

Morphological Diversity and Linguistic Cognition

Morphological structures interact dynamically with lexical processing and storage, with the parameters of morphological typology being partly dependent on cognitive pathways for processing, storage, and generalization of word structure, and vice versa. Bringing together a team of well-known scholars, this book examines the relationship between linguistic cognition and the morphological diversity found in the world's languages. It includes research from across linguistic and cognitive science subdisciplines that looks at the nature of typological diversity and its relationship to cognition, touching on concepts such as complexity, interconnectedness within systems, and emergent organization. Chapters employ experimental, computational, corpus-based, and theoretical methods to examine specific morphological phenomena, and an overview chapter provides a synthesis of major research trends, contextualizing work from different methodological and philosophical perspectives. Offering a novel perspective on how cognition contributes to our understanding of word structure, it is essential reading for psycholinguists, theoreticians, typologists, computational modelers, and cognitive scientists.

ANDREA D. SIMS is an Associate Professor at The Ohio State University (USA). Her recent publications include *Inflectional Defectiveness* (Cambridge University Press, 2015), and *Understanding Morphology* (co-authored, Routledge, 2nd ed. 2010). She is co-editor of the journal *Word Structure* and serves on the editorial board of the journal *Morphology*.

ADAM USSISHKIN is a Professor at the University of Arizona (USA). His research focuses on the structure of the lexicon, morphology, and phonology. He has published in *Morphology*, *Language*, *The Mental Lexicon*, and *Language, Cognition and Neuroscience*. He is also an associate editor for *Language and Speech*.

JEFF PARKER is an Assistant Professor at Brigham Young University (USA). His research investigates inflection class systems from psycholinguistic, computational, and typological perspectives. He has published in morphology journals such as *Morphology* and *Word Structure*, the psycholinguistic journal *The Mental Lexicon*, and the Slavic-focused *Slavic and East European Journal*.

SAMANTHA WRAY is a Lecturer at Dartmouth College (USA). She focuses on spoken and written word processing for nonstandard dialects of Arabic, and other widely spoken underresearched languages of the world, such as Tagalog. She also works on the creation and improvement of computational resources for underresourced language varieties.

Morphological Diversity and Linguistic Cognition

Edited by

Andrea D. Sims

The Ohio State University

Adam Ussishkin

University of Arizona

Jeff Parker

Brigham Young University

Samantha Wray

Dartmouth College



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press & Assessment
 978-1-108-79094-9 — Morphological Diversity and Linguistic Cognition
 Andrea D. Sims, Adam Ussishkin, Jeff Parker, Samantha Wray
 Frontmatter
[More Information](#)



Shaftesbury Road, Cambridge CB2 8EA, United Kingdom
 One Liberty Plaza, 20th Floor, New York, NY 10006, USA
 477 Williamstown Road, Port Melbourne, VIC 3207, Australia
 314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
 103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
 Information on this title: www.cambridge.org/9781108790949

DOI: 10.1017/9781108807951

© Cambridge University Press & Assessment 2022

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

First published 2022
 First paperback edition 2025

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data

Names: Sims, Andrea D., editor. | Morphological Typology and Linguistic Cognition (Workshop) (2019 : Lexington, Kentucky)
 Title: Morphological diversity and linguistic cognition / edited by Andrea D. Sims, University of Arizona [and three others].
 Description: Cambridge ; New York, NY : Cambridge University Press, 2022. | Chapters based on presentations held at the workshop, Morphological Typology and Linguistic Cognition, held in July 2017 at the University of Kentucky. | Includes bibliographical references and index.
 Identifiers: LCCN 2021061917 | ISBN 9781108479899 (hardback) | ISBN 9781108807951 (ebook)
 Subjects: LCSH: Grammar, Comparative and general – Morphology – Congresses. | Typology (Linguistics) – Congresses. | Cognitive grammar – Congresses. | BISAC: LANGUAGE ARTS & DISCIPLINES / Linguistics / Morphology | LCGFT: Conference papers and proceedings.
 Classification: LCC P241 .M5985 2022 | DDC 415/.9–dc23/eng/20220105
 LC record available at <https://lcn.loc.gov/2021061917>

