and Pickering’s suggestion is to abandon this approach and instead to define production processes as those that map higher-level representations onto lower-level ones, and to define comprehension as processes that do the opposite. On this view, then, the production and comprehension systems interact with each other, but the production and comprehension systems themselves are not interactive. For example, self-monitoring,
the ability of speakers to evaluate the quality of their utterance before overtly producing them, is a process that takes place during the production of an utterance but which involves looping the comprehension system in at a specific point during planning. The chapter includes a summary of Pickering and Garrod’s self-monitoring theory, which provides a specific example of this approach. Their theory also captures the phenomenon of prediction during comprehension as another example of how the production and comprehension systems work interactively (and also imply that prediction effects are not evidence for interactive comprehension systems). If as a listener I am able to anticipate your next word, it is because I have invoked my production system to model what I would say in that specific context. This proposal is consistent with evidence suggesting that an individual’s production skills correlate with that person’s ability to predict effectively during comprehension.

But perhaps the most well known contribution these models make, as Gambi and Pickering argue in their chapter, concerns the insights they provide about the fundamental nature of dialogue. As many researchers studying language production have argued, the standard psycholinguistic model that treats production and comprehension as separate systems makes dialogue somewhat of a mystery. Indeed, many researchers who focus exclusively on production have argued that production is hard, but that is not the intuition most of us have when we talk to someone—instead, our sense is that production is pretty easy, and we sometimes marvel at the way our ideas flow out as speech without our awareness of the unfolding processes and without the need for conscious planning. Indeed, Churchland (2013) in her recent book *Touching a Nerve: Our Brains, Our Selves* describes this phenomenon very compellingly, based on personal experience. She notes that not only is production usually quite easy, requiring little conscious planning, it is often precisely when we become conscious of how we are talking that we find ourselves struggling, and in these circumstances we often become disfluent as well as communicatively ineffective. Thus, twenty-first century psycholinguistic theory must explain what makes dialogue easy, at least most of the time. The answer that models like Garrod and Pickering’s provide is that it is based on rapid coordination between the production and comprehension systems, which in turn is likely grounded in humans’ ability to generate recursive models of other minds and intelligent agents.

**Themes, resolutions, and challenges**

As I hope this overview makes clear, the chapters in this section on language production lay out some exciting, important new perspectives. At the same time, notably absent from this section is any chapter discussing the processes that support the generation of prosodic forms during production. The rich interplay among semantic, syntactic, and prosodic sources of information is not addressed in these discussions, which is unfortunate. The fault, however, is not with the editors of the volume but rather the lack of interest in the topic in the field more generally. In my own view as someone who has worked on this issue,
the questions and perhaps also the answers are simply not provocative enough for a field that has far too often exaggerated theoretical distinctions as a way of generating controversy. But if there is one basic fact about language production, it’s that speakers generate a prosodic form each and every time they utter even a single syllable. They mark syntactic and semantic structure, and they even mark discourse constituency using prosodic features such as pitch and intensity (Tyler, 2013, 2014). This issue, then, should be the target of active investigation.

Another area in which the field has not made enough progress, in my view, is in developing clear, specific theories about exactly how referential forms are chosen. There is research demonstrating that people take into account the needs of their listeners, respond to immediate feedback, and so on, and much of that research is summarized in the chapters in this section. But what exactly are the mechanisms that support these abilities? On this question, we have little information. Another example: Decades ago, Levelt (1982) conducted a series of clever experiments using simple figures consisting of connected colored circles to assess speakers’ ability to make macro decisions about how to structure a discourse, and how they keep track of what had been said and what still needs to be communicated. Levelt gave the example of describing a house or an apartment, which requires the speaker to decide where to begin and how to proceed when there is a conceptual choice point (e.g., when a hallway splits off into two wings). The speaker also must keep track of what has been described and what has not. This work showed that speakers attempt to minimize their memory load, beginning with the discourse segment that is shorter and less complex. This strategy enables them to plan the longer, more complex segment during articulation, as we showed in our own work following up Levelt’s (Ferreira & Henderson, 1998). We also suggested that both speaker and listener benefit from this strategy because both need to use working memory resources as efficiently as possible. Unfortunately, we still know about as much concerning discourse planning today as we did 30 years ago, suggesting that the topic is under-investigated.

Continuing with this theme, it appears that we have made a great deal of progress in understanding production since the days of the speech error models developed by Garrett and Fromkin in the 1970s. Our experimental methods permit us to isolate specific bits of the production process and determine the factors that influence it, whether the process is computing agreement, managing multiple languages, or coordinating a conversation. The availability of huge corpora makes it possible to conduct large-scale data analyses of very specific phenomena, including things like phonological reduction. But perhaps we are now missing something that those global speech error models gave us, and that was a road map for the entire system. A researcher might focus on segmental speech errors, but given the constant backdrop of the global models, discussions would ultimately come back around to the big questions concerning the architecture of the production system itself. Now, the connections from specific empirical phenomena to global models of production are somewhat less clear. For example, how does alignment promote communication and influence phonological reduction? More importantly, as researchers in this area, what do we now believe
about the overall structure of the language production system, from discourse planning all the way down to segment retrieval and articulation? Some might say that the original speech error models no longer hold up given findings from controlled experiments or corpus analyses, which is a reasonable point. However, it would be useful if today’s researchers would try to come up with some sort of alternative models that have the same scope and ambition as the ones our pioneers gave us decades ago. Psychology is already known for being a field that sometimes seems too focused on techniques and effects at the expense of theories and mechanisms. Psycholinguistics should be an exception given its rich theoretical history.

At the same time, the study of language production has clearly advanced in many significant ways. Investigations of topics such as multilingualism, signing, and comprehension-production interactions are genuinely novel and exciting. Much more is known also about more traditional issues such as syntactic planning. There is no doubt that there has been a major increase in methodological and statistical rigor. My plea is simply that we not turn our backs on the twentieth century as our field continues to progress, but that we build on previous insights and ideas as we continue to investigate production in the twenty-first century and beyond.

REFERENCES


Introduction

When Nim, the chimpanzee raised at Columbia University, produced one of his longest sentences: *Give orange me give eat orange me eat orange give me eat orange give me you*, he succeeded in conveying his desire to be given an orange and to eat it. However, his sentence could, in principle, have had several other meanings: for example, that he was eating the orange that was given to him by his interlocutor, that he wanted an orange to give to his interlocutor, or that his interlocutor has given him an orange that is now eating him. These alternative interpretations can reasonably be ruled out by our knowledge of the world and the situation (oranges don’t eat monkeys). Nevertheless, they all seem equally compatible with the concatenation of words that Nim produced. Syntax is the component of language that allows us to express ideas with as little indeterminacy as possible, by constraining dependencies between words in the sentences. It does so by two major devices: constraints on the ordering of the words and grammatical morphemes linking words together. Nim was able to develop an idea, learn signs to express the various pieces of that idea, and concatenate them to convey it within a (rudimentary) social interaction. This is an amazing achievement, demonstrating that he was able to communicate. But Nim never managed to develop syntactic knowledge, necessary to express his ideas with the precision that human language users routinely exhibit. Systematic failures from the various programs dedicated to teaching syntax to animals stand in sharp contrast with the ease with which humans deal
with the syntax of their sentences. Within three years, little humans master most of the syntax of their mother tongue. Children spontaneously create syntactic devices when the input language is impoverished (like pidgins or first generations of newly created sign languages; Bickerton, 1984; Senghas & Coppola, 2001; Singleton & Newport, 2004). Adults produce sentences almost flawlessly, with less than 1 error every 1000 words (Levelt, 1989), and the great majority of these errors preserve the syntactic well-formedness of the sentence.

Despite broad agreement that grammar is what makes human language different from other communication systems (Hauser, Chomsky, & Fitch, 2002) and despite its intriguingly powerful machinery that by far surpasses that of lexical processes (most errors in production are lexical errors), psycholinguistic research on language production has for the most part focused on single words. The reason is no doubt the challenge of grasping the abstract, relational nature of syntactic structures. Psycholinguists interested in syntactic encoding face two major questions. First, how can we characterize the syntactic representations underlying the sentences that speakers build? This question has to do with the shape of these representations, and whether they look like the formal hierarchical structures proposed in syntactic theory. Second, how can we characterize the processes that deal with these representations? This question relates to the identification of the functional components involved in syntactic encoding, and the relationships between them, that is, relations between lexical and syntactic processes, and between syntactic and non-syntactic levels of representation.

This chapter addresses these two major questions through the lens of studies of agreement production. Agreement is a syntactic phenomenon par excellence: it ties words together in virtue of their syntactic status, for the most part independently of their semantic or morpho-phonological content. The majority of natural languages have agreement constraints, which may involve features like number, gender, and case. Through several properties, agreement offers an especially revealing perspective on our two questions. The first property is that agreement is structure-dependent. In many languages the verb agrees in number with its subject and sometimes with its object, but there is not a single language in which the verb agrees with, for example, a noun inside a clausal modifier (e.g., *The goat that ate the radishes are mean). That is, not any element in the sentence may enter into an agreement dependency. The second property of agreement is that it shows, to some extent, autonomy from the semantic representation of the words that constitute sentences. We are perfectly able to produce correct agreement in a sentence that does not make sense like The subversions of the boy are genuine or even in a Jabberwocky sentence with senseless words like Which rabun did you say that the livols are eating? The autonomy of agreement with respect to other levels of representation makes it possible to investigate the key question of the modularity of syntactic encoding. Moreover, agreement is at the crossroads of lexical processes (features are retrieved from the lexicon) and syntactic processes (features are transmitted to the targets in virtue of their structural positions), allowing us to explore the relations between lexical and syntactic processes involved in syntactic encoding.
Agreeing units can be contiguous in the sentence, or far apart, in the same clause or in separate clauses; yet, speakers as early as age three (Keeney & Wolfe, 1972) are able to produce multiple agreement dependencies within a single second, without effort, and usually without errors (younger children tend to produce singular verbs as “default,” that is, singular verbs in the context of a plural subject, but they do not produce erroneous plural verbs in the context of a singular verb, which shows that they do not produce verbal agreement morphology at random; Clark, 1998). Nevertheless, agreement computation sometimes fails. Following the long tradition of spontaneous speech error research, Bock and Miller (1991) initiated a fruitful line of experimental work on a particular kind of errors called attraction errors. Attraction refers to the erroneous agreement with an element that is not the agreement controller, as in *The key to the cabinets are rusty, where the verb are incorrectly exhibits plural number marking as though it were agreeing with the plural modifier, cabinets, rather than with the singular head key. Identifying the conditions that modulate attraction errors has turned out to be an extremely rich way to explore both the shape of syntactic representations and the processes that build them. In about two decades of experimental research, an extensive body of evidence has accumulated showing influences from semantic, syntactic, morphological, and morphophonological properties of the words and sentences. The prominent Marking and Morphing model (M&M) of agreement attraction (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001, further developed in Eberhard, Cutting, & Bock, 2005) is founded on this empirical work. It employs the core features of Garrett’s seminal model of language production with a modular, two-level functional architecture in which semantic information penetrates the first level only, while the second level deals with structural and morphological information (Garrett, 1975, 1980, 1989). As such, it provides a representative illustration of what has become the standard psycholinguistic approach to syntactic encoding, and it allows us to see both the strengths and weaknesses of its assumptions.

