The Oxford Handbook of
THE MENTAL LEXICON
THE OXFORD HANDBOOK OF

THE MENTAL LEXICON
THE OXFORD HANDBOOK OF

THE MENTAL LEXICON

Edited by
ANNA PAPAFRAGOU
JOHN C. TRUESWELL
and
LILA R. GLEITMAN

OXFORD UNIVERSITY PRESS
In loving memory of Lila
Contents

List of Abbreviations xi
Contributors xv

1. Introduction 1
   Anna Papafragou, John C. Trueswell, and Lila R. Gleitman

PART I REPRESENTING THE MENTAL LEXICON

PART IA FORM

2. Phonological abstraction in the mental lexicon 11
   Eric Baković, Jeffrey Heinz, and Jonathan Rawski

3. Phonological variation and lexical form 33
   Ruaridh Purse, Meredith Tamminga, and Yosiane White

4. Neural encoding of speech and word forms 53
   David Poeppel and Yue Sun

PART IB MEANING

5. Morphology and the mental lexicon: Three questions about decomposition 77
   David Embick, Ava Creemers, and Amy J. Goodwin Davies

6. Syntax and the lexicon 98
   Artemis Alexiadou

7. Lexical semantics 126
   Ray Jackendoff

8. Logic and the lexicon: Insights from modality 151
   Valentine Hacquard
PART IC INTERFACES AND BOUNDARIES

9. Pragmatics and the lexicon
   Florian Schwarz and Jérémy Zehr
   173

10. Efficient communication and the organization of the lexicon
    Kyle Mahowald, Isabelle Dautriche, Mika Braginsky, and Edward Gibson
    200

11. Compositionality of concepts
    Gala Stojnic and Ernie Lepore
    221

12. Language and thought: The lexicon and beyond
    Barbara Landau
    236

PART II ACQUIRING THE MENTAL LEXICON

PART IIA FORM

13. Infants’ learning of speech sounds and word-forms
    Daniel Swingley
    267

14. Learning words amidst speech sound variability
    Sarah C. Creel
    292

PART IIB MEANING

15. How learners move from sound to morphology
    Katherine Demuth
    313

16. Systematicity and arbitrariness in language: Saussurean rhapsody
    Charles Yang
    327

17. Children’s use of syntax in word learning
    Jeffrey Lidz
    356

18. Easy words: Reference resolution in a malevolent referent world
    Lila R. Gleitman and John C. Trueswell
    378

19. Early logic and language
    Stephen Crain
    401
PART IIC  INTERFACES AND BOUNDARIES

20. Contributions of pragmatics to word learning and interpretation 421
   Myrto Grigoroglou and Anna Papafragou

21. Differences in vocabulary growth across groups and individuals 438
   Christine E. Potter and Casey Lew-Williams

PART III  ACCESSING THE MENTAL LEXICON

PART IIIA  VIA FORM

22. Spoken word recognition 461
   James S. Magnuson and Anne Marie Crinnion

23. Word-meaning access: The one-to-many mapping from form to meaning 491
   Jennifer M. Rodd

24. Learning and using written word forms 506
   Rebecca Treiman and Brett Kessler

PART IIIB  VIA MEANING

25. The dynamics of word production 521
   Oriana Kilbourn-Ceron and Matthew Goldrick

26. The neural basis of word production 536
   Nazbanou Nozari

PART IIIC  INTERFACES AND BOUNDARIES

27. The structure of the lexical item and sentence meaning composition 561
   Maria Mercedes Piñango

28. On the dynamics of lexical access in two or more languages 583
   Judith F. Kroll, Kinsey Bice, Mona Roxana Botezatu, and Megan Zirnstein
29. Lexical representation and access in sign languages
   Rachel I. Mayberry and Beatrijs Wille

30. Disorders of lexical access and production
   Daniel Mirman and Erica L. Middleton

References

Index
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>±1</td>
<td>First person feature</td>
</tr>
<tr>
<td>±2</td>
<td>Second person feature</td>
</tr>
<tr>
<td>±N</td>
<td>Nominal feature</td>
</tr>
<tr>
<td>±past</td>
<td>Past tense feature</td>
</tr>
<tr>
<td>±perf</td>
<td>Perfect aspect feature</td>
</tr>
<tr>
<td>±pl</td>
<td>Plural number feature</td>
</tr>
<tr>
<td>A.pass</td>
<td>Adjectival passive</td>
</tr>
<tr>
<td>A1</td>
<td>Primary auditory cortex</td>
</tr>
<tr>
<td>ACC</td>
<td>Anterior cingulate cortex</td>
</tr>
<tr>
<td>Acc</td>
<td>Accusative</td>
</tr>
<tr>
<td>Act</td>
<td>Active</td>
</tr>
<tr>
<td>AG</td>
<td>Angular gyrus</td>
</tr>
<tr>
<td>ALE</td>
<td>Activation likelihood estimate</td>
</tr>
<tr>
<td>ASD</td>
<td>Autism spectrum disorders</td>
</tr>
<tr>
<td>ASL</td>
<td>American Sign Language</td>
</tr>
<tr>
<td>ATL</td>
<td>Anterior temporal lobe</td>
</tr>
<tr>
<td>AxS</td>
<td>Analysis by synthesis</td>
</tr>
<tr>
<td>BSL</td>
<td>British Sign Language</td>
</tr>
<tr>
<td>CC</td>
<td>Compositionality constraint</td>
</tr>
<tr>
<td>DAS</td>
<td>Deletion, addition, or substitution</td>
</tr>
<tr>
<td>DGS</td>
<td>Deutsche Gebärdensprache (German Sign Language)</td>
</tr>
<tr>
<td>DIVA</td>
<td>Directions into velocities of articulators</td>
</tr>
<tr>
<td>DN</td>
<td>Double negation</td>
</tr>
<tr>
<td>DS</td>
<td>Down syndrome</td>
</tr>
<tr>
<td>DS</td>
<td>Default to stereotype</td>
</tr>
<tr>
<td>EARSHT</td>
<td>Emulation of Auditory Recognition of Speech by Humans Over Time</td>
</tr>
<tr>
<td>ECoG</td>
<td>Electrocorticography</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ERF</td>
<td>event-related field</td>
</tr>
<tr>
<td>ERP</td>
<td>event-related potential</td>
</tr>
<tr>
<td>F0</td>
<td>fundamental frequency</td>
</tr>
<tr>
<td>FC</td>
<td>composition function</td>
</tr>
<tr>
<td>FI</td>
<td>interpretation function</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FWNP</td>
<td>frequency-weighted neighborhood probability</td>
</tr>
<tr>
<td>GODIVA</td>
<td>gradient order directions into velocities of articulators</td>
</tr>
<tr>
<td>HG</td>
<td>Heschl's gyrus</td>
</tr>
<tr>
<td>IAC</td>
<td>interactive activation and competition</td>
</tr>
<tr>
<td>IFG</td>
<td>inferior frontal gyrus</td>
</tr>
<tr>
<td>IFS</td>
<td>inferior frontal sulcus</td>
</tr>
<tr>
<td>Intr</td>
<td>intransitive</td>
</tr>
<tr>
<td>IPA</td>
<td>International Phonetic Alphabet</td>
</tr>
<tr>
<td>IPC</td>
<td>inferior parietal cortex</td>
</tr>
<tr>
<td>ITG</td>
<td>inferior temporal gyrus</td>
</tr>
<tr>
<td>JSL</td>
<td>Japanese Sign Language</td>
</tr>
<tr>
<td>L1</td>
<td>First Language</td>
</tr>
<tr>
<td>L2</td>
<td>Second Language</td>
</tr>
<tr>
<td>LCFcC</td>
<td>lexically mediated compensation for coarticulation</td>
</tr>
<tr>
<td>LSC</td>
<td>Llengua de signes catalana (Catalan Sign Language)</td>
</tr>
<tr>
<td>LSE</td>
<td>Lengua de Signos Española (Spanish Sign Language)</td>
</tr>
<tr>
<td>LSF</td>
<td>Langue des Signes Française (French Sign Language)</td>
</tr>
<tr>
<td>LSTM</td>
<td>long short-term memory</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>MMN</td>
<td>mismatch negativity</td>
</tr>
<tr>
<td>MSS</td>
<td>metrical segmentation strategy</td>
</tr>
<tr>
<td>MTG</td>
<td>middle temporal gyrus</td>
</tr>
<tr>
<td>Nact</td>
<td>non active</td>
</tr>
<tr>
<td>NAM</td>
<td>Neighborhood Activation Model</td>
</tr>
<tr>
<td>NC</td>
<td>negative concord</td>
</tr>
<tr>
<td>NegP</td>
<td>negation phrase</td>
</tr>
<tr>
<td>Nom</td>
<td>nominative</td>
</tr>
<tr>
<td>Pass</td>
<td>passive</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PC</td>
<td>predictive coding</td>
</tr>
<tr>
<td>Perf.</td>
<td>perfective</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>POA</td>
<td>place of articulation</td>
</tr>
<tr>
<td>PRED</td>
<td>predicate phrase</td>
</tr>
<tr>
<td>PreSMA</td>
<td>pre-supplementary motor areas</td>
</tr>
<tr>
<td>PWA</td>
<td>people with aphasia</td>
</tr>
<tr>
<td>SD</td>
<td>semantic dementia</td>
</tr>
<tr>
<td>SES</td>
<td>Socioeconomic status</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SMA</td>
<td>supplementary motor area</td>
</tr>
<tr>
<td>SMG</td>
<td>supramarginal gyrus</td>
</tr>
<tr>
<td>SMP</td>
<td>simplified mapping perspective</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SpT</td>
<td>Sylvian parietal temporal (parietal temporal junction)</td>
</tr>
<tr>
<td>SSM</td>
<td>selection modification model</td>
</tr>
<tr>
<td>STG</td>
<td>superior temporal gyrus</td>
</tr>
<tr>
<td>STS</td>
<td>superior temporal sulcus</td>
</tr>
<tr>
<td>SUB</td>
<td>subject phrase</td>
</tr>
<tr>
<td>svPPA</td>
<td>semantic variant of primary progressive aphasia</td>
</tr>
<tr>
<td>SWR</td>
<td>spoken word recognition</td>
</tr>
<tr>
<td>TİD</td>
<td>Türk İşaret Dili (Turkish Sign Language)</td>
</tr>
<tr>
<td>TMS</td>
<td>transcranial magnetic stimulation</td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
</tr>
<tr>
<td>V.pass</td>
<td>verbal passive</td>
</tr>
<tr>
<td>VLPFC</td>
<td>ventrolateral prefrontal cortex</td>
</tr>
<tr>
<td>vMC</td>
<td>ventral primary motor cortex</td>
</tr>
<tr>
<td>VOT</td>
<td>voice onset time</td>
</tr>
<tr>
<td>vPMC</td>
<td>ventral premotor cortex</td>
</tr>
<tr>
<td>WS</td>
<td>Williams syndrome</td>
</tr>
</tbody>
</table>
Artemis Alexiadou is Professor of English Linguistics at the Humboldt University in Berlin and Vice Director of Leibniz-Centre General Linguistics (ZAS) in Berlin. Her research is concerned with the syntax and morphology of noun phrases and argument alternations, on which she has published several articles and books. Alexiadou currently serves on the editorial board of Glossa, English Language and Linguistics, Linguistic Analysis, Linguistic Inquiry, Natural Language and Linguistic Theory, and Syntax. She is a winner of the 2014 Gottfried-Wilhelm Leibniz Prize, awarded by the German Research Foundation. She has served as Chairperson of GLOW.

Eric Baković is Professor and Chair of the Linguistics Department at UC San Diego. His research is in phonological theory, with a focus on comparing and assessing how different theoretical frameworks address foundational questions of phonological representation and computation.

Kinsey Bice, formerly a postdoctoral fellow at the University of Washington's Institute for Learning and Brain Sciences and the Department of Psychology, now works as a language data researcher on the Alexa AI team at Amazon. Her research focuses on individual differences in adult language learning, their neural underpinnings, and how to improve learning for adults using neurotechnology. She is a member of the Cognitive Science Society and the Cognitive Neuroscience Society.

Mona Roxana Botezatu is Assistant Professor in the Department of Speech, Language and Hearing Sciences at the University of Missouri. Her research focuses on understanding the dynamics of bilingual lexical access, the cognitive mechanisms that support L2 acquisition, and the consequences of bilingualism on native language performance.

Mika Braginsky, a graduate student in Brain and Cognitive Sciences at MIT, works in Ted Gibson's language lab and is interested in language development, psycholinguistics, and developing computational tools in those areas.

Stephen Crain is Emeritus Professor of Linguistics at Macquarie University. He was an Australian Research Council (ARC) Federation Fellow, and Director of the ARC Centre of Excellence in Cognition and its Disorders. He is a fellow of the Academy of Social Sciences in Australia. He previously served as head of the Linguistics Department at the University of Connecticut and the University of Maryland. His research focuses on children's acquisition of logical expressions, with an emphasis on the logical structures of two typologically distant languages, English and Mandarin Chinese.
Sarah S. Creel is Professor of Cognitive Science at UC San Diego where she directs the Language Acquisition and Sound Recognition Lab. Her research centers around how people form memory representations of sound patterns, including speech sounds, words, voices, accents, and music, and how sound pattern representations change from childhood to adulthood. She uses a combination of eye-tracking and behavioral measures to address these research avenues. Creel has served as Associate Editor at the Journal of Experimental Psychology: Learning, Memory, and Cognition and organized the 2017 conference of the Society for Music Perception and Cognition.

Ava Creemers is a postdoctoral researcher in the Psychology of Language Department at the Max Planck Institute for Psycholinguistics. She received her PhD in linguistics from the University of Pennsylvania. Her dissertation research focused on lexical processing and, in particular, the processing of multi-morphemic words that are semantically opaque. She is interested more broadly in language comprehension, including language processing above the word-level.

Anne Marie Crinnion is a graduate student in the Department of Psychological Sciences at the University of Connecticut, where she is a Jorgenson Fellow and a member of the Cognitive Neuroscience of Communication and Science of Learning & Art of Communication interdisciplinary training programs. Her research focuses on how acoustic, lexical, and semantic levels of processing interact in the time course of speech processing, from behavioral, neural, and computational perspectives.

Isabelle Dautriche is a cognitive scientist in the Centre National de la Recherche Scientifique (CNRS) at Aix-Marseille University (France). Her research focuses on language acquisition and its link to the structure of languages and uses a range of computational and experimental methods with human infants and animals.

Katherine Demuth is Distinguished Professor of Linguistics at the Macquarie University where she is Director of both the Child Language Lab and the Centre of Language Sciences (CLaS). Her research focuses on how children acquire morphology, showing how their growing competence with phonological/prosodic structure influences their comprehension, processing, and production of grammatical morphology. She is an ARC Laureate Fellow, a fellow of the Academy of Social Sciences in Australia (FASSA), and a fellow of the Royal Society of NSW (FRSN).

David Embick is Professor and Chair in the Linguistics Department at the University of Pennsylvania. His research in theoretical linguistics concentrates on syntactic approaches to morphology and their connections with other parts of grammar, including phonology and argument/event structure. Other research interests include questions concerning the architecture of the language faculty, experimental approaches to lexical access and representation, and employing neuroimaging and experimental techniques to examine language impairments in different clinical child populations.
Ted Gibson is Professor of Cognitive Science at the Department of Brain and Cognitive Sciences at MIT. He competed for Canada at the 1984 Olympics in rowing, coming seventh in the coxless 4. His research examines how language is processed and how language processing constraints affect language structure.

Lila R. Gleitman is Professor Emerita of Psychology and (by secondary appointment) Linguistics at the University of Pennsylvania. She has conducted foundational work on language acquisition, and the relationship between language and other cognitive systems. She is an elected member of the National Academy of Sciences, an elected fellow of the American Association for the Advancement of Science, a Fyssen Foundation laureate, and a winner of the Rumelhart Prize in Theoretical Foundations of Human Cognition. She has served as past President of the Linguistic Society of America and the Society for Philosophy and Psychology.

Matthew Goldrick is Professor in the Department of Linguistics and (by courtesy) Psychology at Northwestern University. His research draws on behavioral experiments, acoustic analysis, and computational modeling to develop theories of the cognitive and neural mechanisms underlying the production, perception, and acquisition of sound structure in mono- and multilingual speakers. He currently serves on the editorial boards of Cognitive Neuropsychology, Journal of Memory and Language, Laboratory Phonology, and Phonology.

Amy J. Goodwin Davies’ research in linguistics focuses on the mental representations involved in morphological processing, integrating approaches from psycholinguistics and theoretical morphology. She completed her PhD at the University of Pennsylvania in 2018. Goodwin Davies applies the research and data analysis skills that she developed in her graduate work to her current role as a data scientist at the Children’s Hospital of Philadelphia.

Myrto Grigoroglou is Assistant Professor in the Department of Linguistics and the Cognitive Science Program at the University of Toronto. She obtained her PhD in Linguistics and Cognitive Science at the University of Delaware and was a postdoctoral fellow in the Department of Applied Psychology and Human Development at the University of Toronto (OISE). Her research focuses on understanding how children acquire meaning in their native language, and how children and adults produce and comprehend language in conversation.

Valentine Hacquard is Professor in the Department of Linguistics at the University of Maryland, College Park. Her research focuses on formal semantics and pragmatics, with a focus on modality and attitude reports, and their acquisition. Hacquard is currently an Associate Editor at the Journal of Semantics, and on the editorial board of Semantics and Pragmatics.

Jeffrey Heinz is Professor of Linguistics at Stony Brook University with a joint appointment in the Institute for Advanced Computational Science. He conducts research in theoretical phonology and several related areas including theoretical and
mathematical linguistics, theoretical computer science, computational learning theory, cognitive science, and artificial intelligence.

**Ray Jackendoff** is Seth Merrin Professor Emeritus of Philosophy and former Co-director of the Center for Cognitive Studies at Tufts University, and Research Affiliate in the Department of Brain and Cognitive Sciences at MIT. He has written extensively on syntax, semantics, morphology, music cognition, social cognition, consciousness, and the architecture of the language faculty and its place in the mind. He is a recipient of the Jean Nicod Prize in Cognitive Philosophy and of the Rumelhart Prize in Theoretical Foundations of Human Cognition, and he has served as President of both the Linguistic Society of America and the Society for Philosophy and Psychology.

**Brett Kessler** is Associate Professor Emeritus at Washington University in St. Louis, having taught in the Linguistics and the Philosophy-Neuroscience-Psychology programs. Currently he serves as Senior Research Scientist in Psychological and Brain Sciences, where he studies the acquisition of reading and writing systems, with emphasis on the implicit learning of statistical patterns.

**Oriana Kilbourn-Ceron** is a postdoctoral fellow in the Department of Linguistics at Northwestern University. Her research focuses on phonetic and phonological variation, and how pronunciation is shaped by the process of speech planning. She has published quantitative analyses of spontaneous speech corpora, behavioral experiments on speech production, and theoretical linguistic analyses. Her 2017 dissertation was recognized with McGill’s Arts Insight Award for excellence in doctoral research in the social sciences.

