

Exploring Spoken English Learner Language Using Corpora

‘Finally, some principled empirically-based information on qualities of spoken language in context! For several decades, a promise of second language (L2) corpus linguistics has been to revolutionize ways of teaching English to speakers of other languages. But prior to this book’s publication, most L2 corpus resources have focused on genres of the written language. As a result, specialists in research and teaching of the spoken language have felt somewhat frustrated. We are intrigued by the great potential corpus tools offer since we witness the many exciting ways in which they are applied to the written language. Partly because spoken corpora are notably more difficult to generate and analyze, the infusion of corpus tools into research and teaching of the spoken language has been limited. This book goes far in alleviating such concerns since it expands the landscape of corpus studies to include several core genres of the spoken language.’

—John Murphy *Georgia State University, USA*

‘This is a long-awaited volume presenting a brief introduction to corpus linguistics and a variety of excellent corpus-based studies on spoken learner language in the university setting. The authors provide a historical overview of the research in this area, offer a range of new approaches to the analysis, introduce accessible learner corpora, and discuss pedagogical applications. The reader finds a state-of-the-art picture of research and plenty of ideas for future directions to analyze spoken learner language. I highly recommend this volume to researchers and students alike.’

—Eniko Csomay *San Diego State University, USA*

Eric Friginal
Joseph J. Lee • Brittany Polat
Audrey Roberson

Exploring Spoken English Learner Language Using Corpora

Learner Talk

palgrave
macmillan

Eric Friginal
Applied Linguistics and ESL
Georgia State University
Atlanta, Georgia, USA

Brittany Polat
Georgia State University
Atlanta, Georgia, USA

Joseph J. Lee
Ohio University
Athens, Ohio, USA

Audrey Roberson
Hobart and William Smith Colleges
Geneva, New York, USA

ISBN 978-3-319-59899-4 ISBN 978-3-319-59900-7 (eBook)
DOI 10.1007/978-3-319-59900-7

Library of Congress Control Number: 2017946322

© The Editor(s) (if applicable) and The Author(s) 2017

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: © chipstudio / Getty Images

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Summary

As second language (L2) corpus studies expand into their third decade, innovations in computational technology and corpus creation have facilitated unprecedented access to authentic language in the classroom, including among non-native speakers (NNSs) of English. This book focuses on corpus-based analyses of learner oral production in university-level English or English as a Second Language (ESL) classrooms. Our analyses highlight three specialized corpora collected for the three empirical parts of this book, explored using a range of corpus approaches and methods: (1) learner talk in the English for Academic Purposes (EAP) classroom, (2) learner talk in English language experience interviews, and (3) learner talk in peer response/feedback activities. Historical and methodological perspectives in exploring spoken learner corpora, pedagogical applications, and future directions in studying learner language are discussed. A synthesis of corpus-based research of spoken learner language, list of available corpora and online databases, and an introduction to corpus linguistics and corpus tools and approaches are provided in the first two chapters of the book.

Acknowledgement

We would like to thank our mentors and colleagues at the Department of Applied Linguistics and ESL at Georgia State University (GSU), especially Gayle Nelson, John Murphy, and Sara Cushing for their guidance and critical perspectives in developing the three empirical studies presented in this book. Thanks to Mike Cullom for his valuable insights and reviews of earlier drafts of this book, our Palgrave Macmillan commissioning editors and reviewers, and the staff of the Longview Public Library, Longview, WA. We recognize Douglas Biber and Randi Reppen at Northern Arizona University; Lucy Pickering at Texas A&M, Commerce; John Swales and Rita Simpson-Vlach at University of Michigan and for their work with MICASE; Laurence Anthony at Waseda University; and the Learner Corpus Association (Founding Members: Gaëtanelle Gilquin, Sylviane Granger, Fanny Meunier and Magali Paquot at the Centre for English Corpus Linguistics, Université Catholique de Louvain) for leading the way with their corpus tools and seminal studies of learner language. Much appreciation to our GSU colleagues: Diane Belcher, Stephanie Lindemann, Scott Crossley, Youjin Kim, Ute Römer, Jack A. Hardy, Pam Pearson, Nic Subtirelu, Cassie Leymarie, and many other collaborators who have assisted in the data collection and analyses in various parts of this book. We are grateful to all our study participants and especially the students, instructors, and administrators at GSU's Intensive English Program (Cheryl Delk-Le Good, John Bunting, Debra Snell, Louise

Gobron, and Alison Camacho). Data collection in Part 3 of this book was supported by grants from the Educational Testing Service (ETS) and *Language Learning*. Finally, we dedicate this book to our families and friends, and thank them for their love and support: Mike and Beth Cullom, Donna and Ela Friginal; Chang Keun Lee and Joanne Y. Lee, and Helen, Hetty, and Jules Lee; Ali, Guinevere, Thomas, and Charlie Polat; Susan and Jim Roberson, and Michael Mills.