The aim of this chapter is to establish a new pathway to the analysis of syntactic representations and processes involved in agreement production. The framework, Feature and Controller Selection, takes insights from both syntactic theory and the psychological theory of memory retrieval in order to capture the wide empirical range of attraction effects that have been observed. Like Marking and Morphing, the model cuts the pie into two parts, but in a fundamentally different way. The first process, Feature selection, retrieves functional units from the long-term memory lexical store. It is responsible for selecting the grammatical features associated with the nouns in the sentence. This process crystallizes several of the semantic, morphological, and morphophonological influences reported in research on agreement errors. It operates under the guidance of multiple, statistically distributed cues, in line with interactive models of production (Dell, 1986; Goldrick, 2006). Interestingly, most of the data points from the attraction literature focus on this fundamentally lexical component. Even though these effects are manifest in the context of an attractor element, I argue here that they arise on top of attraction, independently of it. In that regard, they only represent the cherry on the pie. The
pie, that is, the core syntactic component of agreement, is the second process, Controller selection. This process is responsible for retrieving the controller in the sentence in order to copy its features onto the target. It is the locus of what I consider to be “attraction proper,” conceived of as the incorrect identification of the attractor as the controller, in line with the initial suggestion of Badecker and Kuminiak (2007). I suggest that by positing a process of controller selection, and properly distinguishing it from feature selection, we can unite a range of syntactic effects showing the sensitivity of attraction to major syntactic constructs (intervention, movement, c-command, hierarchical depth), as well as another set of semantic and morphological effects, different from those arising during feature selection, and that appear to lie in the similarity between the attractor and the controller. The syntactic, semantic, and morphological effects arising at this level are argued to reveal the inner workings of a cue-based retrieval process (Lewis & Vasishth, 2005) in which these factors act as cues to retrieve the controller.

In the first section of the chapter, I sketch the standard model of syntactic encoding in psycholinguistics (Garrett, 1975) and the M&M model of agreement developed by Bock and her colleagues, which instantiates one of the most elaborate illustrations of the standard approach. In the second section, I lay out a typology of attraction effects, organized in terms of my proposed Feature and Controller Selection account. In the third section, I present new evidence showing a close alignment between syntactic structure and memory retrieval in attraction effects, and suggest that syntactic theory describes the strength of memory representations. The chapter ends with some challenging issues for future research.

**Syntactic encoding in the M&M of agreement**

The approach of sentence production initiated by Garrett in the 1970s assumes that the syntax of the sentence is encoded at two separate, successive levels, each of them responsible for a set of lexical and syntactic processes (Garrett, 1975, 1980, 1989). The Functional level, which executes first, ensures the retrieval of words and the construction of a functional, hierarchical structure specifying the words’ syntactic role (like subject, object) and the relations between them. The Positional level, which executes second, is responsible for retrieving word forms and inserting them within a frame of grammatical morphemes (like determiners, inflections) arranged in the linear order in which words will be pronounced. The major evidence in favor of this model comes from the observation that speech errors are distributed in two broad classes. The first class involves whole words (e.g., *She sings everything she writes*; Fromkin, 1971). In the great majority of these errors, the words exchanged are of the same grammatical category, they show no phonological similarity, and they may be part of a rather large unit (e.g., the clause). The second class of errors involves units smaller than words. Most of these errors involve phonemes (e.g., *plit spea soup*; Fromkin, 1971). They do not respect the grammatical category of the words; rather, they respect phonological and prosodic constraints (like syllable position or the vocalic versus consonantal status of the
sounds). A smaller subset of these sublexical errors, so-called stranding errors, involve the exchange of two lexical roots while the closed class morphology of the words is left in place (e.g., *That's why they sell the cheaps drink*; Garrett, 1989). Like phonological errors, stranding errors appear to follow sound-level constraints; they involve words of different grammatical categories, lexical roots are often phonologically similar, and they usually arise within smaller units (phrases). Garrett suggested that the first class of errors arises during the operation of the Functional level, when words are retrieved and assigned a syntactic function, while the second class arises during the operation of the Positional level, when words are specified both morphologically and phonologically. In Garrett’s view, the production system is fundamentally sequential: the Functional level takes as input semantic information from the Message, which provides the content the speaker is wishing to convey. The Positional level takes the output of the Functional level as its input, and it sends its output to the articulatory system responsible for encoding the surface phonetic forms of phrases. The strict seriality of this framework is responsible for its modularity: semantic information only penetrates the Functional level while the Positional level is immune to direct conceptual influences (level n-1 influences level n, not level n+1), and the Functional level is immune to the morphological and phonological specifications that are only specified subsequently at the Positional level (level n+1 cannot influence level n).

Nearly four decades of experimental research has not challenged the major claims of Garrett’s account: the separation of the system in two functional levels, and their serial order with the resulting modular architecture (although see for example, Dell, 1986, for an alternative, interactive approach). I now consider how these two properties are implemented in the M&M model of subject-verb agreement production (Bock et al., 2001; Eberhard et al., 2005). The description summarizes the major assumptions of the model and the types of data it explains, and then highlights the challenges the model is confronted with.

**Two functional levels: Marking and Morphing**

The model involves two functional components. Marking takes place in the syntactic component of the first level of Functional assembly. It is responsible for translating the number notion from the Message into a linguistic feature. At the level of the message, a process of notional number valuation takes place by which notional singulars are distinguished from notional plurals in the speaker’s reference model. Although notional number is continuous in that entities can be conceived as more or less single or multiple, Marking receives the output of number valuation and translates it into the selection of a syntactic feature (singular or plural). The site of Marking is not the subject head but the root of the whole subject phrase. At the same level, lexical processes also take place. These are responsible for recruiting nouns with meanings that are consistent with the notional number of the corresponding concepts. For example, *clothing* and *clothes* equally express a notion of multiplicity, and are therefore equivalent options for the number Marking process: their grammatical properties differ, but these grammatical properties only
come into play during the next stage, Morphing. Morphing takes place at the level of Structural integration, which binds together lexical forms (morphemes), and structural forms (the hierarchical representation of the sentence). Morphing reconciles the syntactic features selected during Marking and number specifications from the lexicon, which are argued to percolate the tree up to the subject root, where reconciliation takes place. Morphing also ensures that the feature selected by the reconciliation process will be transmitted to the agreement target. The model adopts two other assumptions. One assumption is that single count nouns are unspecified, or weakly specified, for number. As a result, only plural nouns have the possibility to percolate the tree and enter into the reconciliation process. The other assumption is that if an inconsistency is encountered between number marking and morpheme specifications (e.g., in collectives, like *army*), the morpheme specifications prevail.

**Serial order: Marking before Morphing**

The model adopts the standard assumption of seriality, such that Marking takes place before Morphing. This architecture has two consequences for the information flow in the system. The first consequence is that Morphing is insensitive to semantic information from the message: semantic information only penetrates the first stage of Marking. The second consequence is that Marking is insensitive to morphological information, which is specified after its output has been sent to Morphing.

This architecture accounts for three major sets of facts: some semantic effects on agreement, the asymmetry between singular and plural attractors, and the profile of differences and similarities between verb and pronoun agreement. The semantic effects considered for the model have their locus in the Marking component. Notionally plural but grammatically singular subjects (e.g., a collective noun like *army* or a distributed subject like *The label on the bottles*) trigger the selection of a plural marking. However, they carry a singular morpheme specification. These two features have to be reconciled during Morphing, and the stronger power of morphological specification is such that the plural marking feature will be overridden most of the time by the more powerful singular morphological specification. Still, on some occasions, the plural feature will win, giving rise to plural agreement on the verb. Attraction results from the contamination from the attractor’s feature of the subject node during the process of reconciling number marking and number specification. Morpheme specifications anywhere in the structure have the potential to percolate, but their influence depends on the structural proximity of the attractor to the subject’s maximal node, that is, the locus of agreement control. In this way, the model correctly predicts that the head noun’s feature will usually be the controller of agreement, given its privileged structural position, closest to the maximal node, involving the smaller percolation path. The finding that most attraction effects, at least in English, arise from a plural attractor while virtually no attraction arises from singular attractors results from the lack of (or weak) morpheme specification of singulars: only plural, morphologically
specified features have the potential to percolate and therefore influence the reconciliation process. The model accounts for the finding that pronouns are more sensitive than verbs to the notional number of the subject (Bock, Eberhard, & Cutting, 2004) by the fact that pronouns receive their number from the semantics, through Marking, whereas verbs’ number is assumed to be semantically empty, such that they can only receive their agreement feature through Morphing. The finding, in contrast, that pronouns and verbs are equally sensitive to the presence of a plural attractor, and equally insensitive to the notional plurality of the attractor, is explained by the fact that attraction takes place at a stage that is common to both pronoun and verb agreement, which is sensitive to morpheme specifications but insensitive to notional representations: the stage of Morphing.

Although M&M has the merit of accounting for a wide range of data points reported in the literature, the model is challenged by two major issues. The first issue concerns the locus of semantic influences. In M&M, semantics only affects agreement by way of its influence on the selection of the agreement feature for the whole subject phrase via Marking. This property of the model fails to account for the finding that semantic information that has no consequence for the whole subject noun phrase also influences agreement. For example, more attraction is found when the subject head noun and the attractor noun semantically overlap (Barker, Nicol, & Garrett, 2001) or when the attractor is a plausible semantic subject (Hupet, Fayol & Schelstraete, 1998; Thornton & MacDonald, 2003). The model does not capture either effects found in gender agreement like the finding that heads with semantic gender are more resistant to attraction than those with grammatical gender (Vigliocco & Franck, 1999, 2001). This shows that notional gender information that is tied to the head itself, but critically not to the subject phrase, also affects agreement. Hence, semantic influences are found outside of the realm of the Marking process, which can therefore not be considered as the sole locus of semantic influences on agreement. In the alternative model proposed in the next section, I suggest decomposing semantic effects in three types: effects due to the notional valuation of the subject phrase (in line with Marking), effects due to the feature stability of the nouns (features with semantic correlates are more stable), and effects due to the semantic similarity between the attractor and the head. Whereas the first two effects are argued to have their loci in one component of agreement production (Feature selection), the latter has its locus in another component (Controller selection).