ISBN 978-1-108-47989-9 Hardback
 ISBN 978-1-108-79094-9 Paperback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>List of Figures</i>	<i>page</i> vii
<i>List of Tables</i>	x
<i>List of Contributors</i>	xiii
<i>Preface</i>	xv
<i>List of Abbreviations</i>	xvii
1 At the Intersection of Cognitive Processes and Linguistic Diversity	1
ANDREA D. SIMS, ADAM USSISHKIN, JEFF PARKER, AND SAMANTHA WRAY	
Part I In What Ways Is Language Processing Tuned to the Morphological Structure of a Language?	
2 Tuning Language Processing Mechanisms to a Language’s Morphology without Decomposition: The Case of Semantic Transparency	31
LAURIE BETH FELDMAN AND FERMÍN MOSCOSO DEL PRADO MARTÍN	
3 Productivity Effects on Morphological Processing in Maltese Auditory Word Recognition	56
SAMANTHA WRAY AND ADAM USSISHKIN	
4 Phonotactic and Morphological Effects in the Acceptability of Pseudowords	79
JEREMY M. NEEDLE, JANET B. PIERREHUMBERT, AND JENNIFER B. HAY	
	v

Part II What Role Does Cue Informativity Play in Learning and How the Lexicon Evolves Over Time?

5	How Agglutinative? Searching for Cues to Meaning in Choguita Rarámuri (Tarahumara) Using Discriminative Learning	121
	GABRIELA CABALLERO AND VSEVOLOD KAPATSINSKI	
6	Words, Probability, and Segmental Information: Less Probable Words Have More Informative Segments	160
	ADAM KING AND ANDREW WEDEL	
7	Learning Complex Morphological Patterns: The Role of Syncretism and Markedness	182
	SARA FINLEY	

Part III How Do System-Level Principles of Morphological Organization Emerge?

8	Morphology Gets More and More Complex, Unless It Doesn't	209
	ERIC MEINHARDT, ROBERT MALOUF, AND FARRELL ACKERMAN	
9	Network Structure and Inflection Class Predictability: Modeling the Emergence of Marginal Detraction	247
	JEFF PARKER, ROBERT REYNOLDS, AND ANDREA D. SIMS	
10	Rule Combination, Potentiation, Affix Telescoping	282
	GREGORY STUMP	

<i>References</i>	307
<i>Language Index</i>	347
<i>General Index</i>	349

Figures

2.1 Original and corrected funnel plots with CI for transparent and opaque facilitation based on differences between related and unrelated conditions.	<i>page 46</i>
2.2 Original and corrected funnel plots for the difference between transparent and opaque facilitation.	48
3.1 Experiment 1 stimuli base × surface frequency.	65
3.2 Experiment 1 RT from item onset × base frequency and RT × surface frequency.	68
3.3 Experiment 2 stimuli base × surface frequency.	71
3.4 Experiment 2 RT from item onset × base frequency and RT × surface frequency.	73
4.1 Example of graphone mappings.	90
4.2 Mean wordlikeness rating by log phonotactic probability scores.	97
4.3 Boxplots of phonotactic scores and wordlikeness ratings, separated by word length: (a) biphone scores, (b) triphone scores, (c) wordlikeness ratings (mean rating per item across participants).	98
4.4 Effect of a compound parse on the distribution of wordlikeness intercepts, for pseudowords of length four to seven.	102
4.5 Effect of a suffix parse on the distribution of wordlikeness intercepts, for pseudowords of length four to seven.	103
4.6 Interaction of vocabulary level with compound type as captured in Decomposition models, for pseudowords of length seven.	106
4.7 Interaction of vocabulary level with suffixation as captured in Decomposition models, for pseudowords of length seven.	107
4.8 Interactions of phonotactic score with suffixation.	108
5.1 The equilibrium equation model (NDL).	143
5.2 Rescorla-Wagner training with a slow learning rate.	144
5.3 Logistic perceptron training.	145
6.1 Hypothetical example of the distribution of information in words.	166
6.2 Segmental information in English monomorphemes of length nine or shorter.	170