Eric Friginal
Joseph J. Lee
Brittany Polat
Audrey Roberson

Contents

Part I	Introduction	1
1	Exploring Spoken English Learner Language Using Corpora	3
2	Corpora of Spoken Academic Discourse and Learner Talk: A Survey	35
Part II	Learner Talk in the Classroom	65
3	Learner (and Teacher) Talk in EAP Classroom Discourse	67
4	Hedging and Boosting in EAP Classroom Discourse	77
5	<i>You, I, and We</i> : Personal Pronouns in EAP Classroom Discourse	95
6	<i>This/That, Here/There</i> : Spatial Deixis in EAP Classroom Discourse	115

Part III	Learner Talk in Language Experience Interviews	129
7	Exploring Learner Talk in English Interviews	131
8	Thematic Cluster Analysis of the L2 Experience Interview Corpus	141
9	Psychosocial Dimensions of Learner Language	153
10	Profiles of Experience in Learner Talk	167
Part IV	Learner Talk in Peer Response Activities	197
11	Understanding Learner Talk About Writing: The Second Language Peer Response (L2PR) Corpus	199
12	Social Dynamics During Peer Response: <i>Patterns of Interaction</i> in the L2PR Corpus	219
13	Linguistic Features of Collaboration in Peer Response: Modal Verbs as Stance Markers	245
Part V	Conclusion and Future Directions	261
14	Corpus-Based Studies of Learner Talk: Conclusion and Future Directions	263
	Appendix A: Transcription Conventions for the L2CD (Adapted from Jefferson 2004; Simpson et al. 2002)	277

Appendix B: Hedges and Boosters Investigated (Adapted from Hyland 2005, pp. 221–223)	279
References	281
Index	297

About the Authors

Eric Friginal is Associate Professor of Applied Linguistics at the Department of Applied Linguistics and ESL, and Director of International Programs, College of Arts and Sciences, at Georgia State University. He specializes in (applied) corpus linguistics, sociolinguistics, cross-cultural communication, and the analysis of spoken professional discourse. His recent books include *Talking at Work: Corpus-Based Explorations of Workplace Discourse* (2016, Palgrave Macmillan), co-edited with Lucy Pickering and Shelley Staples; *Studies in Corpus-Based Sociolinguistics* and *Corpus Linguistics for English Teachers* (2017–2018, Routledge).

Joseph J. Lee is the Assistant Director of the ELIP Academic & Global Communication Program, and Director of ELIP Center for Academic Communication: Tutoring Services in the Department of Linguistics, Ohio University. His research and teaching interests include ESP/EAP, genre studies, classroom discourse, advanced academic literacies, applied corpus linguistics, and teacher education. His recent publications include research articles in *English for Specific Purposes*, *Journal of English for Academic Purposes*, and *Journal of Second Language Writing*.

Brittany Polat is an independent ESL researcher based in Lakeland, Florida. Her research interests include second language acquisition, pragmatics, and corpus linguistics. Her research has appeared in journals such as *Applied Linguistics*, *Journal of Pragmatics*, and *Corpus Linguistics Research*.

Audrey Roberson is Assistant Professor of Education at Hobart and William Smith Colleges in Geneva, New York, where she oversees TESOL certification in the department's Teacher Education Program, as well as directs a certificate program in TEFL. Her research interests include language teacher preparation, applied corpus linguistics, interaction in second language learning, and second language writing. She has co-authored articles in *Corpora* and in the composition journal *Across the Disciplines*.

List of Figures

Fig. 2.1	Major stance features across registers (Adapted from Biber 2006a)	41
Fig. 2.2	Comparison of student texts in Dim 4: Personal narrative vs. non-narrative discourse (Adapted from Friginal and Polat 2015)	49
Fig. 10.1	Comparison of psychosocial features in all clusters	169
Fig. 10.2	Significant features of Narrative cluster	173
Fig. 10.3	Significant features of Cognitive cluster	175
Fig. 10.4	Significant features of Affective cluster	180
Fig. 11.1	Storch's (2002) <i>Patterns of Interaction</i>	201