The second issue concerns the underspecification of the structural conditions over which Morphing takes place (Franck, 2011). In M&M, attraction is a function of the structural distance between the attractor and the subject node. Despite the critical role that structural distance is assumed to play, the theory does not provide a description of what the structure looks like, let alone a tool to measure structural distance. In contrast to Garrett’s early model which represented hierarchical structure at the first level of encoding (Garrett, 1989), that is, at the level of Functional assembly in M&M, current models assume that hierarchical structure is built at the second level of encoding, that is, during Structural integration in M&M (e.g., Bock & Levelt, 1994; Bock & Ferreira, 2014; Eberhard et al., 2005). In these models,
syntactic units are directly assigned to their surface hierarchical position. This position is based on evidence by Bock and colleagues (Bock, Loebell, & Morey, 1992) who argued that the hierarchical structure of a passive sentence is constructed with the patient in the subject position right away, without transiting through a deep hierarchical structure in which it occupies the position of complement of the verb, as assumed in movement-based linguistic theory (Chomsky, 1981). On this view, the speaker builds a single hierarchical structure, at the same stage as the linear structure, while the representation of the first stage is reduced to the “flagging” of the units for a particular syntactic function, assigned on the basis of the message. The consequence for the analysis of attraction patterns is that if attraction occurs at the second stage of Structural integration, it is predicted to occur in the same way for structures that have identical surface hierarchical structures, and differently for structures that have different surface hierarchical structures. Empirical evidence casts doubt on that prediction. For example, significant attraction is found from a plural subject modifier in the interrogative structure (e.g., *Are the helicopter for the flights safe?; Vigliocco & Nicol, 1998), whereas no attraction is found in the superficially identical structure in Italian involving free inversion (e.g., Telefonera l’amica dei vicini, Will-phone-Sg the friend of the neighbors, Franck et al., 2006). Similarly, the moved object of the target verb triggers attraction (patients in Jean parle aux patientes que le médicament guéris-sent, John speaks to the patients that the medicine cures-Pl), while the object of the main verb situated in the same surface position does not (patients in Jean dit aux patientes que le médicament guérit, John tells the patients that the medicine cures-Sg, Franck et al., 2010; Franck, Colonna, & Rizzi, 2015). That is, two structures that are superficially identical but have different underlying structures show different attraction profiles. Moreover, two structures that are superficially different but have identical underlying hierarchical structure show similar attraction profiles. For example, the English interrogative structure generates similar attraction to the corresponding declarative (*The helicopter for the flights are safe), despite their different surface structures. These data suggest that it is not the properties of surface hierarchical structures that account for attraction profiles but rather properties of their underlying hierarchical organization.

In the next section, I sketch the alternative model of Feature and Controller selection that makes use of the fine constructs from linguistic theory to describe underlying hierarchical structures and capture these syntax-based attraction patterns. The model aims at capturing a wide range of attraction effects and opens new windows to the understanding of syntactic encoding and its relation to both syntactic theory and the theory of memory processes.

**Syntactic encoding in the Feature and Controller selection model of agreement**

The model involves two functional processes: Feature selection, which retrieves nominal features, and Controller selection, which selects the agreement controller that will transmit its features onto the target. In this model, attraction does not
arise because of the incorrect percolation of a feature into the tree, but because the attractor has incorrectly been selected as controller and as a result its features have incorrectly been transmitted on the agreement target. The model is an extension of the Selection and Copy model proposed in our previous work (Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Franck, 2011). I suggest that two classes of factors influence Feature selection: factors that modulate the lexical stability of the head and attractor features (depending on the strength of their association to semantic, morphological and morphophonological correlates) and factors that modulate the notional representation of the subject (lying in properties of the head or its relation to the attractor). Following Badecker and Kuminiak (2007), Controller selection is argued to be a retrieval mechanism operating on the basis of cues. It is modulated by the similarity between the attractor and the controller at the semantic, morphological and syntactic levels.

In the following section, the various factors that have been shown to modulate attraction are organized into classes and subclasses, in the manner of a typological classification. The highest classes are those defined by the two functional processes, Feature Selection and Controller Selection. The subclasses are defined by the nature of the factors they involve. Note that I only consider here studies manipulating agreement in the context of an attractor word. A growing set of empirical evidence showing grammatical modulations of agreement production in the context of sentences with no (or no clear) attractor should ultimately be incorporated into the picture, and the question of their relation to attraction should be discussed (like constructions involving conjunctions, disjunctions, pseudo-partitives, or quantified noun phrases, e.g., Haskell & MacDonald, 2005; Haskell, Thornton & MacDonald, 2010; Mirković & MacDonald, 2013; Marušič, Nevins & Badecker, 2001; Marušič, Nevins & Saksida, 2007; Smith, Franck & Tabor, 2016).

**Effects on Feature selection**

Feature selection is the process responsible for selecting the grammatical features of the nouns in a sentence. It is fundamentally a process of lexical retrieval by which nominal features are selected from the functional lexicon. The process shows the property of interactivity widely reported in the literature on lexical retrieval of content words, in that it is influenced by semantic and form information (e.g., Dell, 1986; Goldrick, 2006). Two types of factors affect feature selection. The first factor is *feature stability*. If a grammatical feature is regularly associated with converging semantic and/or form (morphological or morphophonological) information, it is more stable and has more chance to be selected. As a result, converging correlates of the head’s feature back up its grammatical feature and thus reduce the risk of an agreement error. In contrast, if these correlates diverge from the grammatical feature, that is, point to the opposite direction, the feature is less stable and more susceptible to a selection error. Feature stability similarly influences the selection of the attractor’s feature; however, its effect on agreement is diluted given that it only shows up if the attractor is incorrectly selected as
controller. The second factor influencing feature selection is the *notional representation of the subject phrase* at the message level. The conceptual representation of the numerosity of the phrase depends on factors like the distributivity of the subject, the semantic integration between the head and attractor nouns or the spatial distribution of units, which all have the potential to influence the number feature that will eventually be selected.

The key difference with M&M is that whereas M&M attributes semantic and formal effects to two separate functional components of agreement, respectively Marking and Morphing, the semantic and form effects grouped in this first category all arise at the level of the same functional component of Feature selection. Moreover, whereas in M&M form effects are intrinsically linked to attraction, the current model assumes that the semantic and form effects affecting Feature selection are independent of attraction proper (by “attraction proper,” I mean the erroneous selection of the agreement controller), although their influence is sometimes only detectable in the context of an attractor noun, either because the attractor directly modulates the notional representation of the subject phrase or because it indirectly boosts error rates, allowing for these factors to show up. In other words, the effects of the factors listed here arise on top of attraction.

**Effects of feature stability**

**Semantic stability**

Semantic correlates boost agreement errors when they diverge from the grammatical feature. For example, more plural verb agreement is found with grammatically singular collective heads denoting a plural entity (e.g., *The cast in the weekend performances*) than with notionally singular nouns (*The actor in the weekend performances*, e.g., Bock, Nicol, & Cutting, 1999; Bock et al., 2004; Haskell & MacDonald, 2003). Similar effects are found for gender agreement with epicene nouns that have a fixed grammatical gender but can refer either to a feminine or to a masculine entity. Speakers produce more erroneous masculine agreement on the predicative adjective when a grammatically feminine epicene head (e.g., *La victime*, The victim-F) refers to a man than when it refers to a woman (Vigliocco & Franck, 2001). In contrast, semantic correlates reduce agreement errors when they provide converging information to the grammatical feature. For example, head nouns with semantic gender (e.g., *La jument*, The mare-F referring to a female horse) give rise to fewer agreement errors than those with a purely grammatical gender feature (e.g., *La méduse*, The jellyfish-F; Vigliocco & Franck, 1999). Heads with regular plurals (like *bubbles*) generate fewer erroneous singular verbs than invariant plurals (like *suds*) (Middleton & Bock, 2004). Regular plurals are judged conceptually more plural than invariant plurals, suggesting that the presence of a clear semantic correlate of plurality backs up the grammatical feature (Haskell & MacDonald, 2003). Finally, the observation that gender attraction tends to be weaker than number attraction (e.g., Eberhard et al., 2005; Lorimor et al., 2008) may also be related to the fact that grammatical gender lacks semantic correlates. When
Syntactic Encoding

semantic correlates are manipulated on the attractor nouns, the same factors turn out to have a much weaker effect (e.g., Bock et al., 2004; Deutsch & Dank, 2009; Haskell & MacDonald, 2003); the reason is that for these effects to show up, the attractor first needs to be incorrectly selected as controller, which only arises in a small portion of the sentences produced.

Morphophonological stability

Several studies across various languages (Italian, Spanish, French, Dutch, German) have shown that the strength of the association between the controller’s feature and its morphophonological realization modulates agreement production. Heads with nominal endings or determiners that carry converging morphophonological cues (e.g., nouns ending in –o in Italian, which are usually masculine) are less prone to attraction than heads lacking these cues (nouns ending in –e in Italian, which can be masculine or feminine, see also Hartsuiker, Schriefers, Bock, & Kikstra, 2003; Vigliocco et al., 1995). Heads carrying morphophonological information diverging from the grammatical feature are particularly sensitive to attraction (e.g., a masculine noun ending in –a in Spanish, although most nouns ending in –a are feminine, Franck et al., 2008). Again, when manipulated on the attractor noun, the same factors typically show either a weaker effect on agreement, or no effect at all (e.g., Bock et al., 2001; Bock & Eberhard, 2003; Hartsuiker, Anton-Mendez, & van Zee, 2001; Hartsuiker et al., 2003; Meyer & Bock, 1999; Vigliocco et al., 1995).

Morphological stability

One major consistent finding of the number attraction literature, cross-linguistically, is that attraction is often stronger in sentences with a singular head and a plural attractor than in sentences with a plural head and a singular attractor. Similarly, attraction from a gender mismatching noun appears to be stronger in sentences with neutral heads, followed by masculine heads and finally by feminine heads (e.g., Badecker & Kuminiak, 2007; Malko & Slioussar, 2013). These asymmetries have been classically interpreted as stemming from the morphological markedness of plural and possibly also feminine attractors. According to M&M, singular nouns carrying no feature do not have the potential to percolate and therefore be erroneously transferred to the verb. That is, the asymmetry is explained by the properties of the attractor noun. However, these asymmetries may also lie in the markedness of the head, since it systematically co-varies with that of the attractor in conditions where attraction can arise (that is, when the two nouns have mismatching features). In other words, the asymmetry may stem from the fact that plural (or feminine) heads, being marked, are more likely to be selected. One argument in favor of that account comes from the finding that semantic and morphophonological correlates of number or gender (described in the next sections), which do not systematically co-vary on the head and attractor nouns as is the case of markedness, show much clearer effects when manipulated on the head than when manipulated on the attractor. Another argument comes from the fact that languages have grammatical structures like pseudo-partitives, superficially
similar to prepositional phrase modifiers, but which nevertheless take plural verbs (e.g., *A bunch of people is/are demonstrating*). Haskell, Thornton, and MacDonald (2010) estimated that, in English, complex subjects with a singular first noun and a plural second noun take plural agreement in more than 20% of the cases, while subjects with a plural first noun and a singular second noun show less than 3% singular agreement. These authors reported experimental evidence that the grammatical plural in pseudo-partitive constructions primes attraction errors in constructions with PP modifiers (e.g., *A cluster of reporters were… primes *The pencil in the gift bags were…*). In sum, it seems plausible that markedness effects do not arise from the markedness of the attractor (and thus have nothing to do with attraction per se), but rather from the markedness of the head: a marked feature on the head would be stronger/more stable, increasing the chances that it is correctly selected.1

**Effects of the notional representation of the subject phrase**

**Distributivity**

The presence of a plural local noun in some structures sometimes forces the interpretation of the subject as distributed (e.g., *The label on the bottles*). Speakers produce more plural verbs in sentences containing distributive subjects, as compared to sentences with non-distributive ones (e.g., *The key to the cabinets*) (e.g., Eberhard, 1999; Foote & Bock, 2012; Vigliocco, Butterworth & Garrett, 1996; Vigliocco, Butterworth & Semenza, 1995; Vigliocco, Hartsuiker, Jarema & Kolk, 1996). Distributivity can also be a function of the preposition; in *The gang on the motorcycles*, *on* promotes a distributive reading in which each member of the gang seems to be understood as capable of independent action, whereas in *The gang near the motorcycles*, *near* promotes a more collective reading, where the gang members are viewed as a unit. Again, more plural verbs were found with distributive subjects (Humphreys & Bock, 2005).