viii List of Figures

6.3	Segmental information compared against $-\log$ trigram probability.	171
6.4	Histogram of the R^2 values for the original lexicon and 1,000 modified lexicons where the frequency has been scrambled.	173
6.5	Histogram of y-intercepts for length six words.	174
6.6	Predicted y-intercept for the per-word linear regressions over segmental information.	177
6.7	Difference between predicted early information for words calculated in the forward and reverse condition.	178
7.1	Examples of picture–meaning pairings, presented simultaneously.	188
7.2	Example of two-alternative forced-choice items.	190
7.3	Means and standard errors of correct forms selected for singular and dual conditions.	190
7.4	Means and standard errors of correct forms selected for the plural condition.	192
7.5	Means and standard errors of correct forms selected for Experiment 2.	194
8.1	A graphical illustration of drift acting on a small population with two variants.	225
8.2	Each plot shows the trajectories (under drift alone) over 20 generations of 10 simulated populations with population sizes (indicated on the right) varying from 20 to 1,000,000.	236
8.3	Each plot shows the trajectories (under drift alone) over 1,000 generations of 10 simulated populations with the population size (indicated on the right) varying from 20 to 1,000,000.	237
8.4	Each plot shows the trajectories (under drift and a moderate amount of selection) over 20 generations of 10 simulated populations with the population size (indicated on the right) varying from 20 to 1,000,000.	239
8.5	Each plot shows the trajectories (under drift and a moderate amount of selection) over 1,000 generations of 10 simulated populations with the population size (indicated on the right) varying from 20 to 1,000,000.	240
9.1	Undirected graph of the toy inflection class system.	251
9.2	Inflection class graph of Russian nouns.	252
9.3	Inflection class networks for Icelandic verbs (left) and Seri verbs (right).	253
9.4	Effect of log type frequency on entropy difference in nine natural languages.	256
9.5	Network graphs for ten model input conditions.	262
9.6	Mean number of classes in each generation for each condition.	265

List of Figures	ix
9.7 Input network structures for four conditions, with sample output network structures (each network is from a single agent in a single run of the model).	267
9.8 Effect of log type frequency on entropy difference in six model conditions.	270
10.1 The parts of the wordform ‘caused to become a rhino’ in Turkish.	291
10.2 The parts of the wordform ‘having caused to become an antelope’ in Turkish.	292

Tables

3.1	Maltese <i>binyanim</i> (verbal paradigms), adapted from Twist (2006).	page 59
3.2	Hapax-conditioned productivity of Maltese <i>binyanim</i> .	60
3.3	Experiment 1: Sample stimuli from six points in the frequency continua shown in Figure 3.1.	65
3.4	Experiment 2: Sample stimuli from different points in the frequency continua shown in Figure 3.3.	71
3.5	Stimuli for Experiment 1.	76
3.6	Stimuli for Experiment 2.	77
4.1	Summary of principal reported results for word perception, production, and acceptability of pseudowords in English.	80
4.2	Sample stimuli for the twenty-eight cells of the current study.	91
4.3	Sample parses for stimuli by length and parse type.	92
4.4	Summary of factors for the four Replication LMER models.	99
4.5	Summary of factors for the four Decomposition LMER models.	104
4.6	Examples of P2G/G2P instability results.	116
4.7	Drop-1 model comparison statistics for Replication models.	117
4.8	Drop-1 model comparison statistics for Decomposition models.	118
5.1	Agglutination versus flexion (Plank 1999).	128
5.2	Suffix positions and categories of the Choguita Rarámuri verb.	132
5.3	Stem levels of the Choguita Rarámuri verb.	132
5.4	Position of applicative, causative, desiderative, and future singular within the CR verb structure.	133
5.5	CR valence stem allomorphy.	135
5.6	Surveyed CR morphological constructions with suppletive allomorphs and regular morphophonological changes.	139
5.7	The best cues to each meaning according to each model.	146
5.8	The best cues to each meaning according to each model when all verbs are assumed to have the same token frequency ($n = 1$), rather than having a Zipfian token frequency distribution, but the order of verbs varies across runs of a model.	147