**Judith F. Kroll** is Distinguished Professor in the School of Education at the University of California, Irvine and former Director of the Center for Language Science at Pennsylvania State University. Her research uses the tools of cognitive neuroscience to examine bilingualism. She is a fellow of the American Association for the Advancement of Science, the American Psychological Association, the Association for Psychological Science, the Psychonomic Society, the Society of Experimental Psychologists, and the American Academy of Arts and Sciences. She was a founding editor of *Bilingualism: Language and Cognition* (Cambridge University Press), and one of the founding organizers of Women in Cognitive Science, a group developed to promote the advancement of women in the cognitive sciences.

**Barbara Landau** is the Dick and Lydia Todd Professor of Cognitive Science at Johns Hopkins University (secondary appointments in Psychological and Brain Sciences and Neuroscience). Her research focuses on spatial representation, language, and the relationship between these two systems of knowledge in typically and atypically developing individuals. She is an elected member of the National Academy of Sciences, an elected fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science, and the Cognitive Science Society. She was named a
Guggenheim Fellow and received the William James Fellow Award from the Association for Psychological Science.

**Ernie Lepore** is Board of Governors Professor of Philosophy at Rutgers University. He is the author of numerous books and papers in the philosophy of language, philosophical logic, metaphysics, and philosophy of mind, including *Imagination and Convention* (with Matthew Stone), *Meaning, Mind and Matter: Philosophical Essays* (with Barry Loewer), *Liberating Content and Language Turned on Itself, Insensitive Semantics* (both with Herman Cappelen), *Donald Davidson: Meaning, Truth, Language and Reality*, and *Donald Davidson’s Truth-Theoretic Semantics* (both with Kirk Ludwig), and *Meaning and Argument*. He has co-authored, with Jerry Fodor, *Holism: A Shopper’s Guide* and *The Compositionality Papers*; and with Sarah-Jane Leslie *What Every Student Should Know*.

**Casey Lew-Williams** is Professor in the Department of Psychology at Princeton University. He works with postdocs, graduate students, and undergraduates in the Princeton Baby Lab to study how young children learn, and his research focuses in particular on language learning. He is a Chief Editor of *Frontiers for Young Minds* and a Co-founder of ManyBabies.

**Jeffrey Lidz** is a Distinguished Scholar-Teacher and Professor of Linguistics at the University of Maryland. His research explores language acquisition from the perspective of comparative syntax and semantics, focusing on the relative contributions of experience, extralinguistic cognition, and domain-specific knowledge in learners’ discovery of linguistic structure. Lidz was the Co-editor of the *Oxford Handbook of Developmental Linguistics* (2016) and was the Editor in Chief of *Language Acquisition* from 2012 to 2020. He currently serves on the editorial boards of *Syntax, Journal of Semantics, Semantics & Pragmatics, Frontiers in Language Science*, and the *Journal of South Asian Linguistics*.

**James S. Magnuson** is Professor in the Department of Psychological Sciences at the University of Connecticut and the Director of UConn’s interdisciplinary program in Science of Learning & Art of Communication. He also leads the Computational Neuroscience group at the Basque Center on Cognition, Brain, and Language in Donostia-San Sebastián, Spain, and is a Basque Foundation for Science (Bilbao, Spain) Ikerbasque Research Professor. His research focuses on theoretical, computational, and empirical aspects of spoken and written language.

**Kyle Mahowald** is Assistant Professor in Linguistics at the University of Texas at Austin. His research interests are in computational psycholinguistics, the relationship between NLP and linguistics, and experimental methods in the behavioral sciences.

**Rachel I. Mayberry** is Professor in the Department of Linguistics at the University of California San Diego and directs the Laboratory for Multimodal Language Development funded by NIDCS and NSF. She is also affiliated with the Centers for Research on Language, Research and Teaching in Anthropogeny, Human Development Science, and the Joint Doctoral Program in Language and Communication Disorders.
Her primary research investigates the psycho- and neurolinguistic correlates of the critical period for language using sign language; other research investigates language, gesture, and reading development. Mayberry serves on the editorial board of the *Journal of Child Language* and received the Distinguished Alumni Award for Research Leadership from the SCSD of McGill University.

**Erica L. Middleton** is Institute Scientist and Director of the Language and Learning Lab at Moss Rehabilitation Research Institute. Her research focuses on how words are mentally represented and produced, both in healthy speakers and in people with language impairment due to stroke (aphasia). A major emphasis in her work is to delineate how treatments for language impairment can be designed in accord with fundamental principles of human learning to maximize and sustain recovery. Dr. Middleton’s research is supported by the National Institutes of Health and the Einstein Society (Einstein Healthcare Network, Philadelphia, PA).

**Daniel Mirman** is Senior Lecturer in Psychology of Language at the University of Edinburgh. His research focuses on the neuroanatomy of spoken language processing, language deficits in post-stroke aphasia, and the organization of semantic knowledge. Mirman currently serves on the editorial boards of *Cognitive Science* and *PLOS ONE*.

**Nazbanou Nozari** is Associate Professor in the Department of Psychology at Carnegie Mellon University. Her research focuses on the cognitive and neural basis of language production and monitoring. She is currently Associate Editor for *Psychonomic Bulletin & Review*, and the 2020 winner of the American Psychological Association’s APA Distinguished Scientific Awards for an Early Career Contribution to Psychology in Cognition and Human Learning.

**Anna Papafragou** is Professor in the Department of Linguistics at the University of Pennsylvania and the Director of the University’s interdisciplinary graduate program in Language and Communication Sciences. Her research focuses on how children acquire meaning in language, how language is used and understood, and how language interfaces with human perception and cognition. Papafragou is currently Associate Editor of *Language Learning and Development* and serves on the editorial board of *Language Acquisition, Semantics and Pragmatics*, and the *Yearbook of Linguistic Variation*. She is an elected fellow of the Association for Psychological Science and serves on the Governing Board of the Cognitive Science Society.

**Maria Mercedes Piñango** is Associate Professor in the Department of Linguistics and the Interdepartmental Neuroscience Program at Yale University. She directs Yale’s Language and Brain Lab. Her research focuses on the structure of linguistic meaning and its neurocognitive embedding through the study of real-time processing, intra and cross-dialectal variation, brain functional organization, and meaning change dynamics. Piñango currently serves as Associate Editor for *Language and Computation*, a component of Frontiers in Artificial Intelligence.
David Poeppel is Professor in the Departments of Psychology and Neural Science at NYU and the Director of the Neuroscience Department at the Max Planck Institute for Empirical Aesthetics in Frankfurt, Germany. He is Co-director of the Max-Planck-NYU Center for Language, Music, and Emotion. His research focuses on the brain basis of hearing, speech, and language. He is an elected fellow of the American Association for the Advancement of Science.

Christine E. Potter is Assistant Professor in the Department of Psychology at The University of Texas at El Paso. Her research explores the role of experience in language learning, with an emphasis on learning across different communities and environments.

Ruaridh Purse is a PhD candidate in the Department of Linguistics at the University of Pennsylvania. His research explores the structural and representational properties of sociolinguistic variables with a focus on the interface between phonetics and phonology. In particular, his work involves using instrumental techniques from articulatory phonetics to investigate the details of phonetic implementation.

Jonathan Rawski is Assistant Professor of Linguistics at San José State University. His research concerns the mathematics of language and learning, applied to both human cognition and artificial intelligence systems. He received his PhD from Stony Brook University, where he was also a member of the Institute for Advanced Computational Science, and the Center for Neural Circuit Dynamics.

Jennifer M. Rodd is Professor of Cognitive Psychology at University College London (UK). She has worked at UCL since 2003. Her research aims to characterize the cognitive and neural processes that support human communication—specifically, to understand the complex computations involved in decoding the meanings of spoken and written words. Her lab has primarily been funded by the UK Economic and Social Research Council with additional funding from the Leverhulme Trust and the Experimental Psychology Society. Prior to joining UCL, she was a research fellow at the University of Cambridge (UK).

Florian Schwarz is Associate Professor and Undergraduate Chair of Linguistics at the University of Pennsylvania, where he also serves as the Associate Director for Education of MindCORE, Penn’s hub for the integrative study of the mind. His research is concerned with the study of meaning, combining formal tools from semantics and pragmatics in theoretical linguistics with experimental methods from psycholinguistics to study phenomena such as presuppositions, implicatures, and definite descriptions. He was co-Editor of the Pragmatics and Semantics section of the Language and Linguistics Compass from 2017–2020, and is an Associate Editor of the open access journal Glossa Psycholinguistics founded in 2021. He serves on the editorial boards of Semantics and Pragmatics as well as Natural Language Semantics. Together with Anna Papafragou, he initiated the interdisciplinary conference Experiments in Linguistic Meaning. With Jérémy Zehr, he maintains PCIbex, a free and open access online experiment platform.
Gala Stojnic is a developmental cognitive researcher, interested in how young children and infants represent the surrounding world in the course of their development. Her research focuses on the nature and development of core cognitive systems that allow for efficient learning about agents, objects, spaces, and other domains of knowledge. Gala is currently a postdoctoral research scientist at New York University. She earned her MS and PhD from Rutgers University, where she studied the intersection between social cognition and word learning.

Yue Sun is a researcher in the Neuroscience Department of the Max Planck Institute for Empirical Aesthetics in Frankfurt. He obtained his PhD in cognitive neuroscience at the University of Paris VI. His current research focuses on linguistic regularities in the lexicon and speech of various languages and how these regularities interface with computational properties of speech operations as well as language learning. His work combines quantitative analyses of linguistic corpora, experimental psychology, and neuroimaging.

Daniel Swingley is Professor and presently Chair of the Department of Psychology at the University of Pennsylvania. His research concerns the beginning of language learning in infancy, particularly in the domains of sound learning and word learning. Swingley is Associate Editor of the journal Language Learning and Development, and Treasurer of the Society for Language Development. He serves on the editorial boards of several journals and is a winner of the Lindback Award, Penn’s highest award for teaching.

Meredith Tamminga is Associate Professor in the Department of Linguistics at the University of Pennsylvania. Her research investigates the psycholinguistic mechanisms involved in sociolinguistic production and perception across different levels of grammar as well as the role of such mechanisms in language change. She directs the Language Variation and Cognition Lab at Penn and is currently the Associate Director of Outreach for MindCORE, Penn’s hub for the integrative study of the mind.

Rebecca Treiman is the Burke and Elizabeth High Baker Professor in the Department of Psychological and Brain Sciences at Washington University in St. Louis. She conducts research on language processing and language development, with a special focus on reading and spelling and their acquisition. She is the past Editor in Chief of the Journal of Memory and Language, an elected fellow of the Association for Psychological Science, and a winner of the Distinguished Scientific Contributions Award from the Society for the Scientific Study of Reading.

John C. Trueswell is Professor in the Department of Psychology at the University of Pennsylvania and the Co-director of the University’s MindCORE initiative in Integrative Language Science and Technology. His research focuses on understanding how children develop the ability to process language in real-time, and how this ability interacts with the acquisition of language. Trueswell is currently Associate Editor of Language Learning and Development and serves on the editorial board of Cognition. He is an
elected fellow of the Association for Psychological Science, the American Association for the Advancement of Science, and the Cognitive Science Society.

**Yosiane White** is a PhD candidate in the Department of Linguistics at the University of Pennsylvania. Her research investigates the mental representation of morphological and phonological variables. She uses experimental auditory priming paradigms to tease apart aspects of shared representation of multi-morphemic words containing variable suffixes.

**Beatrijs Wille** is a postdoctoral researcher in linguistics at Ghent University (Belgium). She has a background in speech, language, and hearing sciences, and in linguistics. During her interdisciplinary doctorate, she conducted foundational work on Flemish Sign Language (VGT) acquisition by focusing on the early development of VGT, and the visual communication of parents with deaf children. Later she joined the Mayberry Laboratory for Multimodal Language Development, and worked as a visiting researcher at other highly regarded research centers such as the VL2 lab at Gallaudet University (USA) and the Language and Communication across Modalities Laboratory (Italy). Her postdoctoral research focuses on early literacy of deaf children in Flanders.

**Charles Yang** teaches linguistics and directs the Program in Cognitive Science at the University of Pennsylvania. His honors include a Guggenheim fellowship and the Leonard Bloomfield Award from the Linguistic Society of America for his most recent book *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language* (MIT Press 2016).

**Jérémy Zehr** is an application developer for the Linguistic Data Consortium at the University of Pennsylvania. His research investigates how the various contexts in which people use words and sentences influence their short- and long-term understanding of those words and sentences. Zehr is also the designer of PCIbex, a platform designed to develop and run online experiments.

**Megan Zirnstein** is Assistant Professor in the Department of Linguistics and Cognitive Science at Pomona College. Her research focuses on the interactions between cognitive control and reading comprehension ability across the lifespan and across monolingual and bilingual contexts. Her interdisciplinary work utilizes behavioral, psychophysiological, and neurophysiological techniques to understand and develop theoretical explanations for the impact that bilingual experience and second language learning have on cognition. Zirnstein was previously a National Science Foundation PIRE and SBE postdoctoral research fellow at the Pennsylvania State University, and research scientist at the University of California, Riverside.
CHAPTER 1

INTRODUCTION

ANNA PAPAFRAGOU, JOHN C. TRUESWELL, AND LILA R. GLEITMAN

And Adam gave names to all cattle, and to the fowl of the air, and to every beast of the field.

*Genesis 2:20*

What’s in a name? That which we call a rose by any other name would smell as sweet.

W. Shakespeare, *Romeo and Juliet*

Don't gobblefunk around with words.

R. Dahl, *The BFG*

Throughout history, human beings have been fascinated, enlightened, misled, and moved by words. From parents marveling at their child’s first sounds, to writers, politicians, and anyone else who has had to engage in a conversation, words are a central part of the human experience. The word, properly understood, lies at the heart of the human social and cognitive mentality. Furthermore, the capacity to invent new words (graphically honored in William Blake's depiction of Adam naming the animals on the cover of this volume) is one of the essential evolutionary gifts to our species. This Handbook is a comprehensive survey of the state of the art on the “mental lexicon,” the representation of language in the mind/brain at the level of individual words and sub-word meaningful units (morphemes).

Despite the centrality of words to our mental life, in the past—even the relatively recent past—the topic of this Handbook would have been considered too particularistic to be of much interest outside the traditional realms of lexicography and etymology. A long line of scholars beginning at least with Saussure (1916) have considered the lexicon to be a repository of exceptions and idiosyncrasies, very much unlike the rule-governed nature of the grammar. In recent years, however, the study of words as mental objects has grown rapidly across several fields, including linguistics, psychology, philosophy, neuroscience, education, and computational cognitive science. As a result, the
mental lexicon is seen in most quarters today as far more structured and integrated with the mental grammar. Relatedly, it is now recognized that many features of language development and psycholinguistic processing have to be couched at the lexical level as well. Despite many present-day theoretical differences within and across individual disciplines, then, by general consensus the lexicon itself has become the nexus of much significant empirical research and theorizing within modern cognitive science. The goal of this Handbook was to connect and synthesize these recent theoretical and empirical advances.

The key issues surrounding the mental lexicon can be articulated in terms of three framing questions: What exactly do you know when you know the lexicon in your native language? How did you come to know it? And how do you put that knowledge to use? Viewed in this way, the mental lexicon epitomizes the very issues that have been posed for the study of human language as a whole (see Chomsky, 1986). The present Handbook is divided into three parts corresponding to the three framing questions above about the representation, acquisition, and processing of lexical knowledge. Each part follows a similar organizational structure: it begins with chapters that discuss issues of form (mostly, phonology); it then moves on to contributions pertaining to meaning (drawn from the domains of morphology, syntax, and semantics); and it concludes by chapters addressing the interface of the lexicon with other linguistic or non-linguistic (cognitive or communicative) systems in ways that typically straddle disciplinary boundaries. Chapters are written by leading authors from a variety of backgrounds who present their own perspective on an aspect of the lexicon within an up-to-date summary of the state of the art.

Part I (‘Representing the Mental Lexicon’) introduces modern linguistic and cognitive theories of how the mind/brain represents words and sub-word units at the phonological, morphological, syntactic, and semantic levels. This part also discusses broad topics at the interface of the lexicon and other linguistic and non-linguistic systems, such as the contribution of pragmatics to lexical interpretation, the role of compositionality, and the relation between words and concepts. Individual chapters comment on highly interrelated questions, including: How are speech and word forms encoded in the mind/brain? How do we map continuously varying acoustic input to discrete mental representations that form the basis for stored words? How abstract are phonological representations of words in the lexicon? How should we capture the fact that words are produced differently by different individuals and by a single individual on different occasions? Do words break down into smaller meaningful units? What is the best way of accounting for the relation between the lexicon and syntax? How does the mind organize our vast knowledge of word meanings, and how does such knowledge relate to the human conceptualization of the world? How is logical vocabulary represented? What is the appropriate division of labor between aspects of meaning that are semantically encoded in words vs. inferred through pragmatic processes in context? What constraints does the process of combining word meanings to form phrasal or sentential meanings place on theories of lexical semantics? What do lexical universalities and diversities say about informational communicative constraints, and generally about human cognition? By adopting distinct but often complementary perspectives, this first
set of chapters outlines core generalizations about the contents of the (adult) mental lexicon across linguistic domains and illustrates the rich connections between the mental lexicon and other linguistic and extra-linguistic systems within mental architecture. This series of chapters also sets the stage for the next sections that focus on how the contents of the mental lexicon are learned and accessed during language processing.

Part II (‘Acquiring the Mental Lexicon’) turns to the process through which children learn the phonological, morphological, syntactic, and semantic properties of words in their native language. This part also discusses contributions of pragmatics to word learning and interpretation as well as the variability observed in lexical learning. Individual chapters address questions such as: How do children acquire the phonological shape of words in their native language, and how do they handle the variability in how words sound across speakers and contexts? How do children begin to learn the internal structure of words and the rules that govern such structure? What is the role of syntax in word learning? What sources of evidence contribute to the acquisition of word meanings through infancy into adulthood? What meanings do child language learners assign to logical expressions, and how do these meanings correspond to classical logic? How do considerations of speaker intention affect how words are learned and used? And what accounts for the fact that vocabulary learning differs both across individual children and across groups of learners? Together, these chapters show how our understanding of the nature and contents of the mental lexicon in adults inform theories of lexical development in children (and, inversely, how developmental data might provide constraints on linguistic and cognitive theories of the mature-state lexicon). The questions at the core of these chapters bear directly on the prior knowledge that the learner does (and does not) bring to the task of language acquisition, and on the relative contribution of learner-driven and environment-driven factors to different aspects of lexical learning. Despite their differences in topics and angles, the authors in this collection of chapters aim to understand how building a mental lexicon requires the coordination of prior knowledge, information processing capacity, and extra-linguistic cognitive and communicative resources; relies on a proper access to and understanding of information available in the linguistic input; and is modulated by the child’s computational and cognitive capacities and limitations at different stages of development.