**Semantic integration**

This factor refers to how closely the head noun and the attractor are linked in the semantic representation of the sentence. For example, in *The drawing of the flowers* a particular, integrated relation of the two referents (drawing and flowers) is implied, whereas in *The drawing with the flowers* the relation is a very generic, non-integrated relation of juxtaposition. The initial set of experiments on this factor showed that speakers tend to produce more plurals with semantically integrated subjects (Solomon & Pearlmutter, 2004), supposedly because in integrated subjects, the head and attractor nouns are more likely to be encoded together (but no timing measure was provided to back-up that claim). However, subsequent studies made contradictory claims, suggesting that non-integrated subjects were actually
more likely to be interpreted as referring to distinct entities, i.e., they are more individuated, and therefore more likely to give rise to more plural agreement than integrated subjects (e.g., Brehm & Bock, 2013; Veenstra, Acheson, Bock, & Meyer, 2013). A clearer theoretical approach of the semantics of these structures is clearly needed to shed light on the reason for these inconsistencies. Nevertheless, these observations show that the way a speaker represents the numerosity of the subject has an effect on the feature that will be selected on the controller.

**Spatial distribution**

Visual cognition research has shown that visual arrays occupying more space are perceived as containing more items. When the head noun of a quantified phrase (e.g., *Each alligator with humungous claws*) is illustrated with the constituent elements (alligators) spread far apart from one another, speakers tend to produce more plural agreement than when the same sentence is illustrated with a more condensed spatial distribution (Brehm, 2015).²

**Effects on Controller selection**

Controller selection is the process responsible for selecting the controller whose features will be copied onto the agreement target. Badecker & Kuminiak (2007) suggested that attraction reveals the incorrect selection of the attractor as controller, via a cue-based retrieval process triggered by the verb (a similar, though different, proposal in comprehension has been proposed in Wagers et al., 2009 and subsequent studies³). Here, I adopt this hypothesis, and suggest that various effects reported in the literature actually show the signature of a cue-based process: similarity-based interference. In this view, attraction errors are similarity-based interference errors; they arise because there is an element in memory bearing some similarity to the controller, such that it is selected for agreement computation. Experimental work suggests that an element triggers stronger attraction if it is similar to the head semantically (in terms of animacy, semantic overlap, and thematic roles) or morphologically (in terms of case marking). I will suggest here that some of the syntactic modulations of attraction reported in the literature may also be interpreted as syntactic similarity effects: attractors in a syntactic position typically occupied by agreement controllers (c-command and hierarchical height) trigger more attraction than those in a position that is not occupied by controllers.

**Effects of semantic similarity**

**Animacy**

Attraction is stronger when the head and the attractor have the same animacy feature (e.g., *The blackboard behind the desks*) than when they differ in animacy (e.g., *The blackboard behind the teachers; Barker et al., 2001*).
**Semantic overlap**

Attractor nouns with high overlap of semantic features with the head, that is, semantically similar to it (e.g., *The canoe by the sailboats*) trigger more attraction than those with lower overlap (e.g., *The canoe by the cabins*; Barker *et al*., 2001).

**Thematic roles**

Attraction is stronger when the attractor noun is a plausible thematic agent for the verb (e.g., *The album by the classical composers was praised*) than when it is not (e.g., *The album by the classical composers was played*) (see also Hupet, Fayol, & Schelstraete, 1998). Along the same lines, the rate of plural agreement found in pseudo-partitive constructions (e.g., *A subset of problems are resolved*) increases with the relative topicality of the attractor with respect to the head (Smith, Franck, & Tabor, 2016). These findings show that an attractor that is a good topic is more likely to be selected as controller.

**Effects of case marking similarity**

Probably the most prominent factor affecting attraction is the case ambiguity of the noun phrases: virtually all studies reporting attraction involve controller and attractor nouns that lack morphological marking of syntactic roles, either because the language lacks case markers (in English and many of the languages tested), or because case markers are present but ambiguous, which happens when the attractor has nominative case like the head (Badecker & Kuminiak, 2007; Hartsuiker *et al*., 2001, 2003). Attraction is virtually nonexistent when the head and attractor are distinctly case-marked (Badecker & Kuminiak, 2007; Lorimor *et al*., 2008; Malko & Slioussar, 2013; Marusic *et al*., 2013). These findings suggest that attractors that are more controller-like in terms of case marking trigger more attraction than those that are less similar to controllers.

**Effects of syntactic similarity**

**C-command**

C-command refers to a particular configuration of two nodes in the hierarchical structure: *X c-commands Y* if it has a sister node that dominates *Y*. C-command plays a crucial role in agreement in that agreement only takes place with a c-commanding head (Chomsky, 2000). Thus, c-command is a property of controllers. Experimental evidence shows that attractors occupying a position of c-command trigger more attraction than those occupying a position of precedence, which is not a position occupied by controllers. The plural accusative clitic *les* in French, which is in a position c-commanding the verb, triggers more attraction (e.g., *Le professeur les lisent*, *The professor them-Pl read-Pl*) than the plural dative clitic...
leur, which precedes the verb (e.g., *Le professeur leur plaisent, *The professor to-them-Pl please-Pl). Moreover, error rates with dative clitics are similar to those with prepositional phrase modifiers (e.g., *Le professeur des élèves lisent, *The professor of the students read), which also occupy a position of precedence to the verb (Franck et al., 2010). Similarly, in sentences with moved complex objects that contain a head and a prepositional phrase moderator, the c-commanding head triggers more attraction (patientes in *Quelles patientes du médecin dis-tu que l’avocat défendent? *Which patients of the doctor do you say that the lawyer defend?) than the modifier that precedes the verb (patients in *Le médecin de quelles patientes dis-tu que l’avocat défendent? *The doctor of which patients do you say that the lawyer defend?) (Franck et al., 2015).

Hierarchical height

Agreement controllers typically occupy a high position in the hierarchical structure. Studies have shown that when the subject contains two embedded prepositional phrase modifiers, attraction is stronger with the modifier situated high in the tree (programs in *The computer with the programs of the experiment are broken) than with the one situated low (experiments in *The computer with the program of the experiments are broken) (Franck, Vigliocco, & Nicol, 2002; Gillespie & Pearlmutter, 2011). The two modifiers occupy a position of precedence with respect to the verb, however, one may entertain the possibility that hierarchical height is a proxy to c-command, since c-commanding elements are higher than preceding elements.

Syntactic structure and memory in attraction: Evidence for a close alignment

The cue-based memory retrieval process underlying Controller selection is assumed to take place on the hierarchical representation of the sentence, such that it is tightly constrained by it. The conception of hierarchical structure adopted here critically differs from syntactic encoding models like M&M in that multiple hierarchical representations are assumed to be encoded successively as elements progressively move to reach their final, surface position (a detailed description of the linguistic formalism as well as illustrations of the hierarchical representation of the structures manipulated can be found in Franck, Frauenfelder, & Rizzi, 2007; Franck, Lassi, Frauenfelder, & Rizzi, 2006; Franck, Soare, Frauenfelder, & Rizzi, 2010). Experimental support to this conception comes from the findings, discussed earlier in the critical analysis of M&M, that two structures that are superficially different but have identical underlying hierarchical structure show similar attraction profiles, whereas two structures that are superficially identical but have different underlying structures show different attraction profiles. Additional evidence comes from the various reports ever since Bock and Miller (1991) of attraction from a moved object (e.g., Franck et al., 2006, 2010; Santesteban, Pickering, & Branigan, 2013; Staub, 2009, 2010).
Such an effect is at first glance unexpected since the subject and the verb are linearly contiguous in the structures tested. Nevertheless, formal syntax has argued that the object leaves a trace not only in its canonical position but also in an intermediate position through which it transits to satisfy syntactic constraints (Kayne, 1989; see Warren & Gibson, 2005 for psycholinguistic evidence for intermediate traces in comprehension). Critically, that intermediate position intervenes between the subject and the verb in the hierarchical structure, and is therefore expected to be visible to the Controller selection process.

In order to test more directly the hypothesis that the memory retrieval process underlying Controller selection operates on hierarchical structure, Matt Wagers and I designed a study that tested the prediction that attractors situated in syntactic positions triggering more attraction are easier to retrieve from memory (Franck & Wagers, 2015). To do this, we combined a grammaticality judgment experiment with a probe recognition experiment using the response-signal speed-accuracy trade-off procedure (SAT). We used a grammaticality judgment task because our previous research indicates that this task replicates the syntactic modulation of attraction found in sentence production (Franck et al., 2015). I will assume here that both tasks tap into the process of Controller selection. We used the SAT paradigm because it investigates the fine-grained time-course of processing and enables separate measures of retrieval speed and retrieval accuracy (e.g., McElree & Dosher, 1989). In such a probe recognition experiment, participants are trained to respond to a signal presented at varying time points after the onset of the recognition probe (spanning the full time course of retrieval between about 100 ms to 3000 ms), indicating whether the probe was in the list. Distribution of accuracy as a function of retrieval time typically shows an initial phase of chance level performance (participants did not have enough time to select the correct answer), followed by a phase of increasing accuracy, followed by an asymptotic period. The asymptote provides a measure of the overall probability of retrieval (accuracy), which is a joint function of the overall quality of the memory representation and cues at retrieval. Retrieval speed is measured by the intercept of the function, indicating when information first becomes available, and the rate of rise, indicating the rate at which accuracy grows from chance to asymptote. These two parameters provide key indicators of the dynamics of retrieval, independently of the quality of memory representations. In contrast to previous SAT studies of sentence processing that all employ a sentence-acceptability judgment task (e.g., McElree et al., 2003; Van Dyke & McElree, 2006), we used such a probe recognition task in order to get a direct measure of retrieval parameters. Participants first read the sentence presented word by word in a Rapid Serial Visual Presentation manner. At the end of the sentence, they were asked to judge whether a probe word was in the sentence or not. Probe words were subjects, attractors, and words that were not in the sentence. SAT parameters were then linked to the attraction rates obtained in the grammaticality judgment experiment on the same items, in order to examine the alignment of the two measures.