List of Tables	xi
5.9 Agglutinativity of each construction in CR according to each model.	153
5.10 Simplest cue–outcome structure that would produce a spurious excitor.	157
6.1 Linear model for –log unigram probability and segmental information, using token frequency to calculate segmental information.	170
6.2 Linear model for –log trigram probability and segmental information, using token frequency to calculate segmental information.	170
6.3 Linear model for –log unigram probability and segmental information, using word type to calculate segmental information.	172
6.4 Linear model for –log trigram probability and segmental information, using word type to calculate segmental information.	172
6.5 Results of linear model to compare forward and reverse lexicon.	176
6.6 Linear model for predicting y-intercept given lexicon type (forward or reverse), –log unigram probability of the word and their interaction.	178
6.7 Linear model for predicting y-intercept given lexicon type (forward or reverse), –log trigram probability of the word and their interaction.	178
7.1 Gender and number markings.	189
7.2 Training lists.	201
7.3 Test items: Dual condition.	202
7.4 Test items: Plural condition.	203
7.5 Test items: Singular condition.	205
8.1 Comparison of esoteric and exoteric situations.	212
9.1 Toy inflection class system.	251
9.2 Distribution of model input sets (conditions) by degree and average edge weight.	263
9.3 Summary of results of multiple linear regression models for node-level network properties as predictors of entropy difference in natural languages.	278
9.4 Summary of results of multiple linear regression models for node-level network properties as predictors of entropy difference in output of model.	279
9.5 Summary of descriptive statistics for the output of the Bayesian learning model.	281
10.1 Tokens of nouns in <i>-ization</i> in COCA for which forms of the corresponding verb in <i>-ize</i> are absent from the corpus.	293

xii	List of Tables	
10.2	The complex inflection of Noon attributive adjectives (attributive prefixes and definite suffixes).	294
10.3	The agent suffixes <i>-ŋ</i> and <i>-m</i> in the positive nonpreterite paradigm of the Limbu verb HU?MA? ‘teach.’	298
10.4	Some rules of Limbu verb inflection.	299

Contributors

FARRELL ACKERMAN Professor, Department of Linguistics, University of California, San Diego

GABRIELA CABALLERO Associate Professor, Department of Linguistics, University of California, San Diego

LAURIE BETH FELDMAN Distinguished Professor, Department of Psychology, University at Albany and Senior Scientist, Haskins Laboratories

SARA FINLEY Associate Professor, Department of Psychology, Pacific Lutheran University

JENNIFER B. HAY Professor, Department of Linguistics and English Language, University of Canterbury

VSEVOLOD KAPATSINSKI Professor, Department of Linguistics, University of Oregon

ADAM KING Computational Linguist, UBTech Robotics North American Research and Development Center

ROBERT MALOUF Professor, Department of Linguistics and Asian/Middle Eastern Languages, San Diego State University

ERIC MEINHARDT Postdoctoral Scholar, Department of Linguistics, University of California, San Diego

FERMÍN MOSCOSO DEL PRADO MARTÍN Lead Scientist, Lingvist Technologies

JEREMY M. NEEDLE Postdoctoral Fellow, Department of Linguistics, University of Toronto

JEFF PARKER Assistant Professor, Department of Linguistics, Brigham Young University

JANET B. PIERREHUMBERT Professor, Department of Engineering Science, Oxford University

Cambridge University Press & Assessment
978-1-108-79094-9 — Morphological Diversity and Linguistic Cognition
Andrea D. Sims, Adam Ussishkin, Jeff Parker, Samantha Wray
Frontmatter
[More Information](#)

xiv List of Contributors

ROBERT REYNOLDS Assistant Research Professor, Office of Digital Humanities, Brigham Young University

ANDREA D. SIMS Associate Professor, Department of Linguistics, The Ohio State University

GREGORY STUMP Emeritus Professor of Linguistics, Department of English, University of Kentucky

ADAM USSISHKIN Professor, Department of Linguistics, University of Arizona

ANDREW WEDEL Professor, Department of Linguistics, University of Arizona

SAMANTHA WRAY Lecturer, Department of Linguistics, Dartmouth College

Preface

The idea for this volume began with a workshop, *Morphological Typology and Linguistic Cognition*, that we organized in July 2017 in conjunction with the Linguistic Society of America's Linguistic Institute at the University of Kentucky. The workshop featured sixteen oral presentations and eight poster presentations by scholars representing twenty-five universities in nine countries. A select group of presenters was invited to submit chapters for this volume, based on thematic connections among their presentations. In most cases the chapters in the present volume reflect further developments in the authors' research since their presentations at the workshop, and the chapters engage with more recent developments in the field. The authors also discuss their ideas in more depth than was possible at the workshop. The present volume is thus an outgrowth of the workshop more than a proceedings volume.