Finally, Part III (‘Accessing the Mental Lexicon’) examines how the mental lexicon contributes to language use during listening, speaking, and conversation, and includes perspectives from bilingualism, sign languages, and disorders of lexical access and production. The chapters in this part ask foundational questions such as: How are spoken words recognized from the speech signal? How are written words recognized from their orthographic form? What are the processes that are involved in reading, and how do children learn to read and spell? How are words accessed in the mind/brain as people prepare to speak? Does lexical access change when speakers plan and produce multiple words in connected speech? How does stored lexical meaning interact with context as people understand language during conversation? How does bilingualism alter the mental lexicon? Does lexical access and use differ in people learning sign languages? How should we characterize difficulties of lexical access in aphasia and other neurogenic and developmental disorders, as well as in typical aging? Together these chapters
emphasize that mobilizing the mental lexicon during listening and speaking is best understood as a dynamic interaction of information from multiple levels of representation that engages stored linguistic knowledge, information processing capacity, and extra-linguistic cognitive resources. These themes, present throughout the volume, take on special significance here because psycholinguistic studies combine evidence from a broad array of tools, methods and populations to investigate the link between the rapidly unfolding processes of lexical comprehension or production and the underlying mechanisms that guide these processes. As a result, such studies provide particularly promising sources of evidence in adjudicating among competing theories of lexical organization.

As is clear from this brief summary, far from being separate, the three framing questions that organize this Handbook are highly intertwined and the corresponding research programs that have been designed to address lexical representation, acquisition, and access can mutually inform and constrain one another. Within and across the three parts of this volume, for instance, specific proposals about lexical knowledge in mature (adult) minds often rely on evidence from vocabulary acquisition or processing and, inversely, theories of lexical acquisition and access are informed by linguistic and cognitive theories of the contents of the lexicon. Furthermore, even though individual chapters were designed to offer a stand-alone overview of the empirical and theoretical state of the art within a particular domain, throughout the entire book similar issues are addressed in considerable depth across multiple chapters. Recurring themes include, among others, the nature of abstractness in representations of lexical form and meaning (especially in the face of rampant surface variability); the role of experience in lexical representation, acquisition, and processing; the boundary between shared and specific properties of the lexicons in languages of the world (both spoken and sign); the relation between stored lexical knowledge and computations performed over such knowledge; the tension between productivity and arbitrariness; the relation between the lexicon and other levels of representation within the language faculty as well as to non-linguistic cognitive and communicative mechanisms; and the nature and extent of individual and population-level differences in lexical knowledge, acquisition, and processing. The present tri-partite organization of each part of the Handbook in terms of form, meaning, and interfaces or boundaries highlights these common themes and makes it easy to track them throughout the book.

The present collection is unique in offering both a comprehensive overview of the set of phenomena that the study of the mental lexicon is responsible for, and a definitive compendium of the kinds of theories at the representational, learning, and processing level that can adequately describe and explain these phenomena. The picture of the lexicon that emerges from these pages reveals a rapidly changing and dynamic field that spans multiple topics, theories, and methods. It covers a variety of different languages and populations of language users, and moves along multiple levels of analysis from the neural to the social underpinnings of words, and multiple timescales ranging from years (as vocabulary grows from infancy to adulthood) to milliseconds (as lexical information is processed on a moment-to-moment basis). Furthermore, it is clear that the study of
the mental lexicon is defined by complex questions about how the lexicon is psychologically represented that lie at the core of the modern cognitive science of language but defy simplifying assumptions or neat disciplinary labels. To navigate this rich intellectual territory, the present volume brings together theoretical and empirical traditions that until recently were independent of one another. This strongly interdisciplinary approach highlights cutting-edge theoretical and methodological advances but, just as importantly, also identifies areas of disagreement, debate, and inconsistencies or gaps across the field that create opportunities for scientific progress. It is our hope that this book will inspire the current and next generation of researchers in ways that lead to new empirical discoveries and theoretical breakthroughs for the study of words in the mind.

* * *

**Postscript**

As this volume was going to press, Lila Gleitman passed away at the age of 91. The loss of our close friend and collaborator was profound. Lila was an inspiration to us and a guiding light to many, both personally and intellectually. A giant in our field, Lila was legendary for her passion for words (and ideas!)—but also for people and friendships. She worked on this book tirelessly through all its many stages, even choosing a beloved painting for the cover. We will miss her beyond words.

Anna Papafragou and John C. Trueswell

Lila Gleitman (1929–2021)
PART I

REPRESENTING THE MENTAL LEXICON
PART I A

FORM
When we try to understand how we speak and how we listen, an unavoidable fact is that we use the same vocabulary items again and again. So, not only are these items stored in memory but also they are learned on a language-specific basis. In this way, we directly confront the issue of mental representations of speech, that is, something about how we store, retrieve, and remember units of speech. What are these units? What are these mental representations? These questions are central to generative phonology, a fact reflected in the title of a collection of important papers by Morris Halle, a founder of the field: From Memory to Speech and Back Again (Halle, 2003).

In this chapter, we review some of the most basic—and most fascinating—conclusions and open questions in phonology regarding how abstract these mental representations of speech are. In an era of big data and data mining, the prevailing attitude in science and scientific observation seems to be to store every detail, never to assume any detail is irrelevant no matter how small because you never know when it may in fact make a difference somewhere. This attitude is also present in the psychology of language, where it has been shown that people are sensitive to astounding details of aspects of speech reflecting people’s age, gender, social status, neighborhood, and health (Coleman, 2002; Pierrehumbert, 2002; Johnson, 2006). Pierrehumbert (2016) reviews this literature and highlights some theoretical accounts (see also Purse, Tamminga, and White, this volume).
In the context of this prevailing attitude, the theory of generative phonology makes some surprising claims. First, it claims it is necessary to abstract away from much of this detail to explain systematic aspects of the pronunciations of the morphemes, words, and phrases we utter. In other words, while a person is sensitive to subtle phonetic details that are informative about various aspects of a speaker’s condition, the theory of generative phonology claims those very same details are irrelevant to the very same person’s mental representations of the pronunciations of morphemes and words. As such, the mental representations of these pronunciations are particular abstractions. We review the arguments for this position in Section 2.3. Section 2.4 shows how these ideas lead to the phoneme and to distinctive features, a necessarily abstract categories that are said to be the fundamental units out of which the pronunciation of vocabulary items is built.

We then turn to a question posed by Paul Kiparsky: how abstract is phonology? (Kiparsky, 1968)—or, more specifically, how abstract are the mental representations of speech? A vigorous back and forth debate over this question followed, but now, over fifty years later, there is still no consensus regarding the answer. A basic problem is that some languages present good evidence that morphemes are specified with phonological content that is never realized as such in any surface manifestation of those morphemes. At issue is precisely what kinds of evidence justifies such abstract phonological content. In our minds, this question, and others related to it, are particularly important and remain among the most interesting and understudied questions in phonology today.

But first, we begin this chapter with some thoughts on the nature of abstraction and idealization.

### 2.2 What is abstraction?

The first thing to point out about abstraction is not how odd it is but how common it is. Orthographic letters are abstractions. The capital letter “A” and the lowercase letter “a” are quite different and yet at some level of abstraction they are referring to the same thing. Money is an abstraction. Whole numbers and fractions are also abstractions. For example, there is no such thing as “three.” There are only examples of collections of three items. Such abstractions are not only taken for granted but they are also valued: the earlier that toddlers and young children learn such abstractions the more we marvel at their intelligence.

A very common approach to problem solving one learns in grade school is shown in the diagram in Figure 2.1. For example, in an elementary-level math class, students are often given a problem expressed in plain language. Their job is to extract the relevant information, organize it into an equation, solve the equation, and report the answer. Figure 2.2 shows a fourth grade math exercise.
Similarly, when responding to challenges that his law of falling bodies did not correspond to the real world, Galileo used an analogy to accounting:

what happens in the concrete […] happens in the abstract. It would be novel indeed if computations and ratios made in abstract numbers should not thereafter correspond to concrete gold and silver coins and merchandise […]

(quoted in Wootton, 2015, pp. 23–24)

Here Galileo mentions an important feature: one can make computations with abstract numbers without directly referencing the particular objects that implement them. Given a proper abstraction, one can make an inference or calculation at the level of numbers with results that then correspond to some specific physical effect or process. This property, that a given abstraction can have many implementations, is known as multiple realizability (Bickle, 2020). An abstract object is multiply-realizable by a number of concrete objects. The concrete objects might differ from each other in various ways, but the ways in which they differ are irrelevant to the abstraction. Orthographic letters and the mental representations of units of speech are abstract in this way, too.

The second important point about abstraction is that there are degrees of abstraction. Even the question in Figure 2.2 as written in plain language is abstract, since, for example, it uses numbers to explain the situation and not, for instance, a photograph of
the two dozen pigs and/or video of seven pigs being born and two others dying. Figure 2.3 illustrates the concept of degrees of abstraction. The only change we would make to it would be to add question marks after each of the subcaptions in the figure. The question is always “Is the level of abstraction too realistic, too abstract, or just right?”

Answering this question is not easy. It is not at all obvious what the right level of abstraction is. It is appropriate to deliberately consider whether a particular degree of abstraction is appropriate or not. We are in an age of big data and the milieu of the age seems to be to err on the side of “too realistic” and to avoid analyses that are “too abstract.” Our own view is that this is shortsighted. Many things were at one time considered “too abstract” but are now recognized as being perfectly reasonable and in fact essential and useful.

When one studies the history of mathematics, it is interesting how notions once considered to be “too abstract” were considered crazy or useless. This includes things like the number zero, real numbers, \( \sqrt{-1} \), uncountable infinity, number theory, and so on. When Pythagoras and his group realized that \( \sqrt{2} \) could not be expressed as a fraction of whole numbers, they called it “irrational,” literally “unreasonable.” The term sticks today. Cantor went mad after realizing that infinitely-sized sets had different degrees of cardinality; today this fact underlies the Church–Turing thesis of computability. The development of the complex numbers caused much consternation in the 17th and 18th centuries but are routinely used today to understand complex physical systems. Developments in number theory in the early part of the 20th century had no purpose other than to satisfy some strange mathematical aesthetic, and these now underlie secure cryptographic communications, protecting business operations, journalists, and other vital communications.

In short, we are sympathetic to the view put forth by Cheng (2015, p. 22): “Abstraction can appear to take you further and further away from reality, but really you’re getting closer and closer to the heart of the matter.”

In this chapter we show that abstractness is a property exploited both by speakers of natural languages and by scientists describing the linguistic knowledge of those speakers.

---

1 One of the clearest recent expositions on abstractness and its virtues is in Chapter 2 of Cheng’s book, *How to Bake π*, which we encourage readers of the present chapter to read.
2.3 Phonology and the mental lexicon

The pronunciation of a word is a sequence of events in time. From the perspective of speech production, these events are defined articulatorily. From the perspective of speech perception, these events can be defined acoustically and perceptually. The International Phonetic Alphabet (IPA) defines categories of speech sounds articulatorily and provides a symbol for each categorized speech sound. For consonants, these symbols specify three aspects of articulation: the place of articulation, the manner of articulation, and the activity of the vocal folds. Symbols representing vowels primarily specify the degree of jaw aperture, how forward the positioning of the root of the tongue is, and any rounding of the lips. For example, the pronunciation of the word ‘math’ is transcribed in the IPA as [mæθ], as there are three distinct speech sounds in sequence: [m], which is articulated by stopping airflow at the lips but releasing it through the nose, with the vocal folds held together such that they vibrate; [æ], which is an open front unrounded vowel; and [θ], which is articulated by constricting but not stopping airflow with the blade of the tongue between the teeth with the vocal folds spread apart. For most speech acts, these articulations are similar across speakers of the same speech variety, despite individual physiological variation.

For a word like ‘tune,’ one transcription of it using the IPA is [tun], often referred to as a broad transcription. In contrast, a narrow transcription of this word using the IPA is [tʰũːn]. The difference between the broad and narrow transcriptions is the degree of abstraction. Both transcriptions reveal systematic aspects of standard American English speech. However, the broad transcription only includes so-called contrastive information and the narrow transcription includes some non-contrastive information as well, indicated in this case with various diacritic marks: the aspiration on the [t], indicated with [ʰ], and the nasalization [~] and extra duration [ː] of the vowel [u]. Both kinds of transcriptions can be used as abstract representations of the pronunciation of this word stored in memory, and one could ask whether the long-term memory representation of the pronunciation of the word ‘tune’ is more like the broad transcription, the narrow transcription, or something else. One can ask if there is only one long-term memory representation of the pronunciation of the word ‘tune,’ or if there are multiple, possibly partially redundant representations. All of these hypotheses are open to study. In the remainder of this chapter, we use broad IPA transcriptions as we discuss representations of the pronunciations of words. The degree of abstractness chosen is not critical, but we settle here to facilitate additional discussion.

With this in mind, we ask the question: what does modern generative phonological theory say about the mental representations of speech? The central empirical fact that informs this question is that, in many languages of the world, morphemes are pronounced in different ways depending on context. To illustrate with an example, Odden (2014) draws attention to the pattern exhibited by the different verb forms in Kerewe shown in Table 2.1.

There is an interesting difference between the first group of verb forms and the second group of verb forms. In the first group, for example, the pronunciation for the verb
stem ‘adorn’ is consistently [paamba] regardless of whether it is in the infinitival form (prefixed with [ku]), the 1sg habitual form (prefixed with [m]), the 3sg habitual form (prefixed with [a]), or the imperative form (not prefixed). It thus makes sense to assume that /paamba/ represents in a Kerewe speaker’s mental lexicon the major features of the pronunciation of the verb stem ‘adorn’. However, when the same kind of morphological analysis is applied to the forms of the verb ‘hunt’ in the second group of verb forms, we find that the verb stem’s pronunciation in the 1sg habitual form is [pii ɡa] whereas in the other forms it is [hii ɡa]. So the question naturally arises, what long-term memory representation of the pronunciation of the verb stem ‘hunt’ do speakers of Kerewe have?

One possibility is that both /piiɡa/ and /hiiɡa/ are stored, along with the knowledge that [piiɡa] is used in the 1sg habitual form and that [hiiɡa] is used otherwise. This is fine as far as it goes, but it is of interest that the other verbs in this second group pattern exactly the same way: they begin with [p] in the 1sg habitual form, and with [h] otherwise. Furthermore, there are no verb stems in Kerewe that begin with [h] in the 1sg habitual form. Taken together, these observations suggest there is something systematic about the variation in the pronunciation of the various forms of this group of verbs in Kerewe, a systematicity that storage of all pronunciations plus information about their distributions does not readily or insightfully capture.

Table 2.1 Kerewe verbs, from Odden (2014, pp. 88–89)

<table>
<thead>
<tr>
<th>Infinitive</th>
<th>1sg habitual</th>
<th>3sg habitual</th>
<th>Imperative</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>[kupaamba]</td>
<td>[mpaamba]</td>
<td>[apaamba]</td>
<td>[paamba]</td>
<td>‘adorn’</td>
</tr>
<tr>
<td>[kupaaŋga]</td>
<td>[mpaaŋga]</td>
<td>[apaaaŋga]</td>
<td>[paaaŋga]</td>
<td>‘line up’</td>
</tr>
<tr>
<td>[kupima]</td>
<td>[mpima]</td>
<td>[apima]</td>
<td>[pima]</td>
<td>‘measure’</td>
</tr>
<tr>
<td>[kupuupa]</td>
<td>[mpuupa]</td>
<td>[apuupa]</td>
<td>[puupa]</td>
<td>‘be light’</td>
</tr>
<tr>
<td>[kupekeʃa]</td>
<td>[mpekeʃa]</td>
<td>[apekeʃa]</td>
<td>[pekeʃa]</td>
<td>‘make fire w/ stick’</td>
</tr>
<tr>
<td>[kupiinda]</td>
<td>[mpiinda]</td>
<td>[apiinda]</td>
<td>[piinda]</td>
<td>‘be bent’</td>
</tr>
<tr>
<td>[kuhiiɡa]</td>
<td>[mpiigga]</td>
<td>[ahiigga]</td>
<td>[hiigga]</td>
<td>‘hunt’</td>
</tr>
<tr>
<td>[kuheeka]</td>
<td>[mpeeka]</td>
<td>[aheeka]</td>
<td>[heeka]</td>
<td>‘carry’</td>
</tr>
<tr>
<td>[kuhaaŋga]</td>
<td>[mpaaŋga]</td>
<td>[ahaɑŋga]</td>
<td>[haaŋga]</td>
<td>‘create’</td>
</tr>
<tr>
<td>[kuheeba]</td>
<td>[mpeeba]</td>
<td>[aheeba]</td>
<td>[heeba]</td>
<td>‘guide’</td>
</tr>
<tr>
<td>[kuhiima]</td>
<td>[mpiima]</td>
<td>[ahiiima]</td>
<td>[hiima]</td>
<td>‘gasp’</td>
</tr>
<tr>
<td>[kuhuuha]</td>
<td>[mpuuha]</td>
<td>[ahuuha]</td>
<td>[huuha]</td>
<td>‘breathe into’</td>
</tr>
</tbody>
</table>

2 Here and elsewhere in this chapter we follow traditional phonological notation of transcribing the mental representation of speech between slashes and the actual pronunciation within square brackets. When the distinction is immaterial, we use italics.

3 More—and more detailed—arguments against this “morpheme alternant theory” can be found in Kenstowicz and Kisseberth (1979, pp. 180–196).
The methods of modern generative phonology lead analysts to posit that the long-term memory representation of the pronunciation of the verb stem ‘hunt’ that speakers of Kerewe have is /hiiɡa/, and that /h/ representations are transformed to /p/ representations when they immediately follow /m/ representations. Consequently, the 1sg habitual form of ‘hunt’ is [ mpiiɡa] because it derives from /m+hiiɡa/ by application of this phonological transformation. This phonological analysis explains the systematic variation observed because it predicts that every /h/-initial verb stem ought to be realized as [p]-initial when it follows any prefix ending with /m/, such as the 1sg habitual.

There are two key reasons why the mental representation of ‘hunt’ cannot instead be /piiɡa/, with /p/ transformed to /h/ when it immediately follows /not-m/. First, it is assumed that phonological transformations cannot make direct reference to the negation of a class of speech sounds, in this case to ‘not-m.’ Second, and more significantly, it has been independently determined that the members of the first group of verb stems in Table 2.1 all begin with /p/, and yet the putative transformation of /p/ to /h/ immediately following /not-m/ does not apply to those verb forms. Positing that members of the first group of verb stems begin with /p/ and that members of the second group begin with /h/ (transformed to /p/ immediately following /m/) succinctly and systematically distinguishes the two groups.

One argument against this position is that there are cases where it does seem like two distinct pronunciations of a morpheme are stored. A well-known example is the nominative suffix in Korean, which has two pronunciations: [ka], which is suffixed onto vowel-final words, and [i], which is suffixed onto consonant-final words. This alternation is phonologically conditioned, but it is nevertheless a stretch of a phonologist’s imagination to concoct a rule that transforms /ka/ to [i], or /i/ to [ka], or some other machination that relates these two forms of the nominative suffix in some phonological way. Furthermore, there are no other places in the Korean lexicon or language that exemplify any k ~ Ø or a ~ i alternation, where ‘Ø’ denotes the absence of a phone. In other words, a single posited underlying form and rules (whatever they are) would only account for the alternation between [ka] and [i] in the nominative suffix.

So, in Korean, the best analysis of the nominative suffix alternation appears to be a long-term memory representation of the pronunciation of the suffix as {/i/, /ka/} with the choice of which one to select being based on the phonological properties of the stem the suffix attaches to. Given that such examples exist in the world’s languages, the question is: why don’t we just use the same kind of analysis for Kerewe, for example? The answer is the one given above. The systematicity observed in the realization of Kerewe verb stems points to the straightforward phonological analysis in that case, and the lack of systematicity observed in the realization of the Korean nominative suffix points to the alternant selection analysis in that case. For more discussion of phonologically conditioned allomorphy, see Nevins (2011).