We also incorporated another novelty into the design: the materials involved French Jabberwocky in which nouns were replaced by pseudo-nouns but
grammatical morphemes and verbs were preserved. This made it so that participants had no difficulty judging the grammaticality of agreement, while semantic similarity influences were out of the way. Two types of structures were manipulated, both involving two attractors. The first structure involved object relatives with complex objects similar to those tested in Franck et al. (2015). These structures contain an attractor c-commanding the verb (dafran in Which dafrans of the brapou do you say that the bostron defends?) and one preceding it (dafran in The brapou of which dafrans do you say that the bostron defends?). The second structure involved two subject modifiers like those tested in Franck et al. (2002). Hierarchical height was manipulated by contrasting a modifier situated higher (dafrans in The bostron of the dafrans of the drapou sleeps) and one situated lower (dafrans in The bostron of the drapou of the dafrans sleeps). Here, both attractors precede the verb.

Results from the grammaticality judgment experiment showed that Jabberwocky elicits attraction. Importantly, the c-command versus precedence contrast in complex objects found in natural language was replicated: more attraction was found with the c-commanding object head than with the object modifier intervening by precedence. We found as much attraction from the low as from the high attractor in double subject modifiers. This might be due to the fact that Jabberwocky promotes a more strictly syntactic computation of agreement, in which all that counts is the distinction between c-command and precedence, while finer distinctions among precedence relations have no role to play in the syntax (its effect may be in the semantics). But the more important finding is that results from the probe recognition experiment showed a close alignment with grammaticality judgments. First, subjects are more accessible (higher asymptote) and retrieved faster (faster dynamics) than the two attractors. Interestingly, the higher accessibility of subjects is found independently of their linear position: it is found equally in the complex object condition, where the subject is linearly just before the probe word, and in the subject modifier condition where the subject is situated linearly far from the probe word. This contrasts with list memorization where accessibility is a function of distance. The finding that subjects are retrieved faster than attractors also contrasts with list memorization where all units (apart from the most recent one) are retrieved at the same speed (McElree, 2006). This suggests that subjects are maintained in the focus of attention, even when separated from their verb by PP modifiers (in line with Wagers & McElree, in press), capturing the fact that in most of the cases, they are correctly retrieved as the controller for agreement computation. Second, the accessibility of the two attractors aligns with their potential to trigger attraction: the c-commanding attractor in complex objects was significantly more accessible (higher asymptote) than the preceding one whereas no difference was found between the two attractors of the subject modifier structure, in line with the finding that a c-commanding element triggers more attraction but PP modifiers trigger similar attraction in the grammaticality judgment task.

These novel results suggest that the memory retrieval processes underlying sentence processing are constrained by the grammar: subjects are especially prominent in memory, followed (by far) by elements c-commanding the verb, and then by those situated in a position of precedence. They bring direct support
to the hypothesis that attraction is a function of the attractor’s visibility to the memory retrieval process, and that this higher visibility itself depends on the attractor’s syntactic position in the hierarchically structured sentence, in keeping with the critical distinction between c-command and precedence. In sum, I tentatively suggest that syntactic theory describes the strength of memory representations.

Summary and future challenges

Capitalizing on the initial proposition by Badecker and Kuminiak (2007) that attraction results from the incorrect selection of the controller, I have proposed a functional model of agreement that involves two components, Feature selection and Controller selection, the latter being considered as the locus of attraction proper. I developed a typology of attraction effects that splits these effects into two classes according to their functional locus in the model: those that arise during Feature selection, and those that arise during Controller selection. The subclasses of the typology are structured according to the nature of the factors that compose them: semantic, morphophonological, and morphological factors that influence the stability of the feature, and notional factors that influence the construal of the subject phrase, all affecting Feature selection; semantic similarity, case marking similarity and syntactic similarity between the head and the attractor affecting Controller selection. This way of cutting the pie is radically different from that proposed in M&M where effects are split according to the nature of the variables that underlie them: semantic variables affect the first stage of Marking, whereas syntactic and morphological variables affect the second stage of Morphing. In the current proposal, different types of semantic factors and different types of morphological factors are argued to influence both processes, although in different ways.

The framework I have proposed here opens two important avenues for future research. The first avenue concerns the fine characterization of syntactic representations as mental objects. The psycholinguistics of sentence production has managed to set aside this core question for 40 years of research, contributing to a drifting-apart between psycholinguistics and linguistics ever since the disillusionment following the failure of the derivational theory of complexity (Fodor, Bever & Garrett, 1972). As clearly expressed in a recent state of the art review of the sentence production literature (Bock & Ferreira, 2014), psycholinguistic accounts of language are set apart from linguistic approaches in that they are concerned with “the situatedness of sentence production in the circumstances of communication. Speakers have to do a whole lot more than create grammatically acceptable sentences. They have to create acceptable sentences that make sense. This means that they have to convey particular notions to particular people in particular circumstances in a particular language” (p. 42). And indeed, the question of the form of syntactic representations is absent from most sentence production studies. Even the broad research program on syntactic priming showing that speakers tend to
re-use a particular syntactic structure (see Chapters 6, 7, 14) remains fundamentally agnostic about the shape of these structures. This line of work is restricted to a few syntactic structures and rests on a superficial description of their properties. That description is sufficient when the prime and target have identical structures, since syntactic identity can easily be assessed without deeper analysis. Nevertheless, a few studies have pushed the question further in exploring the possibility that priming generalizes to other structures (Bock & Loebell, 1990; Bock, 1989; Griffin & Weinstein-Tull, 2003). This approach seems extremely promising in that it reveals an even more abstract level of representation where syntactic similarity, the underlying basis for generalization, has to be characterized. Drawing the map of similarities and differences between structures cannot bypass a fine analysis of their properties.

The second avenue is the study of the relation between sentence production processes and memory mechanisms, which has seldom been raised (in contrast to research in sentence comprehension). In the model of agreement proposed, both Feature selection and Controller selection are memory retrieval processes. I argued that Feature selection, which amounts to retrieving nominal features from the long-term memory store of function words, operates under various types of constraints, in line with the broad literature showing interactivity in single word production. I suggested that the process of Controller selection, responsible for retrieving the controller from the memory representation of the currently built sentence, shows the hallmark of cue-based retrieval: similarity-based interference. Empirical evidence shows that a higher overlap between memorized units endowed with semantic and syntactic controller-specific features creates interference that occasionally manifests in the form of an attraction error. In this view, attraction is the result of the incorrect selection of the controller during agreement computation. I ended with the report of experimental data obtained with a new design allowing us to explore more directly the link between attraction and memory. The data show that syntax-based variations of attraction strength closely align with variations in memory retrieval measures: sentential subjects are more accessible and retrieved faster than attractors, and attractors that generate more attraction are more accessible than those that generate less attraction.

Some major questions remain open. One question is whether an error in Controller selection arises because an erroneous syntactic tree has been built, or whether it is independent of the overall structure building process, as suggested here. It has often been informally observed that the production of an attraction error does not entail that the speaker has reached the wrong interpretation of the sentence in which the attractor is the subject, and experimental evidence seems to support that claim (Lau et al., 2008). Nevertheless, it seems premature to date to firmly conclude in that direction, and more direct empirical tests need to be designed. Another question concerns the precise identification of the aspects of the memory retrieval process that are affected by the syntactic position of the attractor. The results of my work with Matt Wagers suggest that subjects remain in the focus of attention, however, the memory mechanisms underlying attractor’s access are less clear. One possibility is that syntactic position affects the decay rates of the
controller and attractors in the sentence: the c-commanding object may be reactivated when reaching the verb, boosting its level of activation (McElree, 2000). In order for the model to account for the data, decay rates would then need to be a function of syntactic structure (e.g., Lewis & Vasishth, 2005; McElree, 2000). Another possibility is that syntactic position affects retrieval cues: c-commanding positions and more generally high positions in the tree are typical positions of subjects, such that attractors occupying these positions would be more prone to interfere in Controller selection. Translating syntactic position into retrieval cues is challenging given the fundamentally relational nature of positional information, but proposals have been made along those lines (e.g., Kush, 2013; Wagers & McElree, in press). I pinpointed the relevance of further examining the role of similarity among different types of structures in syntactic priming. The same logic could be applied to the study of attraction, in identifying what syntactic properties (beyond morphological similarity) make an attractor controller-like. Such a view links grammatical factors to surface properties that probabilistically correlate with them, opening the possibility that these correlates play a role in driving the production system as well (e.g., Bever & Poeppel, 2010). It also introduces the intriguing possibility that the fine syntactic constructions characterized by linguistic theory are represented as continuous mental objects, organized along a structural proximity metric (Tabor, Cho, & Szkudlarek, 2013). These questions constitute an interesting program of future research.

The finding that attraction patterns may stem from variations in memory retrieval is not the end of our journey. What remains to be understood is why memory is organized the way it is, which means identifying the cognitive constraints that shape natural language grammars. Nevertheless, the line of research sketched here paves the way for a new relationship between syntactic theory and cognitive psychology, and raises the hope that it will (re)open the debate on the possibility that the theory of competence is also a theory of performance.

Acknowledgments

The author is grateful to Whit Tabor for the energizing and thought-provoking discussions on the ideas developed in that paper.

NOTES

1 Some languages like Slovenian show the opposite profile with singular verbs being required with quantified plural heads referring to more than five units (in which case the sentence amounts to saying five people has arrived, Marušič et al., 2011). Phenomena of closest conjunct agreement also give rise to cases where agreement takes place with the linearly closest conjunct, even if this conjunct is not marked (e.g., Marušič et al. 2011; Haskell & MacDonald, 2005). More work is needed to estimate the role of statistical distributions in attraction asymmetries.
Recent evidence from structures without attractors suggests that it is the degree of individuation of the head and not the number of units that it involves that counts (Mirković & MacDonald, 2013). In Serbian, quantified noun phrases take singular verbs. However, more plural verbs are found with a quantifier like several, which is judged as more individuated, than with many. Similar differences are found between agentive and existential verbs. The former are claimed to promote a more individuated interpretation, and indeed more plural verbs erroneously occur with the former.

The retrieval approach proposed in these studies of attraction in sentence comprehension differs from that of Badecker and Kuminiak (2007) and the one advocated here in assuming that retrieval is selectively triggered when an agreement error is encountered. This conclusion was reached on the basis of the finding that a number mismatch between the controller and the head only affects verb processing in sentences containing an agreement error. The proposal here is that both the generative component of the production process and the predictive component of the comprehension process involve retrieving the controller to compute agreement (even though comprehension involves some specificities, see Franck, Colonna & Rizzi, 2015 for a discussion).

In two relevant studies involving on-line response time measures, Staub (2009, 2010) argued in favor of two distinct causes underlying attraction in object relatives and in prepositional phrase modifiers (in line with Bock & Miller, 1991). Evidence comes from the distribution of response times in the production of the verb in these two structures: whereas prepositional modifier attraction shows a small but systematic effect across trials, object attraction shows an irregular, strong effect on only a subset of trials. This finding remains unexplained in the current framework, which suggests that a single mechanism underlies both types of attraction.