List of Tables

Table 1.1	Collocations of the word <i>know</i> (first left and first right)	27
Table 1.2	Comparison of the most common 4-grams in call-taker and caller interaction in business call centers	29
Table 1.3	Biber's (1988) co-occurring features in Factor 1	33
Table 2.1	MICASE word counts by speech event type and student/faculty and staff 'participation' percentages	37
Table 2.2	Demographic groups in MICASE	38
Table 2.3	Composition of the T2K-SWAL Corpus (spoken texts)	40
Table 2.4	Linguistic composition of Dim 4 from LINDSEI (Friginal and Polat 2015)	47
Table 2.5	Spoken and written registers of the International Corpus of English	52
Table 2.6	ICE components tagged results using the Biber Tagger (data normalized per 1000 words)	54
Table 2.7	Spoken English learner corpora from research groups around the world	56
Table 3.1	Description of the L2CD corpus (Lee 2011)	73
Table 3.2	Description of the L2CD-S and L2CD-T sub-corpora	74
Table 4.1	Comparison of hedges and boosters in the two sub-corpora	82
Table 4.2	Top five most frequent hedging devices in the two sub-corpora	84
Table 4.3	Comparison of hedge sub-functions in the two sub-corpora	86

xviii **List of Tables**

Table 4.4	Top five most frequent boosting devices in the two sub-corpora	90
Table 4.5	Comparison of booster sub-functions in the two sub-corpora	92
Table 5.1	Framework for personal pronoun classification	99
Table 5.2	Comparison of personal pronouns in the two sub-corpora	102
Table 5.3	Comparison of ‘we’ in the two sub-corpora	105
Table 5.4	Comparison of ‘you’ in the two sub-corpora	108
Table 6.1	Comparison of proximal and distal deixis in the two sub-corpora	121
Table 6.2	Comparison of demonstratives in the two sub-corpora	123
Table 6.3	Comparison of “here” and “there” in the two sub-corpora	125
Table 7.1	Native languages represented by participants	133
Table 7.2	Academic disciplines of participants	134
Table 7.3	Interview protocol	135
Table 7.4	Summary of research analyses using L2 Experience Interview Corpus (Polat 2013a)	138
Table 8.1	Most representative lemmas in Classroom cluster	142
Table 8.2	Most representative lemmas in Communicating cluster	144
Table 8.3	Most representative lemmas in Studying cluster	146
Table 8.4	Comments on grammar-translation teaching methods	149
Table 8.5	Comments reflecting changing L2 learning experience	150
Table 9.1	Rotated component matrix of psychosocial features	154
Table 9.2	Component features	155
Table 9.3	Positive psychosocial features of Dimension 1	156
Table 9.4	Positive psychosocial features of Dimension 2	159
Table 9.5	Positive psychosocial features of Dimension 3	160
Table 9.6	Positive psychosocial features of Dimension 4	162
Table 10.1	Comparison of clusters by nationality	170
Table 10.2	Comparison of clusters by geographic region	171
Table 10.3	Comparison of clusters by academic discipline	172
Table 10.4	Positive psychosocial features of Narrative cluster	173
Table 10.5	Positive psychosocial features of Cognitive cluster	176
Table 10.6	Positive psychosocial features of Affective cluster	181
Table 10.7	Profiles of L2 learning experience	184
Table 10.8	Descriptive statistics for TOEFL scores	185
Table 10.9	Means and standard deviations of TOEFL score by cluster	185

Table 10.10	ANOVA summary table for analysis of TOEFL scores by cluster	185
Table 10.11	Interview length by cluster	194
Table 10.12	ANOVA summary table for interview length by cluster	194
Table 11.1	Features of Storch's (2002) <i>Patterns of Interaction</i>	205
Table 11.2	Participant characteristics	208
Table 11.3	Transcription conventions for peer response transcripts (Adapted from Ellis and Barkhuizen 2005)	209
Table 11.4	L2PR corpus composition	209
Table 12.1	Patterns of interaction in the L2PR corpus (Features from Storch 2002; Zheng 2012)	220
Table 12.2	Patterns of interaction for each transcript, across three sessions	223
Table 12.3	Mean number of turns and length of turns by pattern of interaction	224
Table 12.4	Provision and implementation of specific, revision-oriented comments, by writer role	233
Table 12.5	Mean score gains from first to second draft, by writer role	242
Table 13.1	Sub-corpora of the L2PR corpus (Roberson 2015)	248
Table 13.2	Distribution of modals by class, raw/normed per 10,000 words	250
Table 13.3	Distribution of frequent modals (raw/normed per 10,000 words)	251
Table 14.1	Summary of analyses and findings of Chaps. 8, 9, and 10	267