4 The “+” indicates the presumed morphological boundary between the 1sg habitual prefix /m/ and the verb stem /hiiɡa/.

PHONOLOGICAL ABSTRACTION IN THE MENTAL LEXICON 17
The argument just made is the basic one for the position that morphemes are stored in long-term memory with a single representation, known as the underlying form. The fact that the same morpheme can be pronounced in different ways depending on context is due to phonological transformations of this underlying form into its various surface forms. The fact that these transformations are systematically applied explains the systematicity in the alternations of the morphemes across the vocabulary of the language.

The phonological analysis in the generative tradition thus comes with two symbiotic parts. The first posits, where systematicity and its explanation demand it, a single mental representation in long-term memory of the pronunciation of a lexical item—that item's underlying form. The second is the transformational part of a phonological grammar that defines how underlying forms are mapped to surface forms, which are more concrete representations of the pronunciation of the lexical item in the particular context in which it is realized. For example, in the case of Kerewe, the underlying form for ‘hunt’ is /hiiɡa/; the underlying form for the 1sg habitual form of ‘hunt’ is /m+hiiɡa/. There is a phonological transformation changing h to p immediately following m, and so the surface form in this case is [mpiɡa].

Once underlying forms and transformations making them distinct from some of their corresponding surface forms are posited, a natural question arises: how distinct can underlying forms be from surface forms? This is the question of abstractness in phonology. We approach this question first from the perspective of two interrelated, fundamental concepts in phonological theory, phonemes and distinctive features, in Section 2.4, setting the stage for discussion of more specific examples of evidence for different types of abstractness in analyses of phonological patterns in Section 2.5.

### 2.4 Phonemes and features

Much of the question of abstractness in phonology concerns the individual units of speech whose concatenation makes up the pronunciation of lexical items. For example, the speech sounds identically transcribed as k in the Kerewe verb stems heeka ‘carry’ and pekeʃa ‘make fire with stick’ are the same at some level of abstraction; they are just being utilized in different contexts in different lexical items. Precise aspects of the articulation or these two instances of k may well exhibit systematic differences, parallel to the way that the (narrowly transcribed) [tʰ] in [tʰɪk] ‘tick,’ the [t] in [stʊk] ‘stick,’ and the [ɾ] in [ˈæɾɪk] ‘attic’ differ systematically in English but are also the same, and at the same level of abstraction as the k in Kerewe.

The idea that people’s mental lexicons make use of a mental alphabet of abstract speech sounds has a very long tradition in linguistics, dating back at least to Pāṇini’s grammar of Sanskrit, likely written in the 6th–5th century BCE. A much more comprehensive historical context can be gained from Anderson (1985); see in particular pp. 270–276.
theories aside, the speech sounds of this abstract mental alphabet are referred to as phonemes. Phonemes are obviously abstractions: they represent groups or categories of speech sounds, abstracting away from many phonetic particulars, such as the systematic and relatively easily perceptible differences between [tʰ], [t], and [ɾ] in the English example above. In this section, we trace the evidence for the psychological reality of phonemes and highlight two prominent views of phonemes: one as an indivisible alphabetic symbol, and the other as a local confluence of more basic units or properties known as distinctive features.

Readers interested in a more comprehensive review and discussion of the phoneme are directed to Dresher (2011), who distinguishes three views of the phoneme as an entity: the phoneme as physical reality, the phoneme as psychological concept, and the phoneme as theoretical fiction. Dresher (2011, p. 245) concludes, “Once we abandon empiricist assumptions about science and psychology, there is no obstacle to considering the phoneme to be a psychological entity.” This is an important framing, because it means that any linguistic work describing the nature and content of the phoneme is by necessity a statement about mental representation, in much the same way as the concept of underlying forms discussed in Section 2.3. So the question becomes, how much abstraction or idealization is involved in such a psychological entity?

Edward Sapir was the first to present evidence for the psychological reality of phonemes. What is perhaps Sapir’s better-known article, “The Psychological Reality of Phonemes” (Sapir, 1933), was preceded by his article in the first issue of the journal Language, “Sound Patterns of Language” (Sapir, 1925). Together, these two articles establish the perspective that (1) the phoneme is a psychological unit of speech, and (2) the character of the phoneme requires taking into account how it functions in the larger phonological context of the language; it cannot be understood solely from investigation of the articulatory or acoustic properties of its surface realizations. Sapir (1925) argued that languages with the same surface inventory of speech sounds could be organized differently at the underlying, phonemic level. Sapir (1933) argued that the same phoneme-level organization can explain the behavior of native speakers of a language, whether it be with respect to errors they make in the perception or production of speech or in terms of the choices they make when devising or using a novel writing system to represent the speech sounds of their language.

Nearly a century later, psychological evidence for phonemes continues to be found and presented in the literature. The first issue of Language in 2020, 95 years after the publication of Sapir’s (1925) article in the very first issue of the same journal, includes an article by William Labov arguing that the regularity of a sound change in Philadelphia is understandable to the extent that speakers of this dialect of English have an abstract mental representation of the front unrounded mid vowel (Labov, 2020).

Jumping back in time to observations by Bloomfield (1933) and Bloch (1941), among others, English unstressed vowels reduce to schwa [ə] in many contexts, making [ə] an allophone of every English vowel phoneme. Thus we have [ˈfɔɡəˌref] ‘photograph,’ with a primary-stressed [o], an unstressed [ə], and a secondary-stressed [æ], alternating with [ˈfətʰəɡəˌfi] ‘photography,’ with primary-stressed [ə] flanked by unstressed [ə]s. It
follows that phonemic representations of English morphemes must include unreduced vowels only, with reduction to [ə] being due to a phonological transformation, and moreover that the underlying form of many morphemes, such as /ˈfɒtɒɡræf/ 'photograph,' will not directly correspond to any of their complete surface manifestations due to the nature of stress assignment in English.6

Phonemic analysis is the set of methods by which a language’s underlying inventory of phonemes is induced from the distribution and behavior of its surface speech sounds, which can of course differ from language to language. Borrowing an example from Hayes (2009, pp. 31–34), English and Spanish have a set of surface speech sounds that can be broadly transcribed as [t d ɾ r], but their distribution and behavior in each language is such that:

a. /ɾ/ is a phoneme distinct from /t/ and /d/ in Spanish, but not in English, where [ɾ] is sometimes a surface realization of /t/ (e.g. [bæt] ‘bat,’ [’bærəɾ] ‘batter’) and other times a surface realization of /d/ (e.g. [sæd] ‘sad,’ [’særəɾ] ‘sadder’), and

b. /ð/ is a phoneme distinct from /t/ and /d/ in English, but not in Spanish, where [ð] is always a surface realization of /d/ (e.g. [un’dɪsko] ‘a record,’ [los’dɪskos] ‘the records’).

From the same surface inventory of speech sounds [t d ɾ r], then, we arrive at a different phonemic inventory in each language: /t d ɾ/ in English and /t d r/ in Spanish, with [ɾ] a conditioned variant of /t d/ in English and [ð] a conditioned variant of /d/ in Spanish.8

These examples illustrate the basis of an important development in 20th-century linguistics: the idea that the phoneme is not the minimal unit after all, and that instead there are subphonemic units—distinctive features—out of which phonemes are built. Generally speaking, distinctive features are used to describe phonologically relevant phonetic distinctions between speech sounds: those phonetic distinctions that are common to similarly behaving sounds, and those that differ between the conditioned variants of phonemes. In Spanish, for example, the complementary distribution of [d] and [ð] is matched by complementary distribution of [b] and [β] and of [ɡ] and [ɣ], and the members of each of these pairs are related to each other in precisely the same way: [b d ɡ] are voiced stops (vibrating vocal folds and stopping of airflow at three different places of articulation), while [β ð ɣ] are voiced continuants (constriction but not stopping of airflow at corresponding places of articulation). In English, the related speech sounds [t

6 Bloch (1941, pp. 281–283) observed that this represents a potential learning challenge, because arriving at the “right” underlying form for a morpheme requires exposure to a sufficient variety of its surface manifestations.

7 Narrower differences include the precise tongue tip position for the articulation of [t d] (alveolar in English, dental in Spanish) and the degree of constriction of [ɾ] (more closed in English, more open in Spanish).

8 Whether /d/ or /ð/ is the right representation of this phoneme in Spanish is a matter of some debate. Harris (1969) says /d/, Baković (1994) says /ð/, while Lozano (1979) opts for underspecification of the difference between the two speech sounds.
Coronal stops, and the conditioned variant \[ r \] is a coronal continuant.

The distinctive feature idea was explored further by the Prague School of phonologists, notably Roman Jakobson and Nikolai Trubetzkoy, for whom contrast between phonemes was a core theoretical premise. Analyzing the systematicity of such contrasts required viewing phonemes as possessing features distinguishing each from the others. The particular features necessary to distinguish phonemes describe the content of that phoneme. In this sense,

\[ \text{any minimal distinction carried by the message confronts the listener with a two-choice situation. Within a given language each of these oppositions has a specific property which differentiates it from all the others. The listener is obliged to choose either between two polar qualities of the same category [...] or between the presence and absence of a certain quality [...]}. \]

The choice between the two opposites may be termed distinctive feature. The distinctive features are the ultimate distinctive entities of language since no one of them can be broken down into smaller linguistic units. The distinctive features combined into one simultaneous or [...] concurrent bundle form a phoneme.

(Jakobson, Gunnar, Fant, and Halle, 1952, p. 3)

Phonological features were intended to be the cognitive connection between the articulatory and perceptual speech systems.

\[ \text{The distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems. In speech, perception detectors sensitive to the property detectors [...] are activated, and appropriate information is provided to centers corresponding to the distinctive feature[s] [...]}. \]

This information is forwarded to higher centers in the nervous system where identification of the utterance takes place. In producing speech, instructions are sent from higher centers in the nervous system to the different feature[s] [...] about the utterance to be produced. The features then activate muscles that produce the states and configurations of different articulators[.]

(Halle, 1983, p. 95)

Over a quarter century later, the same idea informs the neurolinguistics and biolinguistics literatures.

The [...] featurally specified representation constitutes the format that is both the endpoint of perception—but which is also the set of instructions for articulation.

(Poeppel and Idsardi, 2011, p. 179)

Features serve as the cognitive basis of the bi-directional translation between speech production and perception, and are part of the long-term memory representation.
for the phonological content of morphemes, thus forming a memory-action-perception loop [...] at the lowest conceptual level.

(Volenec and Reiss, 2017, p. 270)

Just as a phonemic analysis of a language's phonological patterns reveals its phonemic inventory, so does a featural analysis reveal its distinctive feature inventory. A phoneme consists of an individual combination of distinctive feature values, but there will be some particular distinctive feature value combinations that do not correspond to phonemes in a given language. For example, the phonemic inventory of Russian includes voiceless oral stops /p t k/ and voiced oral stops /b d ɡ/ at the same three places of articulation (labial, coronal, and dorsal), but nasal stops /m n/ at only two of these (labial and coronal) —meaning that the combination of distinctive feature values describing a dorsal nasal stop /ŋ/ does not correspond to a phoneme in Russian.

A phonemic inventory is often presented as a stand-alone entity in the context of a phonological analysis, but phonologists recognize restrictions on the distributions of individual phonemes (see, e.g., Hall, 2013). For example, some argue /ŋ/ is a phoneme of English, but it is only found word-finally (e.g. [sɪŋ] 'sing'), before word-level suffixes (e.g. [ˈsɪŋər] 'singer,' [ˈsɪŋɪŋ] 'singing'), or when followed by /k/ or /ɡ/ ([sɪŋk] 'sink,' [ˈfɪŋɡər] 'finger'). Similarly, /ð/ is a phoneme of English, but its distribution is heavily restricted, being found at the beginning of a handful of function words, mostly determiners ([ˈðɪs, ˈðæt, ˈðɪz, ˈðəz, ˈðəm, ˈðɪ, ˈðæi, ˈðə, ˈðən] 'this, that, these, those, them, thee, thy, the, then'), in a handful of words ending in [ðə] (['lðə, 'bʌðə, 'wɛðə, 'fɛðə, 'mʌðə, ˈbɪðə, ˈfæðə] 'other, bother, weather, feather, mother, brother, father'), and at the end of a handful of verbs, mostly denominal ([ˈbɪðə, ˈbɛθə, ˈʃeθə, ˈwɪðə] 'breathe, bathe, sheathe, writhe').

Conditioned variants of phonemes by definition also have restricted distributions. The distinction in Russian between voiced and voiceless obstruents (a class that includes the oral stops noted above) is found only before sonorants, and is otherwise neutralized in agreement with following obstruents (voiced before voiced, voiceless before voiceless) or to voiceless word-finally.

These restrictions on the distributions of phonemes and of their conditioned variants, and those on the recombination of distinctive features mentioned earlier, provide fodder for the kinds of abstract analyses that command particular attention in phonology, to which we now turn.

2.5 How abstract are mental representations?

The question posed in this section title mainly lurks under the surface of modern debates in phonological theory, such as the extent to which phonology can be reduced to physiological principles governing articulation and perception (Ohala, 1981, 1997;
Hale and Reiss, 2000; Hayes, Kirchner, and Steriade, 2004; Blevins, 2004; Heinz and Idsardi, 2013; Reiss, 2018). The question was asked more directly by Kiparsky (1968), as betrayed by this classic paper’s title: How abstract is phonology? The paper addresses concerns about early work in generative phonology, which admitted the possibility of many transformations applying in crucial sequence from underlying to surface representations such that the specifications of a given underlying representation could be quite a bit different from those of its eventual surface representation. The concerns were about the possibility of “excessive abstractness,” though different types of abstractness and how one is to measure them with respect to one another were often more matters of opinion than principle. As discussed further below, Kiparsky’s proposal was not so much a line drawn in the sand as it was the curtailment of one particular type of abstractness. His and subsequent theoretical proposals have placed limits on the possible types of differences that may hold between an underlying representation and its various surface manifestations, but they have not somehow placed limits on the distance between these representations.

The form of the evidence for a relatively abstract phonological analysis is typically one where what otherwise appears to be the same phoneme exhibits two distinctive forms of behavior: one expected (call this one A) and one unexpected (call this one B). The more strands of evidence of this sort that exist in any given analysis, the more compelling the case for it; examples from Yokuts (Kisseberth, 1969), Nupe (Hyman, 1970), and Dida (Kaye, 1980) are particularly compelling cases. The unexpected behavior of B can be approached in one of two ways, both being abstract in some sense. One approach is to posit that A and B are indeed phonemically identical, but that there is some abstract non-phonological diacritic marking X, present in forms with B but not in those with A, to which relevant phonological transformations are sensitive. (X may but need not have other functions in the language.) We call this the abstract diacritic approach.

Kiparsky refers to the other approach as “the diacritic use of phonological features,” and it is instantiated in one of two basic ways. One instantiation is to directly posit that B is a phoneme distinct from A in terms of some feature such that the combination of distinctive features describing B does not surface, either anywhere in the language or in the phonological contexts or lexical items in question. Relevant phonological transformations sensitive to the A/B distinction apply, and this distinction is subsequently neutralized to A. The other instantiation is to posit that A and B are phonemically identical but that there is some other phoneme C, co-present with B but not with A, to which relevant phonological transformations are sensitive. C is either subsequently deleted or neutralized with some other phoneme. We call both of these instantiations the abstract phoneme approach.

9 Or very similar in some respects, by way of a non-spurious there-and-back-again sequence of transformations known as a Duke of York derivation (see Pullum, 1976; McCarthy, 2003; Baković, 2013 for discussion).
Abstract phonemes can be either absolute or restricted. An *absolutely abstract phoneme*—or an “imaginary segment” (Crothers, 1971)—is a combination of distinctive feature values posited as a phoneme in some set of morphemes but that is not realized as such in the surface forms of any morphemes in the language. An example appears in an analysis of exceptions to vowel harmony in Hungarian (Vago, 1976); some stems with the neutral front vowels [i iː eː] unexpectedly condition backing of suffix vowels, motivating the postulation of abstract back vowel phonemes /ɯ ɯː ɤː/ in these stems that are eventually and uniformly fronted.

A *restrictedly abstract phoneme* (O’Hara, 2017) is a combination of distinctive feature values posited as a phoneme in a restricted set of contexts but that is not realized as such in any surface contexts corresponding to that set. An example appears in an analysis of exceptions to vowel coalescence in Mushunguli (Hout, 2017): some stems beginning in the high vowels [i u] unexpectedly block otherwise regular coalescence with preceding /a/, motivating the postulation of glide phonemes /j w/ that respectively never surface before [i u], respectively, but that do surface in other contexts. These glides block coalescence and are subsequently deleted just before [i u].

All abstract phoneme analyses crucially rely on the possibility of opaque interactions between phonological transformations (Kiparsky, 1973). In each of the examples sketched above, the transformation that voids the representation of its abstract phoneme crucially must not apply before application of the relevant transformation(s) sensitive to the abstract phoneme.10 In some cases, the abstract phoneme is needed in the representation to prevent the application of an otherwise applicable transformation, as in the case of Mushunguli. These involve the type of opaque interaction known as counterfeeding. In other cases, the abstract phoneme is needed in the representation to ensure the application of an otherwise inapplicable transformation, as in the case of Hungarian. These involve the type of opaque interaction known as counterbleeding. Abstractness is thus intimately intertwined with opacity: to the extent that a theoretical framework (dis)allows opaque interactions, it (dis)allows abstract phoneme analyses.

Kiparsky’s (1968) proposed principle was meant to ensure that every phoneme specified as a particular combination of distinctive features in a given morpheme’s underlying form will be realized with that precise set of specifications in at least one surface form of that morpheme.11 This effectively excludes abstract phoneme analyses such as those sketched above—but in each of these cases the alternative is to rely instead on diacritic marking, which is, as already noted, also a form of abstractness. More recently, O’Hara (2017) specifies a MaxEnt learning model that assigns sufficiently high probability to a restrictedly abstract phoneme analysis (based on a case in Klamath), and

---

10 We write ‘must not apply before’ as opposed to ‘must apply after’ because these opaque interactions can be had with simultaneous as opposed to ordered application of phonological transformations (Anderson, 1974; Kenstowicz and Kisseberth, 1979; Joshi and Kiparsky, 1979, 2006; Kiparsky, 2015). See also Baković and Blumenfeld (2017) for discussion.

11 We put aside the precise formulation of Kiparsky’s principle, as well as of its ensuing revisions (Mascaró, 1976; Kiparsky, 1982, 1993), focusing instead on its intended function.
significantly lower probability to both absolutely abstract phoneme and abstract diacritic alternatives. It is thus reasonably clear that a hypothetical abstract-o-meter (recall Figure 2.3) would place absolutely abstract phonemes closer to the “too abstract” end of the spectrum than restrictedly abstract phonemes, but where abstract diacritics fall on that spectrum is a largely unresolved matter. We must thus conclude here that “a phonological analysis, independently of its ‘degree’ of abstractness, is (only) as adequate as the motivation and evidence that can be produced in favor of it and against substantive alternatives” (Baković, 2009, p. 183).