REFERENCES


Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and
grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, 114(2), 151–164. doi: 10.1016/j.cognition.2009.08.017


Introduction

Human languages provide speakers with a remarkable degree of flexibility in how to linguistically encode near-meaning equivalent messages. This chapter focuses on what is arguably the most pervasive type of flexibility: flexibility in the amount or quality of the signal that encodes the speaker’s message. Figure 3.1 illustrates this for English (inspired by Friedman, 2013). For example, an element may be mentioned or omitted (e.g., the optional complementizer that, or the argument the World Cup), or the articulatory realization of an element may be more or less detailed (e.g., producing a more centralized vowel or shortening the duration of a word). Such flexibility has been of central interest in psycholinguistic research: speakers’ preferences to encode a message with a more or less reduced signal serve as a window into the architecture underlying the language production system.

Although speakers typically do not become aware of this flexibility while talking, the choice between more or less reduced linguistic forms or signals is ubiquitous within and across languages. Alternations like those illustrated in Figure 3.1 exist across many, if not all, languages. Languages differ, however, in the specific alternations that they afford. For example, many languages allow omission of grammatical subjects in certain contexts (e.g., Italian, Japanese, Russian, and Yucatec Maya), whereas this omission is considered ungrammatical—or restricted to colloquial registers—in other languages (e.g., English). Other examples of reduction include optional mention of case-marking (e.g., in Japanese, Korean, and Turkish) or optional head-marking morphology (e.g., in many languages of the Balkan sprachbund), neither of which are available in English.
Reduction constitutes the empirical focus of this chapter. Specifically, we focus on probabilistic reduction: a large body of work has found that speakers tend to produce shorter linguistic forms and more reduced signals for contextually predictable parts of the message. To the best of our knowledge, a systematic review of research on reduction across different levels of linguistic representations has so far been lacking. We thus begin with a summary of this literature and the questions it raises for future research.

The second part of this chapter reviews competing theories and accounts of the empirical findings discussed in the first part. Although we focus on probabilistic reduction, the discussion bears on more general architectural questions. In particular, we discuss competing views of the link between production and comprehension, as well as the link between online processing and biases implicitly encoded in linguistic representations. We distinguish between three broad classes of accounts. One hypothesis holds that flexibility in encoding a message allows speakers to navigate the attentional and memory demands of language production. This type of explanation is sometimes referred to as “production-internal” (Arnold, 2008), “production-based” (Gahl, Yao, & Johnson, 2012), “production-oriented” (Lindblom, 1990a), or “production-centered” (Watson, Arnold, & Tanenhaus, 2010).

![Figure 3.1](image)

**Figure 3.1** Illustration of a few types of implicit decisions speakers make during linguistic encoding that affect the degree of signal reduction. Reduction can be caused by decisions at multiple levels of linguistic encoding, including sentence, lexical, and phonological planning, as well as articulation. In the appropriate context, the upper and lower utterances encode the same message; yet the lower utterance contains shorter linguistic forms and is realized with a much reduced speech signal, compared to the upper utterance. Inspired by Friedman (2013).
This has been contrasted with the idea that production is affected by communicative considerations. According to the latter view, the mechanisms underlying linguistic encoding are—directly or indirectly—affected by comprehension (e.g., Brennan & Clark, 1996, Clark & Fox Tree, 2002; Jaeger, 2006; Lindblom, 1990a). This alternative idea is variously referred to as, for example, “listener-oriented” (Arnold, 2008), “comprehension-facilitation” (Arnold, Kahn, & Pancani, 2012), “intelligibility-based” (Gahl et al., 2012), or “audience design” (Clark & Murphy, 1982; Galati & Brennan, 2010). Here we use the labels production ease and communicative accounts to refer to these two views.

Independent of what (mixture of) pressures ultimately drive speakers’ preferences, there are questions about whether these pressures operate on-line, directly affecting speakers’ preferences during incremental linguistic encoding, or off-line, changing linguistic representations and thus only indirectly affecting incremental encoding. Explanations that focus on the latter possibility constitute a third type of account, which we will refer to as representational accounts (e.g., Pierrehumbert, 2001; Wedel, 2006). For each of these accounts, we review specific proposals and isolate some challenges we consider particularly pressing for future research. The picture that emerges from this discussion is one in which probabilistic reduction is not driven by any single factor, but rather the result of multiple mechanisms.

A few terminological clarifications Throughout this chapter, we refer to such differences as reduction (and to reduced variants), without meaning to imply a directionality of this process: for many phenomena we discuss, it is an open question whether they are better understood as reduction or enhancement. For example, although it might seem more intuitive to think of the complementizer that in Figure 3.1 as being optionally omitted, there are also arguments as to why it is better thought of as being optionally mentioned. In conversational American English, for example, the complementizer that is absent in about 83% of all complement clauses (Jaeger, 2010, p. 29). Even when the most frequent complement clause embedding verbs are excluded, omission is more frequent (53%) than mention of complementizer that (Jaeger, 2010, Table 1). This makes it difficult to determine whether this alternation is better understood as optional mention or optional omission. Similarly, word durations may undergo reduction (i.e., shortening) or enhancement (i.e., lengthening).

We also distinguish between message components, linguistic forms, and their realization in the linguistic signal. Message components are parts of the message speakers wish to convey (e.g., a specific lexical meaning). Linguistic forms are instances of linguistic categories, such as phonological segments, words, and syntactic structures. These forms are not directly observable. Rather, they underlie the observable linguistic signal. The linguistic signal can be acoustic (in the case of speech) or visual (in the case of gestures, sign language, or writing). We sometimes refer to more or less reduced forms to highlight that reduction goes beyond gradient manipulation of the signal and includes cases where language provides speakers with several more or less reduced linguistic forms (e.g., mentioning or omitting the world cup in Figure 3.1).
Probabilistic reduction: Contextual predictability and signal reduction

As shown in Figure 3.1, reduction can take place at different levels of linguistic encoding. Reduction at many of these levels has been found to be correlated with contextual predictability, so that more probable (and less informative) message components tend to be realized with reduced signal. We begin with a summary of work on phonetic and phonological reduction. Then we summarize work at successively higher levels of linguistic encoding, including morphological contraction, the omission of optional function words, and the realization of referring expressions. We close this section with an overview of open empirical questions.

Phonetic and phonological reduction and omission

A large number of studies have investigated the articulatory or acoustic reduction of phonemes, syllables, and words. This research has found that contextually predictable instances of words tend to be produced with shorter duration (e.g., Aylett & Turk, 2004; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Tily et al., 2009) and less articulatory detail (e.g., Aylett & Turk, 2006; Gahl et al., 2012; Son & Santen, 2005). Such probabilistic phonetic reduction has been observed in conversational speech corpora (e.g., Arnon & Cohen Priva, 2014; Aylett & Turk, 2004; Bell et al., 2003; Gahl et al., 2012; Pluymaekers, Ernestus, & Baayen, 2005a) and in the lab, including read speech (e.g., Arnon & Cohen Priva, 2013; Gahl & Garnsey, 2004; Kurumada, 2011) and unscripted speech (e.g., Watson et al., 2010).

For example, contextually predictable instances of words tend to have more reduced vowels (e.g., Aylett & Turk, 2006; but see null results in Bürki, Ernestus, Gendrot, Fougeron, & Frauenfelder, 2011; Clopper & Pierrehumbert, 2008; Gahl et al., 2012; Scarborough, 2010) and consonants (Rose, 2017: Ch. 3; Torreira & Ernestus, 2009, 2012). Aylett and Turk investigated predictability based reduction in a corpus of citation speech. They measured syllable durations and first and second formant values of vowels within those syllables and binned syllables into high and low predictability based on unigram, bigram, and trigram probabilities. They found that syllables with high predictability were shorter in duration and vowels within those syllables showed more centralization. Contextually predictable words are also more likely to undergo phonological weakening or deletion (Bell et al., 2009, 2003). As an example, many varieties of English favor the reduction of complex codas in some phonological environments. A specific case of this is t/d-deletion, where a t or d that is present in citation form is not produced. Such t/d-deletion is more common in predictable words (Gahl, Jurafsky, & Roland, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001; see also Bybee & Hopper, 2001). Other research has further found that a segment’s informativity about the word affects the segment’s realization even after the word’s predictability is taken into account (e.g., van Son & Pols, 2002; van Son & van Santen, 2005).
Similar reduction effects have also been observed as a function of previous mention of a word (Bard et al., 2000; Bell et al., 2009; Pluymaekers, Ernestus, & Baayen, 2005b; Watson et al., 2010). Since the statistics of human language are such that previous mention generally increases the probability of being mentioned again (e.g., Rosenfeld, 1996, Section 2.3), these effects could at least in part be mediated through effects of previous mention on a word’s contextual predictability (for evidence, see J. R. Heller & Pierrehumbert, 2011; for discussion, see Kahn & Arnold, 2012).

While phonological weakening or deletion has been studied extensively, less is known about phonological insertion. One example comes from optional epenthesis. In epenthesis, speakers insert a reduced vowel into a consonant cluster (e.g., filém). Epenthesis enhances the signal and reduces syllable complexity compared to what would be expected under a faithful realization of the citation form. In a corpus study on conversational Dutch, Tily and Kuperman (2012) found that speakers were less likely to insert the schwa into words that were contextually predictable.

How big are the effects of contextual predictability? Bell et al. (2003) find that the top and bottom 5% most predictable instances of English function words (such as the, I, etc.) differ in their duration by about 20-30 ms, out of a mean duration of about 100 ms. For content words, the most predictable instances of words are sometimes more than 100 ms shorter than their least predictable instances (Demberg, Sayeed, Gorinski, & Engonopoulos, 2012, p. 364). These effect sizes mean that predictability effects tend to be somewhat smaller than, though sometimes comparable to, durational lengthening associated with differences in linguistic structure or meaning (such as contrastive prosodic accents, Berkovits, 1994; phrase final lengthening, Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991). At the same time, these effect sizes imply that at least some probabilistic reduction is clearly perceptible (cf. Beaver, Clark, Flemming, Jaeger, & Wolters, 2007, who report detection of 6 ms durational differences). Indeed, although this is an area that has received surprisingly little attention, there is evidence that the phonetic reduction associated with contextual predictability does affect intelligibility (Bard & Anderson, 1983, 1994; see also Buz, 2016, Ch. 4 for related discussion).