Several core questions emerged as themes at the workshop, three of which form the organizational structure for the volume:

1. In what ways is language processing tuned to the morphological structure of a language?
2. What role does cue informativity play in learning and how the lexicon evolves over time?
3. How do system-level principles of morphological organization emerge?

The chapters all engage with one or more of these questions in some manner. In broad terms, the central hypothesis of the volume is that morphological structures interact dynamically with lexical processing, storage, and learning. Nine of the ten chapters present original research exploring the nature of this interaction from different perspectives. The remaining chapter, which serves as an introduction, highlights connections among the chapters and gives an overview of historical trends in the field that are reflected in the chapters. By bringing together research by different types of linguists – experimentalists, computational modelers, and formalists – our goal was to encourage discussion across subdisciplinary boundaries, to highlight the current state of research, and to help shape a research agenda that integrates these different methods and approaches.

We gratefully acknowledge financial support for the workshop from the National Science Foundation (BCS-1623932, PI Andrea Sims and Co-PI Adam Ussishkin), as well as supplemental support from the following units at The Ohio State University: the Department of Slavic and East European Languages and Cultures, the Center for Cognitive and Brain Sciences, and the Center for Slavic and East European Studies. The Center for Slavic and East European Studies' contribution was funded by the Comprehensive National Resource Center Grant from the International and Foreign Language Education division of the U.S. Department of Education. We also thank the 2017 Linguistic Institute organizers, Rusty Barrett and Andrew Hippisley, and all of the Institute staff – especially Ali Salehi – for in-kind and logistical support. The Brigham Young University Faculty Publishing Service provided copy-editing support for the volume.

We would also like to thank a number of colleagues for their efforts in reviewing abstracts submitted for the workshop and/or chapters for this volume: Farrell Ackerman, Mark Aronoff, Outi Bat-El, Sacha Beniamine, Olivier Bonami, Christina Burani, Gabriela Caballero, Greville Corbett, Micha Elsner, Laurie Beth Feldman, Sara Finley, Laura Gonnerman, Stefan Hartman, Volya Kapatsinski, Greg Kobele, Amy LaCross, Ana Luis, Sven Mattys, Frank Mollica, Mikhail Ordinin, Ingo Plag, Kevin Schluter, Scott Seyfarth, Gregory Stump, Tim Thornes, João Veríssimo, Geraldine Walther, and Kie Zuraw.

Most of all, we are beyond grateful to Ekaterina Kolbasova. As a Graduate Research Associate on the NSF grant, she worked incredibly hard to help run the workshop and coordinated and executed most of its logistics. She was absolutely instrumental to the project, handling issues large and small in an efficient, professional, and cheerful manner. Without her dedication, the workshop would not have been nearly as successful as it was. She went above and beyond the call of duty, even learning to drive on the highway. We are also appreciative of her work to create index terms for all of the chapters in the volume, even though the timeline stretched beyond what we expected. Thank you, Katja.

Finally, we thank Helen Barton at Cambridge University Press for her guidance and support (and patience) as we developed this project, Melissa Ward and Isabel Collins for shepherding it through the publication process, and two anonymous reviewers of the full volume for their detailed comments.

Abbreviations

1	first person
2	second person
3	third person
ADJ	adjective
AG	agent
ANIM	animate
APPL	applicative
CAUS	causative
CER	certainty
CL	end of clause particle
CL	noun class
CONJ	conjunct
DESID	desiderative
DIM	diminutive
DU	dual
EV	evidential
EXCL	exclusive
FUT	future
IMP	imperative
INCH	inchoative
INCL	inclusive
INT	intensifier
MOD	modality
N3	non-third-person
NANIM	non-animate
NDIM	non-diminutive
NONSG	non-singular
NINCL	non-inclusive
OBJ	object
PASS	passive
PAT	patient
PL	plural

xviii	List of Abbreviations
POSS	possessive
POT	potential
PROG	progressive
PROX	proximal
PST	past
PTCP	participle
REV	reversive
SG	singular
SUB	subordination
SBJ	subject
TAM	tense, aspect, mood
TR	transitive
VBLZ	verbalizer