This section and the last focused on the issue of phonological abstractness in the mental lexicon at the level of the phoneme and of the distinctive feature. This is primarily because most relevant research has been focused here, with the possible exception of the syllable (Goldsmith, 2011; Strother-Garcia, 2019). However, it has been argued that other representational levels and systems also play a role in the mental representations of words, notably tonal representations (Yip, 2002; Jardine, 2016) and metrical stress (Hayes, 1995; van der Hulst, 2013). These representations have also been the subject of intense study, including the psycholinguistic and neurolinguistic literature, to which we turn in the next section.

### 2.6 Types of evidence

The preceding sections discussed the perspective of phonological abstractness using structural linguistic arguments from the typology of spoken language. However, it is often claimed that structural arguments are far from a definitive proof, constituting one type of evidence. This is especially true for the psychological reality of these abstract forms. Ohala (1974) makes this point:

> It seems to me that the important question should not be whether phonology is abstract or concrete, but rather what evidence there is for the psychological reality of a particular posited underlying form. If our aim is simply a descriptive one, then of course abstract forms can be posited without any further justification than the fact that they make it easier to state certain regularities. However, if we are interested in reflecting the competence of a native speaker, then we must provide evidence that what we are claiming as part of a speaker’s knowledge is indeed such.

(Ohala, 1974, p. 234)

What other types of evidence are there? This section describes various other types of evidence bearing on the question of abstractness. The first type of evidence comes from expanding the typology to consider non-spoken language, that is, signed and tactile language. This evidence is still structural but provides an important window into the lexicon. How dependent is the long-term memory representation of a word on the physical system that externalizes it? At the same time, Ohala and many others take
behavioral and psychological tests as providing additional necessary, if not sufficient, evidence for the reality of these forms. Recent advances in neuroimaging and increasing ease of use allow for intricate looks into the biological underpinnings of phonological representations, combined with computational models and simulations. However, literature in all these topics is vast, especially with regard to experimental work. Here we provide a sample of work that bears on the questions described earlier.

2.6.1 Evidence from signed and tactile phonology

An important source of evidence for the existence and nature of phonological abstraction comes from languages without a spoken modality—namely, signed and tactile languages. Sign languages arise spontaneously in Deaf communities, are acquired during childhood through normal exposure without instruction, and exhibit all of the facets and complexity found in spoken languages (see Sandler and Lillo-Martin, 2006 for a groundbreaking overview). However, Sandler (1993) argues that if human language evolved without respect to modality, we should find hearing communities that just happen to use sign language rather than spoken language, and we do not. Sign language, she argues, is thus “an adaptation of existing physical and cognitive systems for the purpose of communication among people for whom the auditory channel is not available.”

Sign languages offer, as Sandler (1993) puts it, “a unique natural laboratory for testing theories of linguistic universals and of cognitive organization.” They give insight into the contents of phonological form, and conditions on which aspects of grammar are amodal and which are tied to the modality. They also offer unique opportunities to study the emergence of phonology within a developing grammar (Goldin-Meadow, 2005; Senghas, Kita, and Ozyürek, 2004; Marsaja, 2008; Sandler, Meir, Padden, and Aronoff, 2005).

One crucial contribution of non-spoken phonology is that it switches the issue of “how abstract is phonology?” to “where does abstraction lie, and to what extent is it independent of the modality?” There are generally two directions answers can take. To the extent that a given phonological form or constraint is present across modalities, one may make the case that it is truly abstract, in the sense that it exists without regard to the articulatory system which realizes it. On the other hand, to the extent that a given form or constraint differs, one can ascribe that difference to the modality, to the nature of the articulatory system. In this way, non-spoken phonology provides nuance into the relationship between the pressures of abstraction and the pressures of realization in the mental lexicon.

Sign languages, despite their rich history, were almost totally dismissed as natural and structured languages until the latter half of the 20th century (see van der Hulst, to appear for the history of sign phonology). Stokoe (1960) divided signs into meaningless chunks, showing that sign languages display both morpho-syntactic and phonological levels, a “duality of patterning” or “double articulation” considered previously as a unique property of spoken languages (Martinet, 1960; Hockett, 1960). Stokoe’s phonological system
specified abstract representations for parts of the sign: the handshape, the movement of
the hand, and the location in front of or on the body. van der Hulst notes that Stokoe’s
division of signed forms was designed for transcription, but he regarded these symbols
as abstract representations of the discrete units characterizing the signs.

Much work on the psychological reality of these phonological abstractions in sign
came from the Salk Institute (see Klima and Bellugi, 1979 for an accessible overview,
and Emmorey, 2001 for more recent developments). Studying production errors, they
demonstrated compositionality in both perception and production of sign forms. The
Bellugi and Klima group showed that cross-linguistically, signers make acceptability
judgments about what they consider well-formed or ill-formed, evidence that they
possess intrinsic knowledge of how these smaller units can be combined. As van der
Hulst (to appear) notes, Klima and Bellugi’s (1979) accessibility and cognitive scope
convinced many linguists and non-linguists of the importance of sign language, and
sign linguistics as a proper study in ways Stokoe’s analysis was unable to.

These works galvanized the phonological study of signs, increasingly focused on sequen-
tial aspects of signs and away from simultaneous aspects, which Stokoe had emphasized
Researchers discovered, among other evidence, the salience of the beginning and end points
of the movement of signs for inflectional purposes where referents are marked by discrete
spatial locations, as well as phonological processes like metathesis, a switch in the beginning
and end point of the movement (see van der Hulst and van der Kooij, 2021 for discussion).

The notion of both simultaneous and compositional structure in a sign, as well as se-
quential structure, raises a big question: how modality-dependent are these properties?
Languages in both modalities have sequential and simultaneous structure but exhibit
differences in relative centrality of such structure. Spoken languages vary in syllable
structure, word length, and stress patterns among syllables. Sign languages appear
limited in all these aspects. They are overwhelmingly monosyllabic, have no clusters,
and show extremely simple stress patterns, due to few polysyllabic words apart from
fully reduplicated forms (see Sandler and Lillo-Martin, 2006 for general discussion, and
Wilbur, 2011 for discussion of signed syllables).

A further complication arises from the fact that sequential phonological struc-
ture in sign language appears mostly from morphosyntactic operations concatenating
morphemes and words (e.g. affixation, compounding, and cliticization; Aronoff,
Meir, and Sandler, 2005). In general, sequential affixation is rare across sign languages
(Aronoff et al., 2005), and sign exhibits a strong tendency to express concatenative
morphology through compounding (Meir, 2012). Aronoff et al. (2005) show that affix-
ation usually emerges from the grammaticalization of free words, through a series of
diachronic changes affecting both phonological and semantic factors. They cite the rela-
tive youth of sign languages as a major factor in their lack of affixes. No known sign
languages are over 300 years old, with some like Nicaraguan Sign Language as young as
40 (Woll, Sutton-Spence, and Elton, 2001), and many others in development.

The curious lack of sequential structure in sign languages does not imply struc-
tural degeneracy or simplicity, however. Sign languages routinely demonstrate
nonconcatenative morphology (Sandler, 1989; Meier, 2002), incorporating morphological material simultaneously in the phonology alongside the restricted sequential form. Simultaneous phonological structure exists in all languages but differs across modalities in the amount. For example, while the simultaneous “autosegmental” representations for tone or harmony patterns (Goldsmith, 1976) typically consist of one or two features, the autosegmental representation of hand configuration alone in sign language contains around half of the distinctive features comprising a sign organized in an intricate feature geometry (van der Hulst, 1995; Sandler, 1996). Such tradeoffs in sequential and simultaneous centrality have been argued to stem from a computational restriction that may be realized via different representations in different modalities (Rawski, 2017).

This converging line of evidence suggests that the phonological grammar may leverage the representational abilities of the particular articulatory/perceptual system. Brentari (2002) and Emmorey (2001) argue that visual perception of signs (even with sequential properties) is more “instantaneous” than auditory speech perception. This leads van der Hulst and van der Kooij (2021) to adapt Goldsmith’s (1976) division of phonology in terms of the notions of “vertical and horizontal slicing of the signal.” They state:

an incoming speech signal is first spliced into vertical slices, which gives rise to a linear sequence of segments. Horizontal slicing then partitions segments into co-temporal feature classes and features. In the perception of sign language, however, the horizontal slicing takes precedence, which gives rise to the simultaneous class nodes that we call handshape, movement, and place. Then, a subsequent vertical slicing of each of these can give rise to a linear organization.

Perhaps the most intriguing evidence for or against abstractness comes from studying the degree to which phonetics and phonology differ across modalities. Lillo-Martin (1997) cites Blakemore’s (1974) result that exposure to vertical and horizontal lines in the environment affects development of feline visual perception, and asks “why shouldn’t exposure to the special acoustic properties of the modality affect perception, especially auditory perception?” Sandler and Lillo-Martin (2006) note, for example, that unlike spoken syllables in many languages, sign syllables prohibit location clusters analogous to consonant clusters, as well as diphthong-like movement clusters, and sign physiology requires movement between locations. They additionally note that sign syllables do not have onset-rhyme asymmetries, which affects syllable structure and stress assignment. Many more such differences have been studied, and further work in this area will bring important issues to bear on the nature of abstract representations in and across articulatory systems.

The similarities and differences in phonological abstraction across modalities means that signed languages continue to play an important role as evidence of abstraction in the lexicon. This holds equally true for language expressed by the DeafBlind through the tactile modality, often called tactile or pro-tactile sign languages (see Edwards, 2014 for a recent phonological analysis).
2.6.2 Psycholinguistic and neurolinguistic evidence

As mentioned, behavioral testing has long been argued as necessary evidence for the mental reality of phonological abstraction, in addition to typological evidence. The introduction and improvement of neuroimaging methods enabled correlations between behavioral tasks and gross neural excitation levels associated with them. In addition, recent simulation tools allow for modeling of phonological representations in an idealized \textit{in silico} setting. Here we overview several results via behavioral and neural methods bearing on phonological organization of the lexicon by its salient features into abstract phonemes, as well as work on the temporal abstraction of speech.

One salient question concerns experimental evidence for abstract phonemic and featural representations of words. For example, in Russian, the sounds [d] and [t], which featurally differ in voicing, are contrastive members of different phonemes. In Korean, these sounds do not contrast and are members of a single phoneme. Kazanina, Phillips, and Idsardi (2006) used the neuroimaging method of magnetoencephalography (MEG), which tracks the timecourse of gross, large-scale neural activity to show that Russian and Korean speakers react differently to these sounds. The Russian speakers separated the sounds into two categories corresponding to /t/ and /d/. On the other hand, the Korean speakers did not separate the sounds, again corresponding to the analysis of a single underlying phoneme. From this result, the authors conclude that both phonetic and abstract phonemic analyses necessarily shape the perceptual analysis of speech sounds.

There is much evidence supporting a vast neuronal ensemble for phonological representations in speech production and perception (see Eickhoff, Heim, Zilles, and Amunts, 2009; Jueptner and Krukenberg, 2001 for an extensive overview). In particular, various portions of the superior temporal sulcus are suggested to encode the phonological representations discussed in this chapter. Scharinger, Isdardi, and Poe (2011) used a combination of MEG imaging and statistical modeling to map the entire vowel space of a language (Turkish) onto three-dimensional cortical space, organized by lateral-medial, anterior-posterior, and inferior-superior axes. Their statistical model comparisons showed that, while cortical vowel maps do reflect acoustic properties of the speech signal, articulator-based and featural speech sound information “warp the acoustic space toward linguistically relevant categories.”

Scharinger, Monahan, and Isdardi (2012) used MEG to localize three vowel feature variables (height, frontness and roundness) to the superior temporal gyrus. Mesgarani, Cheung, Johnson, and Chang (2014) used cortical electrode placement to show that the superior temporal sulcus encodes a “manner of articulation” parameter of speech sounds. Intriguingly, different electrodes responded selectively to stops, sibilant fricatives, low back vowels, high front vowels and a palatal glide, and nasals, respectively. Bouchard, Mesgarani, Johnson, and Chang (2013) showed similar results, that the superior temporal gyrus encodes a “place of articulation” parameter, confirming labial, coronal, and dorsal place features across various manner classifications. These results match with Hickok and Poeppel’s (2007) hypothesis that “the crucial portion of the STS
that is involved in phonological-level processes is bounded anteriorly by the most anterolateral aspect of Heschl’s gyrus and posteriorly by the posterior-most extent of the Sylvian fissure.”

Additionally, recent experimental evidence using aphasic patients even supports the existence of abstract phonological rules in processing. Linguistic analysis posits that the English words *pit* and *spit* both contain the segment /p/ in their underlying representations. The surface representation of *pit* has aspirated [pʰ], because the /p/ is in word-initial position, while the surface representation of *spit* has unaspirated [p], because it is preceded by /s/. Buchwald and Miozzo (2011) constructed an experiment using the productions of two aphasic patients who were unable to produce an /s/ in relevant consonant clusters like /sp/ or /st/, and compared them with correctly produced consonants. They wanted to test whether an aphasic would aspirate the /p/ (marking phonological fricative-deletion) or not (the fricative deleted after successful application of the phonological rule).

To analyze it, they compared the voice onset time (VOT) of the two patients on the two instances. VOT provides an acoustic measure of the relative aspiration of the consonant by seeing how much the following voicing is delayed. One patient had a long VOT ([pʰ]) while the other had a short VOT ([p]), confirming the divide between two distinct levels of phonological and phonetic influences in processing. Follow-up work by Buchwald and Miozzo (2012) showed similar results for nasal consonants that deleted in clusters like /sn/ and /sm/. The conclusion drawn from these studies points to abstraction in the units being processed, mentally divorced from their phonetic realization but ultimately driving it.

Apart from the biological underpinnings of the atomic representations characterizing speech units, there is also much work focused on the biological underpinnings of temporal abstractions in speech. Specifically, much of this work focuses on the insula, basal ganglia, and cerebellum, where temporal information is speculated to be filtered in a cortical-subcortical loop for the purposes of motor planning (Eickhoff et al., 2009). In this process, motor sequence plans are filtered through basal ganglia, while the cerebellum converts the sequences into fluent, temporally distributed articulations. Ackermann, Mathiak, and Riecker (2007) underpin this by describing drastic negative effects on speech production consistent with damage to the cerebellum.

While neuroimaging methods brought many advantages, they simultaneously introduced a uniquely thorny issue to phonological abstraction, which Poeppel (2012) divides into the “Maps Problem” and the “Mapping Problem.” The Maps Problem concerns descriptive analysis of the behavioral and neural underpinnings of mental representations, say, the effects and brain areas associated with a particular phonological phenomenon. The Mapping Problem concerns how to take a particular mental representation, perhaps known to correlate with some behavioral entity or brain network, and mechanistically connect it to neuronal function. Neither the Maps Problem nor the Mapping Problem have easy answers (Buzsaki, 2019), and decades of work have led to many open and nuanced questions.

Attempting to address the Mapping Problem, some work seeks a somewhat more explanatory approach to the forces underlying the temporal segmentation of the signal
produced and perceived. One emerging insight is that the perceptual system divides an incoming auditory stream into two distinct time windows (Giraud and Poeppel, 2012; Chait, Greenberg, Arai, Simon, and Poeppel, 2015). What phonological entities do they map to? As Poeppel and Idsardi (2011) put it, there are two critically important windows that appear instantiated in spoken languages: segments and syllables. Temporal coordination of distinctive features overlapping for relatively brief amounts of time (10–80 ms) comprise segments; longer coordinated movements (100–500 ms) constitute syllabic prosodies.

(Poeppel and Idsardi, 2011, p. 182)

A more fundamental question concerns the neural mechanism that drives these windows and their coordination. Oscillatory activity within and between neural populations has been posited (Giraud and Poeppel, 2012). Neural populations that comprise a certain type of neuron may show a stable neural oscillation at certain frequency bands, which varies depending on their excitatory and inhibitory properties (see Buzsaki, 2006 for an accessible overview). Evidence suggests that pyramidal interneuron gamma oscillations, as well as theta oscillations, comprise the segmental vs. syllabic time distinction. These two oscillations funnel an incoming speech signal into time windows of different sizes, computationally represented by the waveform of the population activity.

In silico modeling work reveals another interesting property. While these oscillations do track the signal, a crucial feature is that rhythms of distinct frequencies show specific coupling properties, termed cross-frequency coupling (Hyafil, Giraud, Fontolan, and Gutkin, 2015b). Briefly, this means that the stable oscillatory populations innervate each other, allowing different timescales to track one another, effectively parsing a temporally complex signal efficiently. Specifically for speech perception, Hyafil, Fontolan, Kabdebon, Gutkin, and Giraud (2015a) showed that when a network exhibiting gamma oscillations coupled to a network showing theta oscillations, it was able to effectively segment a corpus of phonological words much better than a network where this coupling was absent.

These results reflect an explosion of work using neurolinguistic and psycholinguistic tests to describe the sorts of representations speakers have. The intersection of experimental results with theory promises many new insights into the mental content of the lexicon (see Poeppel and Sun, this volume). For a further discussion on the particular biological substrate underlying phonological abstraction and how they impact the phonetics-phonology interface see Volenec and Reiss (2017).

2.7 Conclusion

In this chapter, we have endeavored to motivate and review a central idea of modern generative phonology that the fundamental representational units of words in languages are abstract and psychologically real. In particular, the systematic patterning
of the pronunciation of morphemes motivates an abstract mental representation of a morpheme’s pronunciation. These underlying forms are largely regarded as sequences of phonemes, which are themselves abstractions, and which are organized along featural dimensions. These abstract mental representations find support not only from the patterning in morpho-phonological paradigms but also from language change, from sign language linguistics, and from psycholinguistic and neurolinguistic study.

There are many open and fascinating questions regarding the nature of abstract mental representations of words. Many are among the most basic and fundamental. How abstract can they be? How are they learned? How are they realized in the brain? The fact that we simultaneously know both so much and so little about phonological abstractness in the mental lexicon sends a clear message that this will continue to be a fertile and exciting area of research for many years to come.

To conclude this chapter, we can do no better than to repeat the concluding sentences of Labov (2020, p. 57, emphasis added): “We have a common ground in our understanding of what it means to know a language. It involves knowing a vast number of particular things. But at bottom it is reaching down to something very deep, very abstract, and very satisfying.”

Acknowledgments

We thank Aniello De Santo, Charles Reiss, Veno Volenec, and two anonymous reviewers for valuable feedback on earlier drafts. And we dedicate this chapter to Lila Gleitman, whose pioneering work on linguistic abstraction, especially for language learning, has had a major impact on our thinking.
3.1 What is phonological variation?

The mental lexicon is where we store our knowledge of the words in our language. A reasonable starting point is to think of entries in the mental lexicon as form–meaning pairs. The lexical entry for the word *cat*, for example, pairs (a) some semantic information about a household pet that meows (**meaning**), with (b) some information about the speech sounds used to refer to it (**form**). As a first pass, we could say that this form information is stored as a string of **phonemes**: /k æ t/. The form side of this pairing is what allows speakers to externalize meaningful messages to the people around them and allows listeners to retrieve the intended meanings in turn. But a string of phonemes is, of course, an abstraction: what comes out of a speaker’s mouth is sound waves shaped by articulatory gestures, and what a listener encounters is a continuous and complex acoustic signal. The physiological demands of this **phonetic implementation** mean that no two instances of a word in real speech are ever exactly the same, an observation known as the lack of invariance problem. Both speakers and listeners face the challenge of connecting a word’s abstract lexically stored form with the continuous and multidimensional space of the phonetic implementation.