In summary, there is ample evidence that a word’s contextual predictability tends to be correlated with its reduction (for further references, see Ernestus & Warner, 2011; Ernestus, 2014). Here we have focused on evidence from English. This reflects the status of the field, with the majority of existing research on phonetic reduction coming from English and typologically related languages (e.g., Dutch: Kuperman, Pluymaekers, Ernestus, & Baayen, 2007; Pluymaekers et al., 2005a; van Son & van Santen, 2005; French: Bürki et al., 2011; Pellegrino, Coupé, & Marsico, 2011; Torreira & Ernestus, 2009; Italian: Pellegrino et al., 2011; Spanish: Torreira & Ernestus, 2012), with only a handful of comparable studies on other languages (e.g., Cantonese: Zhao & Jurafsky, 2009; Japanese: Kurumada, 2011; Vietnamese, among others: Pellegrino et al., 2011). By taking advantage of language-specific properties, future studies on languages other than English hold
great promise for the study of probabilistic reduction (for the critical importance of cross-linguistic evaluations of psycholinguistic theory, see also Jaeger & Norcliffe, 2009; Norcliffe, Harris, & Jaeger, 2015). For example, Zhao and Jurafsky (2009) investigated the effects of frequency on the realization of lexical tone in Cantonese. They found that frequency influences pitch production: low frequency words are produced with tone contours that are more distinct from each other. Paralleling phonetic and phonological reduction in English, Cantonese speakers thus tend to produce more reduced—or less distinguishable—signals for contextually more expected—and thus less informative—message components.

**Morphological contraction and omission**

Effects resembling probabilistic phonetic reduction have been observed in speakers’ preferences between near-meaning equivalent morphological forms. For example, Frank and Jaeger (2008) investigate morphological contraction in American English conversational speech. Specifically, they focus on not (e.g., *isn’t* versus *is not*), auxiliary be (e.g., *he’s* versus *he is*), and auxiliary have (e.g., *I’ve done that* versus *I have done that*). They find that the rate of morphological contraction increases with the predictability of the meaning of the contractible element (e.g., negation in *isn’t* versus *is not*). This effect holds while controlling for potentially confounding factors such as speech rate, the type of host word preceding the contractible element, and the complexity of the material following the contractible element (see also Frank & Jaeger, 2008). These effects are confirmed by other studies on morphological contraction in conversational English (Bresnan & Spencer, 2016; Bybee & Scheibman, 1999).

More recent research has investigated alternations in which a bound morpheme can be either mentioned or omitted under near-meaning equivalence (Kurumada & Jaeger, 2015; Norcliffe & Jaeger, 2014). For example, Kurumada and Jaeger (2015) investigate optional case-marking in Japanese. Like in other case-marking languages, Japanese has case-marking morphology on the arguments of the verb that encode the grammatical function assignment. For example, the direct object of a transitive verb is marked with the suffix -o. Case-marking is important in understanding Japanese sentences, since Japanese has flexible word order, allowing both subject-before-object and object-before-subject ordering in transitive sentences. Unlike languages in which case-marking is obligatory (e.g., German), informal spoken Japanese allows speakers to omit the case marker without loss of near-meaning equivalence (see also Fry, 2001). In fact, case omission is frequent in informal Japanese (e.g., up to 51% of object case markers are omitted, Fry, 2001). In a spoken recall study, Kurumada and Jaeger find that speakers are more likely to omit the direct object case marker -o when the sentence makes the intended grammatical function assignment contextually predictable (e.g., *grandma* is more likely to be case marked in *The doctor sued the grandma* than in *The doctor treated the grandma*). Related corpus-based research has found that the rate of case-marking depends on how typical an argument is for the grammatical function it carries in the sentence (e.g., for Japanese: Fry, 2001; Korean: H. Lee, 2006; for further
evidence from artificial miniature language learning, see Fedzechkina, Jaeger, & Newport, 2012, Fedzechkina, Newport, & Jaeger, 2016). For example, in conversational Korean, which also has optional case-marking, definite subjects are less likely to be case marked than indefinite subjects, whereas definite objects are more likely to be case marked than indefinite objects (H. Lee, 2006, Table 4). Since subjects are more likely to be definite than objects are, these findings suggest that case is more likely to be omitted when the meaning it encodes is predictable from context (for discussion, see Kurumada & Jaeger, 2015).

Another example comes from a recent study on optional head-marking in Yucatec Maya (Norcliffe, 2009; Norcliffe & Jaeger, 2014). In head-marking languages, grammatical function assignment and other information is encoded through bound morphology or clitics attached to the verb (rather than the arguments, as in the case of case-marking). In Yucatec some of this morphology is optional in certain environments. Norcliffe and Jaeger (2014) provide evidence that this optional morphology follows similar patterns as described for case-marking in Japanese above.

In sum, existing cross-linguistic evidence suggests that speakers' preferences in morphological reduction environments (i.e., contraction and omission) are affected by contextual predictability in ways that are at least qualitatively similar to phonetic reduction. However, compared to phonetic and phonological reduction, relatively little is known about the pressures driving optional morphological contraction and omission. Research on morphological production in morphologically rich languages seems a particularly promising venue for future work.

**Omission of optional function words**

Probabilistic reduction has also been documented for morphologically free function words. For example, English allows the omission of complementizer that, as in sentences like *She certainly knew (that) this was a required test* (Elsness, 1984; Huddleston & Pullum, 2002). This phenomenon is sometimes referred to as optional complementizer *that*-mention or *that*-omission. Speakers are more likely to produce the optional complementizer *that*, when the complement clause is less predictable given the matrix verb (e.g., *knew* in the example above). This effect has been observed in conversational speech (Jaeger, 2010) as well as production experiments (e.g., in written sentence completion, Garnsey, Pearlmuter, Myers, & Lotocky, 1997, Table 5; spoken or written recall, Ferreira, 2008; Jaeger & Grimshaw, 2013).

Optional function word omission is also observed in certain types of relative clauses. For example, in Standard American English, both finite non-subject-extracted non-pied-piped relative clauses (e.g., *That's the way (that) it is done*) and passive subject-extracted relative clauses (e.g., *These are the type of people (who are) not taken seriously*) allow similar omissions. For these environments, too, speakers have been found to be more likely to omit the optional function words the more predictable the constituent they introduce is in context (Jaeger, 2010, 2011; Wasow, Jaeger, & Orr, 2011; see also Melnick, 2011; Wiechmann, 2015).

In sum, speakers' preference to mention or omit optional function words seems to exhibit sensitivity to contextual predictability in ways that resemble phonetic
reduction. However, beyond *that*-omission, the sensitivity of optional function word omission to contextual predictability has remained under-explored. Alternations similar to optional complementizer *that* exist in other languages (e.g., in Danish), though omission is sometimes accompanied by constituent order changes (e.g., in German). English, too, contains a number of additional environments that support optional omission of function words, such as the omission of *to* after verbs like *help* (Rohdenburg, 2004) or in the do–be construction (e.g., *all I want to do is (to) go to work*, Flickinger & Wasow, 2013). Additional examples are observed in non-standard varieties of American English, such as optional copula omission in African American Vernacular English (e.g., *You done yet?*; Bender, 2000, p. 85) or relativizer omission in subject-extracted relative clauses in, for example, the English of the British Isles (e.g., *And there were a wee alarm clock sat on the window*; Tagliamonte & Smith, 2005, p. 87).

It is thus an open question whether the effects of contextual predictability observed in research on *that*-omission in Standard American English will generalize to these similar phenomena and across languages. Preliminary evidence comes from ongoing research on the do–be construction (Wasow, Levy, Melnick, Juzek, & Zhu, 2015). Wasow and colleagues find that speakers are more likely to omit *to* in the do–be construction in lexical contexts that frequently co-occur with the do–be construction.

**Reduction and omission of referring expressions**

Another domain in which languages typically provide multiple near meaning-equivalent forms with more or less reduced signals is referring expressions. For example, in many contexts speakers can choose between a pronoun (e.g., *he*), name (e.g., *John*), or a full lexical noun phrase (e.g., *a colleague of mine*) to refer to the same referent.5

It has long been hypothesized that the choice between these different ways of encoding a reference depends on the referents “accessibility” in context (e.g., Ariel, 1999; Givón, 1983). This includes several factors that make referents more predictable (Arnold, 1998, 2010). For example, previous mention of a referent makes it more likely that it will be referred to in subsequent utterances. Previous mention also makes it more likely that a more reduced form is chosen (Bard et al., 2000 and references therein). Moreover, the probability that a previously mentioned referent is referred to again decreases with increasing distance from its last mention. Similarly, the preference for a pronoun over a longer referring expression decreases with increasing distance from the last mention of a referent (Arnold, 1998; Arnold, Bennetto, & Diehl, 2009; as summarized in Arnold, 2010, p. 190).

Recent work has more directly assessed the effect of contextual predictability on the realization of referring expressions, paralleling research on probabilistic phonetic reduction. Tily and Piantadosi (2009) employed a type of Shannon guessing game (Shannon, 1951) to obtain estimates of the contextual predictability of over 2,000 references in a newspaper corpus. In their version of the Shannon guessing game, raters saw story fragments up to the next referring expression. Their task was to guess which of the previously introduced referents (or possibly a new referent) the
next expression would refer to. Almost 500 raters provided a total of over 70,000 guesses. This made it possible to calculate estimates of the contextual predictability of the actual references made in the corpus. Tily and Piantadosi found that writers had indeed been more likely to use longer linguistic forms (e.g., names rather than pronouns) when the intended reference was less expected given the preceding context. This effect held beyond the effects of previous mention and other previously documented effects (for related results, see also Rohde & Kehler, 2014).

Mahowald, Fedorenko, Piantadosi, and Gibson (2013) investigated speakers’ preference between full and reduced lexical forms with the same meaning, such as *mathematics* and *math*. Mahowald and colleagues found that speakers’ preference for the shorter form increases with the contextual predictability of the concept encoded by either form. In a corpus study, the average informativity (measured as Shannon information) of long forms was significantly higher than for short forms suggesting that short forms tend to be used in contexts where they conveyed less information. In a sentence completion study, Mahowald and colleagues further found that participants chose the short form for sentences with supportive contexts (e.g., *Susan loves the apes at the zoo, and she even has a favorite …*) as compared to non-supportive contexts (e.g., *During a game of charades, Susan was too embarrassed to act like a …*). This preference closely mirrors the preference observed for contractible auxiliaries and negation (Bresnan & Spencer, 2016; Bybee & Scheibman, 1999; Frank & Jaeger, 2008).

A similar preference to produce reduced linguistic signals for contextually more predictable referents is also observed for optional argument omission (Kravtchenko, 2014; Resnik, 1996). In certain lexical environments, speakers of English can decide to omit an entire argument (e.g., *the semi-finals in Germany lost (the semi-finals)*), while maintaining near meaning-equivalence. In his seminal corpus study, Resnik found that verbs that contained more information about the types of arguments they take, thereby making the arguments following them (on average) more predictable, also are associated with a higher rate of argument omission (Experiment 4, Resnik, 1996).

Recent work on optional subject omission in Russian builds on these results (Kravtchenko, 2014). While considered non-standard or ungrammatical in English, many languages allow omission of contextually inferable subjects, sometimes referred to as pro-drop (Dryer, 2013). Using the version of the Shannon guessing game developed by Tily and Piantadosi (2009), Kravtchenko (2014) obtained estimates of the contextual predictability of over 700 subject noun phrases from a Russian corpus. Paralleling the results for the realization of referential expressions in English, Kravtchenko found that Russian subjects are more likely to be omitted when they are contextually predictable.

