World Englishes is one of the most active fields of present-day linguistic research. This book provides a comprehensive overview of variation in English across the globe, by analyzing a large corpus of texts from the English-speaking world. The analysis draws on a larger number of individual texts and linguistic features than previous studies have done. It includes a systematic comparison of the influence of register and national variety on variation in English worldwide, and explores patterns of variation in both qualitative and quantitative terms. The results provide a testing ground for theoretical models, and highlight the importance of register in mediating linguistic differences across national varieties.

Axel Bohmann is Assistant Professor of English Linguistics at University of Freiburg, Germany.
The aim of this series is to provide a framework for original studies of English, both present-day and past. All books are based securely on empirical research, and represent theoretical and descriptive contributions to our knowledge of national and international varieties of English, both written and spoken. The series covers a broad range of topics and approaches, including syntax, phonology, grammar, vocabulary, discourse, pragmatics, and sociolinguistics, and is aimed at an international readership.

Already published in this series:

Christian Meierkord: Interactions across Englishes: Linguistic Choices in Local and International Contact Situations
Haruko Momma: From Philology to English Studies: Language and Culture in the Nineteenth Century
Raymond Hickey (ed.): Standards of English: Codified Varieties around the World
Beneditk Szmreczanyi: Grammatical Variation in British English Dialects: A Study in Corpus-Based Dialectometry
Daniel Schreier and Marianne Hundt (eds.): English as a Contact Language
Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis (eds.): The Verb Phrase in English: Investigating Recent Language Change with Corpora
Martin Hilpert: Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax
Jakob R. E. Leimgruber: Singapore English: Structure, Variation, and Usage
Christoph Rühlemann: Narrative in English Conversation: A Corpus Analysis of Storytelling
Dagmar Deuber: English in the Caribbean: Variation, Style and Standards in Jamaica and Trinidad
Eva Berlage: Noun Phrase Complexity in English
Nicole Dehê: Parentheticals in Spoken English: The Syntax-Prosody Relation
Jock O. Wong: The Culture of Singapore English
Marianne Hundt (ed.): *Late Modern English Syntax*
Irmtraut Taavitsainen, Merja Kytö, Claudia Claridge, and Jeremy Smith (eds.): *Developments in English: Expanding Electronic Evidence*
Arne Lohmann: *English Co-ordinate Constructions: A Processing Perspective on Constituent Order*
Nuria Yáñez-Bouza: *Grammar, Rhetoric and Usage in English: Preposition Placement 1500–1900*
Anita Auer, Daniel Schreier and Richard J. Watts (eds.): *Letter Writing and Language Change*
John Flowerdew and Richard W. Forest: *Signalling Nouns in English: A Corpus-Based Discourse Approach*
Jeffrey P. Williams, Edgar W. Schneider, Peter Trudgill, and Daniel Schreier (eds.): *Further Studies in the Lesser-Known Varieties of English*
Jack Grieve: *Regional Variation in Written American English*
Douglas Biber and Bethany Gray: *Grammatical Complexity in Academic English: Linguistics Change in Writing*
Gjertrud Flermoen Stenbrenden: *Long-Vowel Shifts in English, c. 1050–1700: Evidence from Spelling*
Zoya G. Proshina and Anna A. Eddy (eds.): *Russian English: History, Functions, and Features*
Raymond Hickey (ed.): *Listening to the Past: Audio Records of Accents of English*
Phillip Wallage: *Negation in Early English: Grammatical and Functional Change*
Marianne Hundt, Sandra Mollin and Simone E. Penninger: *The Changing English Language: Psycholinguistic Perspectives*
Joanna Kopaczyk and Hans Sauer (eds.): *Binomials in the History of English: Fixed and Flexible*
Alexander Haselow: *Spontaneous Spoken English: An Integrated Approach to the Emergent Grammar of Speech*
Christina Sanchez-Stockhammer: *English Compounds and Their Spelling*
David West Brown: *English and Empire: Language History, Dialect, and the Digital Archive*
Paula Rodríguez-Puente: *The English Phrasal Verb, 1650–present: History, Stylistic Drifts, and Lexicalisation*
Erik R. Thomas (ed.): *Mexican American English: Substrate Influence and the Birth of an Ethnoloc*
Thomas Hoffmann: *English Comparative Correlatives: Diachronic and Synchronic Variation at the Lexicon-Syntax Interface*
Nuria Yáñez-Bouza, Emma Moore, Linda van Bergen, and Willem B. Hollmann (eds.): *Categories, Constructions, and Change in English Syntax*
Raymond Hickey (ed.): *English in the German-speaking World*
Axel Bohmann: *Variation in English World-Wide: Registers and Global Varieties*