Often, however, words surface with multiple forms in ways that cannot be explained by the physiological demands of speech production. In the basic case, accounting for these forms is the domain of **phonology**. For example, words sometimes appear to change form when they are combined with certain suffixes. Consider the English word *confess*, which can be combined with the suffix *-ion* to make *confession*. The stem *confess*

---

1 Phonemes are the distinctive sound units of a language; see Bacović et al. (this volume). These symbols are from the International Phonetic Alphabet, a system of representing speech sounds in writing.
shows up in different forms: *confession* has a palatal [ʃ] where *confess* has an alveolar [s]. Do English speakers store both forms in their lexicon and know to choose the [ʃ] form with the -ion suffix? Baković et al. (this volume) lay out the standard arguments that many linguists give for saying no, and instead analyzing the [ʃ] as an allophone (a predictable pronunciation alternant) derived by phonological rule rather than stored in the lexicon. A single rule can capture the generalization that [s] becomes [ʃ] in other stems that combine with -ion (e.g., express/expression, compress/compression, and so on). On this view, phonological rules intervene between the lexicon and the phonetic implementation, editing the target segments that go on to be articulated in speech. This is not the only available model of the relationship between the lexicon, the phonology, and the phonetics, but because it is a widely accepted framework, we will build our discussion around it.

This chapter is about phonological variation. But the aforementioned phonological alternation between [ʃ] and [s] is not generally referred to as phonological variation because the rule is obligatory whenever the linguistic conditions that trigger it arise. Thus, the [ʃ] allophone is fully predictable from the linguistic environment. Pronouncing *confession* as [kənfɛʃən] (without having applied the rule) is simply not a well-formed option for English speakers. What, then, is phonological variation? First, while many phonological rules are obligatory, there are also cases where seemingly similar alternations are not fully predictable. To continue our current example, the phrase *impress you* can be pronounced as either [ɪmprɛʃju] or [ɪmprɛʃju] in connected speech. The two forms look suspiciously like the input and output of our [ʃ]-deriving phonological rule, but in this case the speaker has a choice between the options. To give a more intuitively familiar example, words ending in unstressed /ɪŋ/ can optionally be pronounced with [ɪn]: *morning*~*mornin’*, *pudding*~*puddin’*, *jumping*~*jumpin’*, *hypothesizing*~*hypot- hesizin’*, and so on. The same logic that led us to posit a general rule capturing the obligatory pattern of [s] alternating with [ʃ] might lead us to conclude that this optional variability in [s]~[ʃ] and [ŋ]~[n] is also the product of phonological rules—just not obligatory ones. This intraspeaker variation, where a given speaker may say the same thing in different ways, is one kind of phonological variation that we will cover in this chapter. The choice a speaker has between the different options is called a variable, and the options themselves are called variants. Phonological variables are often influenced by social and situational factors; for example, most English speakers will share the intuition that *mornin’* is a more casual way of saying *morning*. However, quantitative

2 More precisely, a rule that palatalizes coronals before /j/-initial suffixes.
3 A class of prominent alternatives is usage-based approaches to phonology, such as Exemplar Theory, in which episodic traces or “exemplars,” prototypically word-level exemplars, are stored in memory and form the basis for the emergence of phonological categories or generalizations. The possibility of “hybrid” abstractionist/episodic frameworks has attracted increasing attention in recent years. See Pierrehumbert (2002), Pisoni and Levi (2007), and Hay (2018) for overviews.
4 Even though colloquially the words “vary” and “alternate” seem to mean approximately the same thing.
5 We use the ~ notation to mean “varies with.”
sociolinguistic research supports the premise of inherent variability (Weinreich, Labov, and Herzog, 1968): that variant choice is not fully predictable, even with a hypothetically exhaustive understanding of an utterance’s social context.

In addition to intraspeaker phonological variation, phonology can also differ across speakers, even when they are nominally speaking the same language. We refer to these differences as interspeaker variation. Speakers can differ from each other in many ways, but we will give special attention to interspeaker variation that involves phonological structure and lexical form. For example, two speakers may have different phonological rules in their grammars, different stored forms in their lexicon, or different phonemic inventories (the set of distinctive sounds in a language). One familiar source of interspeaker variation that we discuss at some length in Section 3.1 is regional dialects. Interspeaker variation may also reflect other aspects of a language user’s background, such as gender, class, or race. From the point of view of language production, one might think that interspeaker variation need not be treated as variation at all: an American English speaker is probably never going to entertain the option of pronouncing the word got with a retroflex consonant, [ɡɑʈ], the way an Indian English speaker might. But from the perspective of language comprehension, differences across speakers are a major contributor to the phonological variation in the input that listeners must accommodate.

Both intra- and interspeaker phonological variation can create a range of mismatches between real utterances and the stored forms in the lexicon. These mismatches pose a substantial challenge for lexical processing, including processes of word recognition (see Magnuson and Crinnion, this volume), word production (see Kilbourn-Ceron and Goldrick, this volume), and word learning (see Creel, this volume). In addition to its processing consequences, phonological variation raises new questions of representation. We have already pointed out that there are some parallels between obligatory phonological rules and intraspeaker phonological variables, but it does not necessarily follow that a non-obligatory phonological rule is the right analysis for any given variable; in some cases, there might be reason to believe that the options are stored in the lexicon, or arise in the phonetic implementation. Given the complexity of the challenges posed by phonological variation, it is unsurprising that models of the mental lexicon have largely set aside variable phenomena and have instead developed on the basis of invariant forms of words in isolation. However, as many other authors in this volume point out, variation is one of the major hurdles standing in the way of modeling the comprehension and production of words in their real-world context of continuous speech. As psycholinguists set their sights on increasingly realistic and dynamic models of how the mental lexicon is structured and used, it will become correspondingly worthwhile to tackle issues of phonological variation.

6 This excludes speakers who command both varieties and may code-switch between them, as well as perhaps a narrow set of circumstances that we think can reasonably be set aside as marginal, such as deliberately performing a different accent.
3.1.1 The scope and aims of this chapter

We have already touched on a number of important themes, each of which alone could each easily fill a chapter on phonological variation: lexicon vs. phonology vs. phonetics; intraspeaker vs. interspeaker variation; comprehension vs. production; processing vs. representation. These dimensions could also be crossed with each other to form a very large space of interacting topics: the processing of intra- vs. interspeaker variation, the mental representation of intra- vs. interspeaker variation, and so on. An exhaustive treatment of this space is clearly beyond what could be accomplished in a single chapter. We therefore pursue a smaller set of more narrowly directed aims. Our overarching goal is to make explicit the relevance of phonological variation to the study of the mental lexicon. To do so, we connect the linguistic properties of phonological variation with their observed or potential consequences for lexical representation and, especially, processing.

First, in Section 3.2, we review experimental work showing that phonological variation comes into play during lexical access, when language users “look up” words in their mental dictionaries for use in production or comprehension. This literature, which focuses mostly on spoken word recognition, demonstrates that variation sometimes but not always interferes with word recognition, that listeners are able to rapidly accommodate even unfamiliar variation, and that listeners may in fact use phonological variation as extra information to guide lexical access. While these results show that there is some relationship between phonological variation and lexical access, they leave many questions about the nature of this relationship unanswered. We suggest that the path forward should take the linguistic properties of different types of phonological variation into account. Different variables may be represented differently, and have different structural consequences, as we outline in Section 3.3. In Section 3.4, we elaborate on how such differences have the potential to impact lexical processing and conclude by sharing our optimism about the advantages that incorporating phonological variation may offer for models of the mental lexicon.

3.2 How does phonological variation affect lexical access?

Although we have noted that psycholinguistic models of the mental lexicon have largely developed on the basis of isolated citation-form words, there are a number of lines of experimental work about how different pronunciations of words might impact lexical access. The evidence from this body of work provides a number of insights. Phonological variation does not necessarily disrupt lexical access; in fact, listeners are quite tolerant of licit variation in isolated words, and can flexibly adapt to novel accents characterized by many co-occurring phonological variables. Further, there is emerging evidence that
listeners can use phonological variation and their knowledge of a speaker’s accent to facilitate spoken word recognition, suggesting that understanding phonological variation may ultimately help us understand lexical access processes.

3.2.1 Variation need not impede word recognition

Variation creates mismatches between the form in a listener’s lexicon and the phonetic form they perceive. Canonical forms are careful (or even hyperarticulated) pronunciations perceived as matching the “dictionary” pronunciation, while non-canonical forms diverge from that ideal in some respect. A reasonable hypothesis following from this point is that a non-canonical pronunciation might delay how quickly a listener can access a word’s lexical entry if the input diverges from the corresponding stored lexical form, whereas canonical forms might diverge less and thus be easier to access. An early example of a study supporting the hypothesis of a canonicality advantage in processing is Andruski, Blumstein, and Burton (1994). This study manipulated the voice onset time (VOT) of voiceless initial consonants in English, which tends to be relatively long in isolated words compared to connected speech. In a semantic priming task, where prior presentation of a semantically related prime word speeds recognition of a target word, primes with longer (i.e., more canonical) VOTs generated more priming than primes with short VOTs. Interestingly, this advantage only arose when the time between the prime and target was very short, pointing to a role for variation in the early stages of spoken word recognition. LoCasto and Connine (2002) found a similar advantage for words like camera with a canonical pronunciation vs. a non-canonical reduced schwa; Racine and Grosjean (2000, 2005) and Racine, Bürki, and Spinelli (2014) reported an advantage for canonical word forms in French with a schwa in the first syllable (e.g., genou = knee). Tucker and Warner (2007) found a small facilitation effect for words pronounced with a canonical word-medial /d/ or /g/ in comparison to reduced forms of those consonants.

Conversely, a number of studies using similar methods with other phonological variables have not found support for a canonicality advantage, instead finding apparently equivalent facilitation for canonical and non-canonical forms. One such variable is place assimilation, where a word-final consonant adopts the place of articulation of the following segment. For example, a word ending in a coronal stop /d/ such as wicked can (but need not) be pronounced with a labial stop [b] when the following word is labial-initial, as in [wikib ɹæŋk] for wicked prank. Gaskell and Marslen-Wilson (1996) show that non-canonical place-assimilated pronunciations yield just as much priming as the non-assimilated forms, contra the predictions of the canonicality advantage. Similar effects in which non-canonical pronunciations did not impede word

---

7 Ultimately, a pronunciation’s “canonicality” is a social construct; see Section 3.4.1 for additional discussion.
recognition have been found for nasal flapping (Ranbom and Connine, 2007; Pitt, Dilley, and Tat, 2011; Sumner, 2013, but cf. Pitt, 2009), voicing assimilation (Snoeren, Segui, and Hallè, 2008), and final /t/ allophony (Deelman and Connine, 2001; Sumner and Samuel, 2005).

Even studies failing to support the canonicality advantage, though, have generally found that listeners’ tolerance for non-canonical pronunciations is not unbounded. Gaskell and Marslen-Wilson (1996) argued that their place assimilation results did not just reflect listeners’ tolerance for mismatch: when the [wɪkɪb] pronunciation of wicked was followed by game, where [b] could not have been generated through place assimilation, it did inhibit lexical access (cf. Gow, 2001). In a slightly different vein, Sumner and Samuel (2005) found that various /t/ allophones produced naturally by speakers were accessed equally by listeners, but there was no priming from a minimally contrastive nonword prime (i.e., [flus] compared to non-canonical forms of flute). These results suggest that successful recognition of non-canonical forms depends on the variants being (a) possible pronunciations that are (b) licensed by context. In other words, the variation needs to represent surface patterns that should be familiar to listeners from their real-world listening experiences.

There is some evidence that the effect of context in facilitating lexical access of different pronunciations may be gradient. Tucker (2011), while finding a general advantage for canonical pronunciations, also found that the predictability of a non-canonical pronunciation in context improved its acceptability and speeded response times. This is supported in the domain of production by findings that French speakers produce words with non-canonical schwa omission faster as the relative frequency (i.e., predictability) of this form, compared to the canonical form, goes up (Bürki, Ernestus, and Frauenfelder, 2010). Another aspect of context that appears to modulate the recognition of non-canonical forms is speech style. In a study probing the conflicting results on whether nasal flapping (e.g., [splɪntə]~[splɪnə] for splinter) exhibits a canonicality advantage, Sumner (2013) observed equivalent priming from naturally produced canonical and non-canonical primes, but no priming from a non-canonical variant spliced into a carefully articulated word frame. In the latter case, the non-canonical variant was not licensed by the context of other acoustic cues to speech style within the word; Sumner proposes that careful and casual speech styles induce different processing modes, which may not be efficient for dealing with variants that are incongruous with the style. However, Bürki, Viebahn, Racine, Mabut, and Spinelli (2018) found that lexical decision latencies for words with and without schwa-reduction were equivalent across careful and casual speech styles.

The full set of results in this literature are challenging to reconcile completely, but that is to be expected for an area of such active inquiry. What we can take away at present is that hearing a word pronounced in a non-canonical way does not necessarily disrupt recognition of that word, and that listeners’ ability to reconstruct or predict the variants from the surrounding context may facilitate the processing of phonological variation. However, this ability is contingent on the variation being consistent with listeners’ social, stylistic, and linguistic experiences in the real world.
3.2.2 Listeners adapt even to pervasive and unfamiliar variation

The mixed support for the canonicality advantage hypothesis suggests that, on the whole, listeners are quite good at coping with variation so long as it is limited and familiar. But we might further inquire how variation influences word recognition processes when listeners encounter many different variable features at once, or when those features are not part of a listener’s own production repertoire. Both of these situations are a consequence of interspeaker variation, particularly between regional varieties of a language. These regional sub-varieties of a language are commonly called dialects, and we can refer to the full set of a dialect’s pronunciation features as an accent. Of course, the term “accent” can also be used to talk about the pronunciation of second-language speakers of a language. The notion of accent variation is intuitively familiar to listeners, who generally think of them in holistic terms; for example, a speaker might be said to have “a Southern accent” or “a French accent” even though the listener is unlikely to identify the cluster of features that give rise to that percept.

Naïve listeners have shown in perceptual categorization experiments that they are broadly able to identify where speakers are from, and that they can use specific acoustic-phonetic properties of talkers’ speech to make their judgments (van Bezooijen and Gooskens, 1999; Clopper and Pisoni, 2004, 2006, 2007). This skill is not limited to adult listeners. Children as young as 12 months are sensitive to dialect differences while listening to speech (Schmale, Cristià, Seidl, and Johnson, 2010). In fact, by four years of age listeners are able to group similar talkers together and distinguish them from dissimilar talkers, with further major developmental improvements in classifying talkers by region happening between the ages of 7 to 11 (Jones, Yan, Wagner, and Clopper, 2017; Evans and Lourido, 2019).

A number of studies have shown that listeners have highly flexible word recognition processes that allow them to quickly adapt to both regionally accented (Best, Shaw, and Clancy, 2013; Maye, Aslin, and Tanenhaus, 2008, i.a.) and foreign accented speech (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Witteman, Weber, and McQueen, 2013, 2014; Vaughn, 2019; Imai, Flege, and Walley, 2003; Bent and Frush Holt, 2013). Specifically, the evidence suggests that listeners experience a temporary disturbance when encountering a new accent, but this is normalized at an early stage of processing and improves over time (Floccia, Goslin, Girard, and Konopczyński, 2006; Goslin, Duffy, and Floccia, 2012). Adaptation is even possible for laboratory-created accents, so long as listeners are given enough exposure to them (Weatherholtz, 2015). Further, listeners’ expectations about what accent they are going to hear from a novel talker can impact how successfully they adapt (Vaughn, 2019).

3.2.3 Listeners may use phonological variation to guide lexical access

If phonological variants do not dramatically inhibit word recognition and listeners can quickly come to understand different accents (even if we are not sure how they do it),
does the study of the mental lexicon really need to deal with phonological variation? The conclusion that real-world, appropriately contextualized variation never derails lexical access is probably premature; in Section 3.4 we will discuss cases where we think such issues are quite plausible. However, the canonicality advantage hypothesis does not capture the only possible way in which phonological variation might be relevant to lexical access. In fact, in this section we turn to the mounting experimental evidence that phonological variation can actually provide information to listeners that may help guide their lexical access processes.

Many of the studies in Section 3.2.1 not only fail to show negative consequences for non-canonical pronunciations but also that listeners actually use phonological variants to anticipate upcoming words (Bürki, 2018; Gow, 2001, 2002; Lahiri and Marslen-Wilson, 1991; Tucker, 2011, i.a.). In an eye-tracking study, Mitterer and McQueen (2009) presented Dutch listeners with words in sentential context, and with a deleted word-final /t/ to make them ambiguous (e.g., *tast* = ‘touch’ sounds like *tas* = ‘bag’). Listeners used probabilistic knowledge of the likelihood of /t/ deletion in different following contexts to anticipate which image to look toward.

Besides cues from intraspeaker variation in the input, listeners can also use knowledge of a speaker’s background to guide lexical access. Listeners presented with a word like *pants*, which has a different dominant meaning in the United Kingdom (*pants* = undergarment) vs. the United States (*pants* = trousers), used the accent of the speaker to facilitate semantic access to the congruent meaning of the word (Cai, Gilbert, Davis et al., 2017). Accents can also help listeners decode non-canonical pronunciations in isolated words. American listeners hearing /r/-final words pronounced either with consonantal /r/ in a General American accent (e.g., *slender* as [slɛndə]), or with vocalized /r/ in either a British or New York City accent (e.g., [slɛndɔ]) were able to use their knowledge of the British English accent to support their understanding of /r/-vocalized words in isolation in a priming task. However, /r/-vocalized words were much harder to recognize in the New York City accent (Sumner and Kataoka, 2013). Sumner, Kim, King, and McGowan (2014) argue that this is because certain language varieties are more salient or idealized than others, which facilitates lexical access.

3.2.4 Dimensions of phonological variation in lexical access

The body of work discussed in this section makes a strong start at examining the ways in which phonological variation affects lexical access. We suggest that a useful next step in understanding how variation is represented and processed could be to consider the ways in which phonological variation is not monolithic. One possible reason that the experimental studies discussed in this section have not reached a consensus on these issues is that they manipulate different phonological variables, which in turn have different representations and may interact differently with lexical access processes. While the
canoncality advantage literature has acknowledged and explored the point that gra-
dient phonetic variation—such as the VOT manipulation of Andruski et al. (1994)—is
probably different in some respect from variation between different phonemes—such as
the variation between /d/ and /b/ in Gaskell and Marslen-Wilson (1996)—there is much
more that could be said about the dimensions along which we could classify different
kinds of phonological variation. In the next part of this chapter, we provide a non-
exhaustive overview of some such dimensions, with an eye to how incorporating a more
linguistically complex understanding of phonological variation might facilitate psycho-
linguistic research on the mental lexicon.