**Reduction beyond the level of the clause**

The majority of psycholinguistic research has focused on linguistic encoding at the level of the clause or below. A few more recent studies have begun to investigate reduction beyond the clause. For example, Asr and Demberg (2015) investigated
the realization of coherence relations in English (see also Asr & Demberg, 2012). Simplifying somewhat, coherence relations are discourse relations between propositions. Asr and Demberg (2015) focused on the so-called Chosen Alternative relation and the coherence marker instead in environments in which it is optional (e.g., They didn’t panic during the first round of selling. (Instead,) they sold into the strength, which kept the market orderly). Asr and Demberg found that instead was more likely to be omitted in the presence of a contextual cue to the Chosen Alternative relation (but see Anibel, 2010, for a failure to find such effects for other types of coherence relations).

Another environment in which speakers have the choice between providing more or less linguistic material to encode a near meaning-equivalent message was investigated by Gallo and colleagues (Gallo, 2011; Gallo, Jaeger, & Furth, 2010; Gallo, Jaeger, & Smyth, 2008). For example, Gallo et al. (2008) had speakers participate in a version of the Map Task (A. H. Anderson et al., 1991; see Pardo, this volume, for a description). Speakers instructed another (confederate) participant to replicate on their screen a specific arrangements of objects seen only by the speaker. Gallo and colleagues coded whether speakers used one or two sentences to convey the same message. For example, participants could say Move the triangle to Central Park or use a more verbose message like Take the triangle. Now move it to Central Park. Gallo and colleagues found that speakers were more likely to split the message across two clauses when the object (e.g., the triangle) consisted of less predictable words (for similar evidence from Spanish, see Gallo et al., 2010). These effects held beyond effects of previous mention, which is known to be correlated with the choice between pronoun versus lexical NPs (cf. Tily & Piantadosi, 2009).

Of the areas summarized here, production planning (including preferences regarding reduction) beyond the clause-level is probably the least understood. Further work is required to see whether the tentative evidence summarized here will confirm that principles similar to those observed in phonological, lexical, and syntactic reduction also operate during planning of larger linguistic chunks.

**Summary and open questions**

Language provides speakers with an astonishing degree of flexibility in the linguistic encoding of messages. Many of the options available to speakers differ in the amount of signal produced by the speaker. Across all stages of production summarized here, speakers’ preferences between different ways of realizing the same message seem to be affected by a similar bias, reflected in a correlation between contextual predictability and reduction. More specifically, it seems that it is the predictability of a linguistic form or message component (roughly, part of the meaning a speaker wishes to convey) that correlates with a preference for shorter linguistic forms at the next lower level and more reduced linguistic signals. For example, the predictability of negation following a lexical context (e.g., President Clinton did …) correlates with an increased preference for morphological contraction (i.e., saying President Clinton didn’t … rather than President Clinton did not …, Frank & Jaeger, 2008). Similarly, it seems to be the predictability of a complement
clause that correlates with an increased preference to omit the relativizer that
(Jaeger, 2010) and the predictability of a lemma that correlates with the reduction
of its word form (Aylett & Turk, 2004; Jurafsky et al., 2001). In this context, a partic-
ularly intriguing piece of evidence comes from research on homophones, such as
time and thyme. While time and thyme have the same phonological citation form,
the actual realization of the two words tends to differ subtly (Gahl, 2008). Speakers
tend to produce the more frequent lemma (time) with a more reduced speech
signal, compared to the less frequent lemma (thyme). To the best of our knowledge,
comparable work on the effects of contextual predictability on homophone
pronunciation has yet to be conducted. Still, this type of effect suggests that it is at
least partly the predictability of a message component (in this case the lemma or
its meaning) that drives the extent to which its realization in the linguistic signal is
reduced (see also Jaeger, 2006, Study 6).

While the inverse correlation between predictability and linguistic signal is
now firmly established, many questions remain about the nature of this relation.
The perhaps most pressing questions regard the processes underlying probabi-
listic reduction and, in particular, the relation between production planning and
the realization of the linguistic signal. Before we address these questions in the
second part of this chapter we briefly summarize outstanding empirical ques-
tions about probabilistic reduction. One question that deserves further attention
is the relation between reduction at different levels of linguistic encoding (e.g.,
phonetic vs. syntactic reduction). Simply put, what determines the level of
linguistic encoding at which speakers reduce or enhance the signal? This question
has received some attention in research on phonetic reduction and phonological
deletion (e.g., Bürki, Ernestus, & Frauenfelder, 2010; Bürki et al., 2011; Hanique,
Ernestus, & Schuppler, 2013; Torreira & Ernestus, 2011). For example, some cases
of omission might be better understood as extreme cases of gradient phonetic
reduction, while others are better understood as originating in categorical phono-
logical representations.

Another open question is what types of cues affect probabilistic reduction. The
majority of previous research on probabilistic reduction has focused on the imme-
diately surrounding lexical context. For example, for phonetic reduction most
research has estimated the word’s predictability based on its surrounding trigram
context (e.g., Aylett & Turk, 2004; Bell et al., 2009, 2003; Gahl et al., 2012; van Son &
the predictability of the final word in a 4gram (e.g., tea in a cup of tea) is correlated
with phonetic reduction, even after bi-, tri-, and unigram frequencies are
accounted for (see also Arnon & Cohen Priva, 2014; Demberg et al., 2012). Similarly,
most research on reduction at higher levels of linguistics encoding has employed
local lexical cues (e.g., Frank & Jaeger, 2008; Jaeger, 2010; Mahowald et al., 2013;
Resnik, 1996).

There are, however, also some studies that have found less local or more abstract
cues to affect reduction. For example, phonetic reduction has been found to be cor-
related to the word’s predictability given its semantic (Sayeed, Fischer, & Demberg,
Less local cues have also been found to affect the omission of optional function words (syntactic context, Jaeger, 2006, Study 5; Levy & Jaeger, 2007; Wasow et al., 2011) as well as the reduction of referring expressions (cloze completions, Kravtchenko, 2014; Tily & Piantadosi, 2009), although some of these studies have not tested whether the same effects could be attributed to more local cues.

A closely related question is whether different types of cues are weighted differently depending on the level of linguistic encoding (e.g., phonological versus morphological contraction). This would arguably be expected under most accounts discussed below. Even accounts of linguistic encoding that assume that information from lower levels can affect earlier stages of production generally assume that these influences are weaker than influences from the current or earlier stages of production (e.g., Dell, 1986; Dell, Chang, & Griffin, 1999; Janssen & Caramazza, 2009). For example, segmental phonological properties generally only weakly affect syntactic preferences (Jaeger, Furth, & Hilliard, 2012a; McDonald, Bock, & Kelly, 1993). Suprasegmental phonological preferences, on the other hand, have been found to affect syntactic production. For example, speakers prefer to insert optional function words or reorder constituents so as to avoid adjacent stressed syllables (Anttila, Adams, & Speriosu, 2010; Jaeger et al., 2012a; M.-W. Lee & Gibbons, 2007). Similar asymmetries in the factors that drive variation have been observed between the phonetic reduction of segments and their omission (for results and discussion, see Bürki et al., 2011; Hanique et al., 2013). Whether similar asymmetries are reflected in what cues affect probabilistic reduction is a question for future research (for preliminary results, see an unpublished study by Jaeger, Snider, Staum, & Jurafsky, 2006, who compared the phonetic reduction and optional omission of complementizer and relativizer that).

Another question is whether and how speakers integrate multiple cues to the same target (e.g., the same word). For example, does such integration follow similar principles that have been observed in comprehension, where comprehenders seem to be able to integrate multiple sources of information (e.g., Hare, McRae, & Elman, 2004; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995)? To the best of our knowledge, there is so far no published work that addresses this question. A few studies have compared the effect of predictability (or surprisal) estimates based on different types of cues (e.g., Demberg et al., 2012, p. 364; Sayeed et al., 2015). But these studies have not directly compared the objective information contained in these cues to their relative importance in the subjective language models that speakers implicitly draw on during linguistic encoding. Preliminary evidence comes from an unpublished study on phonetic reduction in speech (Post & Jaeger, 2010). Post and Jaeger integrated multiple lexical and syntactic cues into a single estimate of a word’s predictability. They found that both types of cues contributed to a word’s phonetic reduction and that they did so proportionally to their contribution to the word’s predictability. If confirmed by future work, results like these would suggest that probabilistic reduction draws on
multiple contextually available cues, weighted by their relative informativity (see also Jaeger, 2006, Studies 3 and 4, for related evidence for optional complementizer and relativizer *that*).

### Theoretical positions

Psycholinguistic accounts of probabilistic reduction tend to come in three broad flavors: production ease, communicative, and representational accounts. Production ease accounts attribute variation in speakers’ preferences to the demands of incremental linguistic encoding. Below we discuss three related classes of proposals about how production ease affects linguistic encoding. Following that, we discuss accounts of linguistic reduction that refer to communicative goals. This includes a discussion of research on *audience design*. We also discuss more recent communicative accounts that either draw on information theoretic considerations (cf. Shannon, 1948) or the concept of rational (J. R. Anderson, 1990) or boundedly rational cognition (e.g., Simon, 1990).

Production ease and communicative accounts share a focus on online processes that affect production as it is unfolding. This contrasts with *representational* accounts, which have focused on changes in the phonetic representations of words over longer periods of time (e.g., the lifetime of a speaker or even generations of speakers). The majority of psycholinguistic work on reduction and omission has interpreted speakers’ preferences in alternations as providing a window into the mechanisms underlying language production, thereby more or less explicitly assuming the former (e.g., Arnold *et al.*, 2012; Baese-Berk & Goldrick, 2009; Bard *et al.*, 2000; Ferreira & Dell, 2000; Gahl *et al.*, 2012). In research on speech production, however, phonological and phonetic reduction is often described as the result of changes to phonological representations (e.g., Bybee & Hopper, 2001; Kohler, 1990; Pierrehumbert, 2001, 2002; Wedel, 2006; Zipf, 1929; for additional references, see Ernestus, 2014). Following our discussion of production ease and communicative accounts, we turn to this third type of account of reduction mentioned above, representational accounts. We discuss the relation between such offline accounts and online accounts of reduction.

Before we turn to these different accounts, we begin with an important caveat.

*Production ease versus communicative goals: Mutually exclusive?*

Although it is helpful for the purpose of exposition to group accounts of reduction into broad classes of competing positions, production ease and communicative accounts are arguably better seen as defining a continuum of perspectives. For example, some communicative accounts do not argue against the idea that the resource demands inherent to linguistic encoding affect speakers’ production preferences. Rather, speakers’ preferences are assumed to also be affected by communicative considerations. Specifically, a long-standing idea holds that language production is subject to competing pressures—on the one hand, speakers want to achieve their communicative goals, on the other, they have limited resources (e.g.,