3.3 The complexity of phonological variation

We have just proposed that in the study of the mental lexicon, our ability to surmount
the difficulties of phonological variation may be dependent on a detailed empirical
understanding of the variability itself. In this section, we therefore outline the ways
phonological systems can differ between speakers, as well as how the types of variation
produced by individual speakers can be classified based on their representation. We pri-
marily draw our examples from varieties of English due to the abundance of work on
variation in Englishes and the fact that the majority of readers will have some frame of
reference for these varieties. However, the typology of phonological phenomena that
we describe is relevant to all language varieties. Taking representational differences into
account can better position us to make realistic predictions about how phonological
variation impacts lexical access. We begin by focusing on the linguistic descriptions in
this section, and then expand on their possible processing consequences in Section 3.4.

3.3.1 Interspeaker variation

Different people speak differently, even when they are speaking what is ostensibly the
same language. Many of the studies mentioned in Section 3.2 explored how different
accents are processed. However, research of this type has not generally asked about the
specific linguistic and (with some exceptions) social properties of accents. Here we cover
four elements that characterize accents and can differ between them: representations for
specific words, the phonetic realization of phonemes, the structure of the phonemic in-
ventory, and allophonic processes.

In some cases, interspeaker phonological variation is lexically specific. A person from
New York City and a person from London could both say the English word tomato, but
the New Yorker is likely to say something like [tɔˈmɛːroʊ], while the Londoner is likely to
say something like [tɔˈmoʊtou]. Both are speaking English, and both are using the same
word in reference to the same (prototypically) edible red fruit, but the pronunciations are different: in most varieties of American English, the second syllable of tomato has a vowel like that in the word face,⁸ but in most British Engishes, the vowel in this syllable matches the one in palm. This is not a phonological difference that generalizes to other words; the British English pronunciation of mate is not [mɑːt]. Therefore, the same word must be represented with different phonemes depending on the variety. Lexically specific phonological differences can be found at a smaller scale, too. So far, we have juxtaposed accents such as “American English” vs. “British English.” However, the Atlas of North American English (Labov, Ash, and Boberg, 2006) actually identifies seven major dialect regions across the United States and Canada (and British English may be even more internally diverse). Similar distinctions can be made between many of these accents for certain words; for example, for speakers in the northern half of New Jersey, the preposition on contains the lot vowel while in southern New Jersey it has the thought vowel (Coye, 2009).

In addition to some lexically specific differences, a common difference between accents concerns the across-the-board phonetic realization of phonemes. In fact, the primary criteria used in the Atlas of North American English (Labov et al., 2006) to draw boundaries between dialect regions concern differences in vowel quality. For example, speakers from the eastern Great Lakes region in the north of the United States produce lot with a fronted vowel compared to speakers from elsewhere. A fronted lot is one that has, over time, developed into a vowel more like the one that most other American English varieties use in trap. In this lot-fronting variety, trap is also shifted from an earlier position, so that lot does not overlap with trap and the two vowels are still distinct. When two vowel shifts seem to push or pull each other along in the same direction like this, it is called a chain shift. Chain shifts can result in dramatic differences in the realization of phonemes while keeping all the categories intact. The two chain links we have described here (lot fronting and trap tensing) are part of a larger chain called the Northern Cities Shift, which characterizes the accent of the Inland North dialect region (see Figure 3.1). A number of US dialects are involved in chain shifts like this one, which Labov (2012) argues are leading to greater regional differentiation than homogenization in the United States.

Beyond the phonetic properties of various phonemes,⁹ we can also ask about the number of phonemic categories in an accent and how they are organized. The main mechanisms by which a language variety comes to have a different number of phonemes are mergers and splits. There are many dialects in which some two phonemes have

---

⁸ In dialectology and sociolinguistics, words in small capital letters are conventionally used to represent vowel phonemes, whatever the actual pronunciation of this vowel in a given variety. This kind of representation is called a lexical set (Wells, 1982) because it picks out the set of words that, historically, share the same phoneme.

⁹ It could be argued that even if all the phonemes of two systems have quite different phonetic properties, it does not constitute a phonological difference so long as every phoneme in one system finds a structural equivalent in the other.
merged to form a single category. A well-known example in English involves the vowels in *lot* and *thought*. In some dialects, the vowels in these words remain distinct. However, many dialects have undergone a merger in which these lexical sets have combined into a single large lexical set, reducing the overall number of contrastive phonemic categories in the inventory (for American English, see Labov et al., 2006, p. 58). Conversely, when one phoneme splits into two, the number of contrastive phonemic categories increases. While mergers and splits constitute qualitative differences in the structure of the phonemic inventory, these effects are often not salient to the speakers themselves. Two speakers of American English varieties might disagree about whether words like *cot* and *caught* are pronounced the same, but they are unlikely to comment on or even notice this disagreement (Labov, 1994, p. 344).

Without reorganizing the underlying phonemic inventory, accents may exhibit different allophonic processes that affect how phonemes are realized in certain contexts (Labov, Fisher, Gylfadottir, Henderson, and Sneller, 2016; Sneller, 2018). This means rules like the variable palatalization process outlined in Section 3.1 may or may not exist in a given dialect. A useful example may be found in rhoticity: the pronunciation of /t/ when there is no following vowel (Scobbie, 2006). Most varieties of English spoken in the United States, Canada, Scotland, and Ireland are rhotic, meaning /t/ is always a consonant, typically an approximant like [ɹ]. However, many varieties of English spoken in England, Wales, Australia, and New Zealand are non-rhotic, meaning there is an obligatory allophonic rule turning /t/ into a vowel when it is followed by a consonant.

---

10 The mechanism for splitting is more complicated. Normally, it takes place in two stages: (1) an allophonic alternation stage, in which one phoneme is realized two different ways according to linguistic context, and (2) the loss of the triggering environment for allophony, so the different realizations are no longer in complementary distribution and must be reanalyzed as contrastive.
or a pause.¹¹ For other varieties still, most notably certain regional varieties of English from areas of the United Kingdom (Wells, 1970; French, 1989; Blaxter, Beeching, Coates, Murphy, and Robinson, 2019) and the United States (Labov, 1972, 2001; Feagin, 1990; Carmichael and Becker, 2019), consonantal or vocalic /r/ are both possible in instances of the same context. For speakers of these dialects, it makes sense to posit a variable rule that probabilistically turns non-prevocalic /r/ into a vowel.

Every person who has acquired language has acquired a particular accent, even if that accent happens to be held up as a prestigious or “standard” way of speaking. Prestigious language varieties are often assigned properties of neutrality or universality, but there is no objective linguistic basis for such a designation. For example, “standard” or “mainstream” American English is a non-uniform collection of varieties typically associated with white speakers from the Midwest or non-urban Northeast of the United States. Not only is the perspective that these varieties are objectively “neutral” born from racist and classist ideology, it conceals assumptions about linguistic representation and processing that should be interrogated. In Section 3.4.1, we unpack some possible consequences of failing to account for interspeaker variation.

### 3.3.2 Intraspeaker variation

We have just seen that a group of people all speaking the same language cannot be treated as uniform. It would also be inaccurate to characterize each individual as an invariant member of this heterogeneous group. Indeed, it is not an exaggeration to say that every utterance involves a number of decisions to produce words in one way and not another. We have already encountered a number of examples where an individual can produce the same utterance in different ways in Sections 3.1 and 3.2, noting that the issue of how these options are represented may not be straightforward. In reality, different types of intraspeaker variation probably work differently in this respect. In a recent review of phonological variation from a cognitive perspective, Bürki (2018) lays out some dimensions for classifying variable phonological phenomena, such as distinguishing between deletions, insertions, and substitutions. Here we build on that foundation by outlining some dimensions of structural classification that could prove relevant for word recognition and other elements of processing. Phonological variation is complex and pervasive, and attending to these complexities may prove fruitful in designing experiments and interpreting seemingly-incongruous results.

In at least some cases of intraspeaker variation, the most parsimonious analysis seems to be to allow for variation inside the lexicon itself, disrupting the basic notion of a form–meaning pair by ascribing multiple forms to a single meaning. For example, many speakers of American English can pronounce the word *economic* with

¹¹ We could entertain the possibility that some set of vowel phonemes (*near, square, start, north, force, cure, letter*) just have different phonetic properties in these dialects. However, this analysis is not as successful at capturing patterns of intervocalic /r/ (e.g., *he is* vs. *here is*).
an initial vowel like that in fleece or dress. Since those same speakers are not free to interchange those vowels otherwise, we might think that their lexicon contains two different forms for economic. This resembles the tomato example in Section 3.1 in that it is lexically idiosyncratic, but speakers actually produce both forms. On the other hand, many patterns of variation apply more generally, affecting all words with the relevant phonological properties. The same representational desiderata that motivate abstraction in the invariant phonology can be applied here. In Section 3.1, we suggested that such patterns could be accounted for through phonological rules that are stipulated to apply with some probability rather than obligatorily. Such a rule is called a variable rule and has long been a prominent way of thinking about phonological variation in quantitative sociolinguistics (Weinreich et al., 1968; Cedergren and Sankoff, 1974; but cf. Fasold, 1991). While the original variable rules were patterned on the phonological rule formalisms of Chomsky and Halle (1968), the modeling of phonological variation across a range of formal frameworks is a vibrant area of active research (for overviews, see Coetzee and Pater, 2011; Nagy, 2013).

In addition to the possibility of representing variation within the lexicon, we must contend with other non-phonological levels of structure. For example, there is some debate around whether phenomena like variable palatalization really constitute the same kind of process as their invariant counterparts, or instead come about as a consequence of how a sequence of segments is sometimes executed in the phonetics. More specifically, when producing the sequence press you, speakers may not always perfectly separate the alveolar articulation of [s] and the palatal articulation for [j]. Instead, these articulatory movements are produced simultaneously and the segments are coarticulated in a way that can acoustically resemble [ʃ]. In order to consider whether a process is phonological or phonetic, we must consider the properties of these two modules (see e.g., Pierrehumbert, 1990; Cohn, 1993 for in-depth discussions). One generally held distinction between phonological and phonetic operations concerns the properties of categoricity and gradience. Phonological processes are typically held to take, as both input and output, some finite number of discrete categories that the language user stores in memory, e.g., /s/, [s], [ʃ]. The phonetics, on the other hand, control all the continuous and infinitely subdivisible dimensions of the physical instantiation of language, e.g., all possible configurations of an idealized target [s]. As mentioned earlier, this is one difference between Andruski et al’s (1994) manipulation of VOT and the other, more categorical, variables laid out in Section 3.2.1.

Exploring the properties of categoricity and gradience in production, Zsiga (1995) uses electropalatography to investigate how speakers articulate underlying /ʃ/ (fresher), word-internal derived [ʃ] (pressure), and word-final derived [ʃ] (press you). She concludes that variable palatization does not result in precisely the same [ʃ] as in underlying /ʃ/ or obligatorily derived [ʃ]. The latter two were indistinguishable from each other, although this is not necessarily always the case for obligatory derivation.

12 Not to be mistaken for invariance versus variability.
either (e.g., Port, Mitleb, and O’Dell, 1981; Ernestus and Baayen, 2006). Similarly, Ellis and Hardcastle (2002) looked at variable place assimilation in sequences like green card, which could be produced with a velar nasal [ŋ] or an alveolar nasal [n] (compare obligatory word-internal assimilation, e.g., enter, amber, prank). They find that while some speakers variably produce a fully velar [ŋ], others retain some residual alveolar articulation. Importantly, just because the phonetic realization of, for example, palatalization in press you is different from that in pressure, it does not necessarily follow that the former is not phonological. The different types of palatalization cannot be represented with a single process anyway, since one is obligatory and one is variable.

As a larger point, phonological processes need not be structure preserving. That is, they do not have to result in sounds that are already part of the underlying inventory (Scobbie, 1995; Bermúdez-Otero, 2010). For instance, most speakers of American English will produce postvocalic /t/ and /d/ as a flap, [ɾ], before unstressed vowels (e.g., city, writer, rider), but /t/ is not generally considered to be a sound that is available for underlying lexical representations because it only occurs in certain predictable environments (Kiparsky, 1979; Kahn, 1980). Structure non-preserving allophony is licit whether it neutralizes the contrast between multiple underlying segments (e.g., latter and ladder become homophones), or there is only one possible underlying representation of a surface segment, like the word-final /t/ manipulations overviewed in Section 3.2.1 (Deelman and Connine, 2001; Sumner and Samuel, 2005). A proper diagnosis of categoricity in this sense, then, does not require the recreation of phonological categories that occur elsewhere in the system. Rather, it is only necessary that the various instances of forms be distributed in discrete categories, whatever the precise nature of these categories turns out to be.

To further exemplify this point, consider British English /t/-glottaling and TH-fronting. Most speakers of British Englishes can optionally realize /t/ as [t] or as a glottal stop, [ʔ], (Stuart-Smith, 1999; Fabricius, 2002) word-finally or before unstressed syllables in the same word (e.g., bottle, butter, bat). Like for flapping, /t/ is not typically considered part of the underlying inventory, that is, the process is structure non-preserving. Plus, instances of British English /t/-glottaling are not equally distributed across phonetic continua like the gradual reduction of a coronal gesture or constriction of the glottis. Rather, there is variable, but categorical, selection between discrete coronal and glottal closure options (Heyward, Turk, and Geng, 2014). In contrast, TH-fronting is a variable feature of many varieties of British English (Kerswill, 2003) and African American Language (Green, 2002; Sneller, 2020), whereby underlying interdental fricatives /θ ð/ can be realized as labiodental fricatives [f v]. Since labiodental fricatives are required for the representations of other words (e.g., free, vine), TH-fronting is a structure preserving process. As with all structure preserving processes, it is also neutralizing, since the contrast between interdental and labiodentals is lost when it applies (e.g., three becomes homophonous with free). The dimensions of structure

---

13 TH here stands in for both voiced /ð/ and voiceless /θ/ interdental fricatives.
preservation and neutralization are relevant for processing because they impinge on matters of the nature and number of stored segments, and potentially the readiness with which underlying forms are retrieved.

Of course, the invariance problem means that it is often not a straightforward task to identify categories. Phonetic variation is a constant, even if it is not solely responsible for a pattern of variation. In practice, this means that phonological categories manifest as individual **distributions** and not individual **points** in phonetic space, since instances where some category is the intended target will inevitably be perturbed by phonetic variation. Moreover, phonological processes are often developed from the stabilization of more gradient phonetic variation (Bermúdez-Otero, 2007), and these physiologically motivated phonetic phenomena can remain and co-exist even after a phonological process is established from them (Bermúdez-Otero, 2013). This means that phonological variables are often accompanied by diachronically related phonetic variation that resembles and even conceals them.

Another factor in determining the structural properties of a potential phonological variable involves the contexts in which it occurs and, specifically, how it interacts with morphology. As mentioned in Section 3.1, words ending in unstressed -ing can variably be pronounced with [ɪŋ] or [ɪn]. As it turns out, the form with a coronal nasal appears much more often in verbal forms, where -ing is a suffix (e.g., *working*~*workin’*), than in nouns (e.g., *awning*~*awnin’*). Even without detailed evidence of categories in phonetic space, many grammatical theories do not allow the phonetics to be directly affected by morphology like this (Fodor, 1983; Bermúdez-Otero, 2010). This kind of a relationship between morphology and phonetics is prevented by the concept of **modularity**, which relegates different kinds of operations to separate parts of the grammar that are strictly ordered: by the time a word or utterance gets to the point of phonetic implementation, its internal morphological structure is no longer relevant. If we apply modular reasoning to variable processes, we have to conclude that variation in -ing is not phonetic. However, there is a different source of ambiguity present in these kinds of variables. Specifically, we can also account for morphologically conditioned patterns in variation by saying the morphology has some capacity for variation itself. Just as [t] and [ʔ] are possible allophones of /t/, perhaps /ɪŋ/ and /ɪn/ are possible **allomorphs** of -ing. Thus, it is not clear whether an instance of the word *workin’* was rendered in this form by way of phonological process or if the speaker selected an alternative form of the -ing suffix with a coronal nasal.

Ultimately, these levels of structure (phonetics, phonology, morphology) are not always neatly separable, and what look like cases of a single variant may actually have a

---

**Table 3.1 Examples of variables classified by structure preservation and neutralization**

<table>
<thead>
<tr>
<th></th>
<th>Structure preserving</th>
<th>Structure non-preserving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutralizing</td>
<td>TH-fronting</td>
<td>Flapping</td>
</tr>
<tr>
<td>Non-neutralizing</td>
<td>–</td>
<td>/t/-glottaling</td>
</tr>
</tbody>
</table>
number of different sources. Furthermore, different phonological processes interact with the underlying inventory of sounds in different ways. Understanding these structural properties, more broadly, helps us to discern the kind of mechanics that are at play when someone produces phonological variation, and the nature of the processing tasks required to interpret the linguistic signal as it is perceived.

### 3.3.3 Non-linguistic factors in phonological variation

Our overview thus far may have given the impression that phonological variation merely injects uncertainty into multiple levels of structure. In reality, phonological variation is systematically conditioned by a number of factors. Weinreich et al. (1968) refer to this systematization as **orderly heterogeneity**. There are observable linguistic differences correlated with many social dimensions such as gender identity (Eckert, 1989; Kiesling, 2004; Zimman, 2009), race/ethnicity (Fought, 1999; Hoffman and Walker, 2010; Bucholtz, 2011; King, 2016; Holliday, 2019), class or socioeconomic status (Labov, 1966; Rickford, 1986; Eckert, 1988; Labov, 1990), and sexual orientation (Gaudio, 1994; Moonwoman-Baird, 1997; Bucholtz and Hall, 2004). All of these elements of a person’s identity are relevant to the language they produce and are negotiated with regard to the particular setting and audience of an utterance. The interplay between various social dimensions is particularly well exemplified in classic observations of the social stratification of phonological variation. Variants whose rate of use is correlated with a language user’s socioeconomic status are typically also correlated with formality. An early demonstration of this effect can be found in Labov’s (1966) investigation of the social stratification of rhoticity in New York City. Speakers increasingly used consonantal realizations of non-prevocalic /r/, the variant associated with American English speakers of a higher socioeconomic status, as they performed tasks inducing more linguistic self-monitoring and a more formal style.

Results like these suggest that speakers understand how phonological variation is socially stratified and can draw on this knowledge to inform their linguistic choices. As such, inter- and intraspeaker variation are not divorced from one another but closely intertwined. This basic premise is the foundation of “Third Wave Sociolinguistics” (Eckert, 2012), which focuses on individuals’ capacity to dynamically construct and perform their identities by using linguistic features that have garnered particular social meanings according to the context they appear in. The fact that individuals have detailed knowledge of how phonological variation is socially organized is further demonstrated by studies that explicitly elicit listener judgments of an interlocutor (Campbell-Kibler, 2009, 2011). These show broad agreement in terms of what social information can be inferred from certain linguistic behaviors. All of this is to say that language is inherently social, and any linguistic processing must occur in tandem with processing of social information.
3.4 Consequences of phonological variation for models of the mental lexicon

Section 3.3 provided a detailed, though not comprehensive, look at the possible grammatical underpinnings and structural outcomes of different types of phonological variation. Equipped with this background, we now turn to consider some additional ways in which these structural aspects of phonological variation might be relevant to processing (that is, beyond the widely explored questions of non-canonicality that we discussed in Section 3.2). We find it useful to frame some of these questions in terms of their methodological implications, in order to highlight the inescapability of these issues even for studies that are not designed to address phonological variation. However, the issues we spell out are not merely methodological, since the potential differences between a model talker and a participant that we will outline could equally well be thought of as potential differences between two real-world interlocutors.

3.4.1 When processing meets mismatching representations

A practical consideration that we have only touched on briefly so far is that experimental research in spoken word recognition must choose what form of each word to use as a stimulus. The usual practice is to create stimuli using a model talker whom the researchers judge to sound “standard,” producing word stimuli with the form that is taken to be “canonical” in some respect. As we discussed in Section 3.2, canonical forms may or may not have a special representational status, but even if they do, we must recognize that standard accents and canonical forms are inherently socially constructed. And even once we acknowledge that these are social constructs, there are cases where different options are equivalently standard or canonical in different varieties of the language. Choices that feel like neutral defaults to a researcher, then, simply have no guarantee of either matching what is in any given participant’s mental lexicon or reflecting the bulk of their real-world listening experience. And as our discussion in Section 3.3 makes clear, there are many dimensions along which the form chosen by the researcher might not align with participants’ mental representations.

Many of the under-explored issues we identify have to do with mismatches between language users with or without some merger: in other words, representational differences across individuals. Consider the word competitors that are entertained and eliminated over time in cohort-based word recognition models (Marslen-Wilson, 1987; see Magnuson and Crinnion, this volume, for more detail). In these models, when a listener hears an initial sound such as [k], they generate a list of possible word candidates starting...
with /k/: capture, kick, cotton, continent, caution, cough, and so on. But if the second sound is [a], a listener with a merged LOT-THOUGHT class (most often realized phonetically as [ɑ]) might eliminate only capture and kick from this (partial) list, while the listener with a LOT-THOUGHT distinction might additionally eliminate caution and cough (because that listener expects those words to contain /ɔ/, not /ɑ/). A related lexical property like neighborhood density,\textsuperscript{14}, which has been shown to influence word identification ease and accuracy (Vitevitch, Luce, Pisoni, and Auer, 1999; Vitevitch and Luce, 2016), might exhibit similar types of differences depending on whether it is calculated over LOT-THOUGHT merged or LOT-THOUGHT distinct lexicons.\textsuperscript{15} Interestingly, this suggests that dialects differing along phonemic inventory lines such as the LOT-THOUGHT merger might offer a useful opportunity to study homophone representation (Swinney, 1979; Caramazza, Costa, Miozzo, and Bi, 2001) because they may allow a minimal comparison between speakers for whom some word pairs are and are not homophones. Another merger-related question is whether nonwords used as experimental stimuli might sometimes have an unrecognized real word status to listeners with different lexical or phonological mental representations. A researcher might construct an intended nonword frind without taking into account that to a listener who has a pin/pen-merger (where dress and kit are merged before nasal consonants) this would be a perfectly good instance of the real word friend. The real-world flipside of this question is whether a pronunciation like [frɪnd] from a Southern-accented talker might be processed as a nonword by a non-Southern listener.

Is it plausible that such representational mismatches intervene in processing in this way, given the literature we already surveyed in Section 3.2? This premise is supported by the observation that misunderstandings arising from phonemic inventory differences are not uncommon in everyday conversational interaction (Labov, 1994, p. 324–327), even in seemingly disambiguating communicative contexts. It is also supported by a number of cross-dialectal comprehension studies targeting specific sources of representational mismatch (Labov, Karen, and Miller, 1991; Flanigan and Norris, 2000; Labov, 2010). In other words, although our Section 3.2 discussion emphasized evidence that listeners are eventually able to overcome the challenges of phonological variation, we should not conclude that those challenges do not arise at all. And methodologically speaking, we cannot safely assume that these issues are minor in scope or easily avoidable. For example, the presence vs. absence of the LOT-THOUGHT merger divides the geographic territory of US English approximately in half across a number of non-contiguous dialect regions (Labov et al., 2006, p. 59).

Beyond these interspeaker differences, we highlighted issues of structure preservation and neutralization in Section 3.3 because intraspeaker variation can give rise to parallel issues when it is structure preserving or neutralizing. For example, the variable deletion of word-final /t/ and /d/ in consonant clusters, which happens in every English dialect we are aware of, can generate homophony (past/pass, mold/mole) and erase morphological

\textsuperscript{14} Neighborhood density measures capture how many other words in the lexicon have a similar phonological shape; see Magnusson and Crinnion, this volume.

\textsuperscript{15} Exactly how these differences play out will depend on the calculation method and the treatment of homophones within that method.
information \textit{(jumped/jump)} but does not always do so \textit{(act, spent)}. These inconsistent lexical consequences pose challenges for listeners and researchers alike. Intraspeaker variation that is structure non-preserving, on the other hand, introduces variants that inherently signal their derived status by virtue of not existing in the phonemic inventory, a signal that could in principle trigger shifts in how lexical access processes proceed. Nonetheless, structure non-preserving intraspeaker variation can additionally generate homophony when it is neutralizing \textit{(latter/ladder)}, reintroducing some of the issues around cohort competitors and neighborhood density that we discussed for more basic merger examples. Empirical questions about how any particular variable is represented, such as whether it involves incompletely neutralized phonetic variants or an alternation between discrete phonemic units, take on new importance in light of the possibility that they may give rise to different processing consequences.

3.4.2 When social information comes into play

Work in the emerging area of \textit{sociolinguistic cognition} \cite{Campbell-Kibler2010,Loudermilk2013,Chevrot2018} suggests that listeners’ navigation of inter- and intraspeaker phonological variation during lexical access is guided by their experience with the social influences on variation such as we discussed in Section 3.3. There is robust evidence that listeners can use social information to make inferences about a speaker’s likely linguistic system.\footnote{It should be noted that listeners’ ability to use such information is modulated by listeners’ social attitudes \cite{Kang2009}.} They can then take into account whatever knowledge they have of that system, instead of relying on their own system, when identifying sounds and words produced by that speaker \cite{Niedzielski1999,Strand1999,Hay2010,DOnofrio2015, Hay2019}. This kind of reverse-engineering appears to be possible based on the presence of linguistic features that tend to co-occur (such as different features of a Southern accent) without extra social information \cite{Dahan2008}, although it remains an open question whether these effects are mediated by social inference \cite{Wade2020}. For example, in the \textit{frind/friend} example above, even a Southern-accented participant might be able to use the model talker’s non-Southern accent to infer that the model talker did not intend the word \textit{friend}. Of course, the fact that this listener might ultimately reach the correct conclusion about the intended nonword status of the stimulus does not rule out the possibility that they first retrieved \textit{friend} and then backtracked, or were otherwise delayed in ways that a non-Southern listener might not have been. The time course of reasoning about an interlocutor’s differing linguistic system, and how it interacts with possibly more basic lexical access mechanisms, is certainly not well established. Furthermore, while there is evidence that listeners can also use linguistic and social information to help guess the intended word when intraspeaker variation is in play \cite{Mitterer2009,Casasanto2008}, inherent variability means that incorporating such information can favor, but not guarantee, the correct outcome.
The fact that phonological variation encodes social information in speech, then, makes it an exciting frontier for understanding the mental lexicon. On the comprehension side, phonological variation is not simply noise that listeners must factor out. Rather, it is a rich source of structured information about speakers themselves and their likely behavior. Recent approaches to incorporating this kind of social information into processing models include the dual-route approach to socially weighted encoding proposed by Sumner et al. (2014) and the ideal adapter model of Kleinschmidt and Jaeger (2015; see also Kleinschmidt, 2019). On the production side, producing phonological variation demands that speakers dynamically shape their own speech to be fluid, connected, and socially and contextually appropriate; Babel and Munson (2014) give a useful overview on production issues related to variation.

However, this same social sensitivity is a reason that experimental work on variation needs to proceed with caution. It is easy to lose sight of the fact that a laboratory on a college campus is itself a social setting, one that for most people is far-removed from everyday life. Research that attempts to investigate socially meaningful phonological variation in the lab, but does not take into account the social properties of the experimental context itself, runs the dual risk of not only drawing scientifically unwarranted conclusions but also propping up the marginalization of “nonstandard” varieties.

3.4.3 Toward new advances in modeling the mental lexicon

The issue of phonological variation arises throughout this volume. Magnuson and Crinnion (this volume) point to the many sources of talker variation and a wide range of phonological processes producing deviations from canonical form as major challenges for current models of spoken word recognition. Creel (this volume) highlights the difficult challenge that pervasive variability poses to word learners. Kilbourn-Ceron and Goldrick (this volume) end their survey of word production by noting that we know very little about how words are produced in sentential contexts (as, indeed, words nearly always are). It appears that phonological variation poses one of the major obstacles preventing word recognition and production models from being able to cope not only with diverse talkers and social contexts but also with connected speech at all. Improving our understanding of phonological variation and its relationship to the mental lexicon thus promises to facilitate the modeling of word recognition in connected speech input and word production in context. The problems at hand are far from simple, but we believe that turning toward a view of lexical access as an inherently socially situated process offers the promise of bringing models of the mental lexicon and its use into a more detailed alignment with what human language users know and do.
4.1 Introduction

The encoding of speech signals and the mapping of speech to words are addressed in this chapter in the context of two parallel research agendas: the processing of speech as the initial step in spoken language comprehension, that is, the traditional domain of speech perception, and the neural implementation of the computations that underpin the encoding of speech and lexical access. The first question has been investigated extensively in cognitive science research, specifically in experimental psychology and psycholinguistics, focusing on how listeners retrieve linguistic information from the physical speech signal. The second question has been examined in cognitive and computational neuroscience, involving studies of the anatomical and functional infrastructure of the human brain to facilitate the processing of speech. This chapter reviews the literature from both research fields and provides a perspective on how to achieve explanatory connections between the two lines of research.

4.2 Computational-representational theories of speech perception

“Speech perception” refers to a set of operations that transform an auditory signal into representations (phonetic, phonological) of a form that can make contact with linguistic information stored in a listener’s mental lexicon (morphemes, roots, words). If we take the end goal (or, in David Marr’s terms, the “computational goal”) of speech perception to be the mapping of sounds to stored representations like words (whether in isolation—“virus”—or in context—“this virus stinks”), then the study of the encoding and decoding of speech needs to make reference to how stored information is represented. For now, we
do not have compelling evidence on exactly how words or morphemes (or any other types of knowledge, for that matter) are stored in memory. But while the details of how linguistic information is stored must remain speculative, one must bear in mind that the mapping of speech to words ultimately requires a commitment to the representational format of the input signals and the pieces of stored linguistic knowledge.

Two major concepts form the basis for much of cognitive science: representation and computation. Embick and Poeppel (2015), referring to computational-representational (CR) theories in language, suggest that existing theories typically make claims about the following questions: first, what representational format do different types of linguistic information take in a speech signal, and how is linguistic information represented in the mental lexicon? Second, what are the computations that are needed to transform acoustic events in a speech signal into phonological representations that subsequently interact with other linguistic dimensions (morphological, syntactic, etc.)? Figure 4.1 summarizes the conceptual architecture of the problem.

**Figure 4.1** Representations, transformations, computations: from the physical speech signal to phonological and lexical representations. Solid arrows: logically required feedforward steps. Dotted arrows: hypothesized feedback mappings. (a) The peripheral auditory system encodes continuously varying acoustic waveforms (left side: x-axis, time; y-axis, amplitude) decomposing the input signal into different bands (right panel: cochleagram) and conveying spectro-temporal information to auditory cortex via the afferent auditory pathway. (b) A series of intermediate representations may be necessary to map from a spectro-temporal representation of the acoustic input signal to the abstract representation of the word. Accordingly, multiple types of computations are required to assure transformations between different intermediate representations. (c) The hypothesized representation of the word ‘cat’ (/kat/) in the mental lexicon of a speaker/listener. Each of the three segments of this consonant-vowel-consonant word is built from distinctive features that, as a bundle, are definitional of phonemes that comprise the lexical item.
Unsurprisingly, the concepts of representation and computation are inherently linked, as theoretical commitment to the format of the mental representations should align with the nature of the relevant computations. To be concrete, two opposing views exist with regard to the format of lexical representation. On the one hand, many linguistically motivated theories argue that words are represented in the mental lexicon in terms of sequences of discrete, abstract segments or phonemes (Dresher, 2011; Jones, 1967), which can be further subdivided and described as bundles of distinctive features (Jakobson, Fant, and Halle, 1951; Stevens, 2002). On this view, the mapping from speech to words involves computations that translate continuous acoustic input into discretized segments, with each assigned a specific label within the phonemic/featural space. On the other hand, psychologically inspired theories propose that words are stored in the mental lexicon as auditory “episodes” that contain fine phonetic details and non-linguistic indexical information, such as voice of the speaker (Goldinger, 1998; Johnson, 1997; Port, 2007, 2010). On that view, speech perception requires accessing acoustically detailed exemplars of words, with little or no need for mapping onto abstract phonological representations.

We first provide a short perspective on the evidence from psycholinguistic research regarding this debate. We then present findings from neurobiological research and explain how these findings complement those from psycholinguistics and help unify features from both approaches to reach a more comprehensive understanding of speech perception.

4.3 Behavioral foundations for competing CR theories

4.3.1 Behavioral evidence of phonological abstraction

Phonological abstraction is a central research question in the perception of individual speech sounds. In such studies, listeners are typically presented with isolated speech sounds such as vowels or syllables and execute behavioral tasks (e.g., identification, discrimination). This line of research has convincingly demonstrated the existence of abstract phonological categories as part of listeners’ linguistic repertoire that drives perceptual processes. A major, well-established finding is the categorical nature of speech sound perception: listeners are perceptually more sensitive to acoustic differences between two speech sounds that belong to two distinct phoneme categories than to the same amount of acoustic difference between two speech sounds from the same phoneme category (e.g., Liberman, Harris, Hoffman, and Griffith, 1957). This phenomenon, referred to as categorical perception, reflects the tuning of listeners’ perceptual system to perceive acoustic differences that contrast different phonemes, which consequently allows the assignment of physically varying sounds to a single, functionally equivalent phonological representation.

Categorical perception supports the relevance of abstraction in the mapping from a continuous spectro-temporal space to a discrete phonological space (i.e., phonemes
and distinctive features) that underwrites lexical representations. Extensive research shows that listeners’ perceptual systems are tuned to the set of phonemes (and distinctive features) used in their native language, such that the perception of phonemes and phonetic features that do not exist in their native language can be compromised (Strange and Shafer, 2008). In fact, it is commonly observed that listeners have difficulty perceiving non-native phoneme contrasts (Best, 1994; Best, McRoberts, and Goodell, 2001). A well-known example consists of Japanese listeners’ persistent difficulty to distinguish between the English liquid consonants /r/ and /l/ (Goto, 1971; Miyawaki, Jenkins, Strange et al., 1975). Specifically, Japanese listeners map tokens from both English consonants onto the same consonant category in Japanese (/r/ or /w/: Best and Strange, 1992; Takagi, 1995; Yamada and Tohkura, 1992) while becoming insensitive to the acoustic difference between them (Iverson, Kuhl, Akahane-Yamada et al., 2003). Such perceptual difficulties for non-native phonemic contrasts, widely observed for both consonant and vowel contrasts across numerous languages (Strange and Shafer, 2008), demonstrate the impact of native phonemic categories in shaping listeners’ perceptual space. The data highlight the efficiency with which listeners’ speech perception system filters out acoustic variability that is not informative for distinguishing native phonemic categories, such that it becomes problematic to perceive and learn new phonemes from non-native languages. Note that this kind of systemic optimization applies to the perception of individual phonemes and to the level of distinctive phonetic features (Brown, 2000; Hancin-Bhatt, 1994; McAllister, Flege, and Piske, 2002). For instance, McAllister and colleagues (2002) observed that learning non-native Swedish long versus short vowel distinctions is difficult for native English and Spanish speakers, whose native languages do not employ “length” as a distinctive feature for vowels, while it is easier for Estonian speakers, who do make use of the length feature to distinguish between Estonian vowels. That is, although none of the participants from the three linguistic backgrounds were previously exposed to the specific Swedish phonemic contrast, Estonian speakers generalized their perceptual sensitivity to vowel length to distinguish between non-native vowel pairs with the same featural difference.

Decades of behavioral research on speech perception provide convincing evidence for the existence of abstract phonological representations in the mental lexicon (Figure 4.1c) as well as for the optimization of listeners’ perceptual systems to robustly map acoustic signals onto phonemic and featural categories used in the construction of the mental lexicon of their native language. This line of empirical evidence supports a correspondence between linguistic theories that describe the structure of language and the set of functional operations that putatively underlie the processing of speech. Findings from such empirical work may not be decisive in showing the necessity for any particular abstract phonological unit to be the foundational primitive representational unit to form words in the mental lexicon (see Kazanina, Bowers, and Idsardi, 2017, for a review of various phonological units that have been proposed for that role and see Samuel, 2020, for potential shortcomings of psycholinguistic work for the purpose of validating linguistic theories). However, the experimental evidence unequivocally demonstrates
that phonological abstraction at multiple levels is part of the functional architecture that maps the speech signal onto internally stored lexical representations. It is worth noting that the functional necessity of a phonological unit in encoding lexical forms can also be argued for from a theoretical perspective. For instance, using a modeling approach, one can show that the use of a certain phonological unit presents compelling computational advantages in the organization of word-forms in the mental lexicon in order to assure robust mapping between word-forms and the acoustic signal as well as with articulatory gestures (see Nozari, this volume, for a similar argument in the case of speech production).

4.3.2 Behavioral evidence for episodic encoding in speech perception

In contrast to phonologically-centered theories, episodic theories posit that detailed acoustic variations of a phoneme (a category that still exists) are functionally informative for the recognition of words and are therefore retained in the sub-lexical and lexical representations (Goldinger, 1998; Johnson, 1997; Port, 2007, 2010). These variations mainly include fine phonetic details, which result from specific articulatory realizations of a phoneme in various phonological environments, and indexical information, which includes acoustic cues for the identity and emotional state of a speaker, speech rate, as well as other prosodic properties of speech. Building on initial proposals of episodic theories, which assumed that the mental lexicon is solely composed of acoustically rich exemplars of different words (Goldinger, 1998; Port, 2007), modern revisions acknowledge the existence and functional relevance of abstract phonological representations in the lexicon while they maintain that acoustic details are encoded as part of the lexicon (Pierrehumbert, 2016). Here we highlight some evidence that supports the inclusion of phonetic variation and indexical information in the mental lexicon, alongside abstract phonological units.

The maintenance of fine-grained phonetic details of speech sounds during speech perception and mapping to words is mainly supported by experimental findings that challenged the notion of categorical perception (e.g., Liberman et al., 1957). Specifically, it has been argued that the appearance of categorical perception is largely due to the binary nature of response alternatives in identification and discrimination tasks (Lotto and Holt, 2000; Massaro and Cohen, 1983). Indeed, Massaro and Cohen (1983) found that when participants were asked to rate the degree to which each consonant instance from a /ba/-/pa/ continuum resembled one or the other category, their ratings appeared to be continuous rather than categorical. Subsequent studies confirmed listeners’ ability to judge the goodness of acoustic variants of the same phonemic category (e.g., Iverson and Kuhl, 1995; Kuhl, 1991). Moreover, recent studies using eye-tracking have revealed that the degree of perceptual goodness of a speech sound instance affects the process of word identification (McMurray, Tanenhaus, and Aslin, 2002; McMurray, Aslin, Tanenhaus, Spivey, and Subik, 2008). Such findings contradict the view that fine-grained