Earlier titles not listed are also available.
VARIATION IN ENGLISH WORLDWIDE

Registers and Global Varieties

AXEL BOHMANN

University of Freiburg
I dedicate this book to my parents.
## Contents

**List of Figures** page xii
**List of Tables** xv
**Preface and Acknowledgments** xvii
**List of Abbreviations** xviii

1 Introduction 1
   1.1 Research Context 1
      1.1.1 World Englishes 2
      1.1.2 Aggregation-Based Linguistics 3
      1.1.3 MD Analysis 4
   1.2 Research Objectives 5
   1.3 Outline 6

2 The World of English: Variation in Geography and Register 9
   2.1 Geographic Variation in English Worldwide 9
      2.1.1 The Circles Model 10
      2.1.2 The Dynamic Model 13
      2.1.3 The World System of Englishes 19
      2.1.4 The Models Against the Empirical Record 23
   2.2 Register Variation in English 24
      2.2.1 Text, Register, Genre, Style: A Clarification of Terms 25
      2.2.2 Thirty Years of MD Analysis: An Overview 27
   2.3 Geography and Register in WE Studies 32
   2.4 The Digital Frontier: Computer-Mediated Discourse and Global
      English Variation 35

3 Quantifying Linguistic Variation 41
   3.1 Language Variation and Change: Labovian Sociolinguistics 41
      3.1.1 The Linguistic Variable 42
      3.1.2 The Speech Community and the Concept of Prestige 43
      3.1.3 The Sociolinguistic Interview: Vernacular and Authentic Data 45
   3.2 Corpora and Frequencies: The Text-Linguistic Perspective 48
      3.2.1 Units of Analysis and Analytical Goals 49
Table of Contents

3.2.2 Patterns and Explanations of Linguistic Variation 50
3.3 Feature-Aggregation Approaches: Dialectometry and Register Analysis 52
3.4 The MD Approach to Aggregate Variation 55

4 The Space of Variation in the Present Study 59
4.1 Corpus Data 59
4.1.1 The International Corpus of English 59
4.1.2 The Global English Twitter Corpus 65
4.2 Linguistic Features Under Analysis 66
4.2.1 Selection of Features 66
4.2.2 The Feature Catalog 73
4.3 Statistical Analysis: Exploratory Factor Analysis 84
4.3.1 General Description of the Method 84
4.3.2 Parameter Settings 87
4.3.3 Overview of the Factor Solution 93

5 General Situational Dimensions of Variation 101
5.1 Involved Versus Informational Production 101
5.1.1 Salient Linguistic Features 102
5.1.2 Patterning of Texts 105
5.1.3 Sample Texts 110
5.2 Collaborative Communicative Orientation 111
5.2.1 Salient Linguistic Features 111
5.2.2 Patterning of Texts 114
5.2.3 Sample Texts 117
5.3 Conceptual Versus Concrete Informational Focus 119
5.3.1 Salient Linguistic Features 119
5.3.2 Patterning of Texts 120
5.3.3 Sample Texts 123

6 Register-Specific Dimensions 125
6.1 Canonical Narrative Focus 125
6.1.1 Salient Linguistic Features 125
6.1.2 Patterning of Texts 127
6.1.3 Sample Texts 132
6.2 Situational Anchoring of Reference 133
6.2.1 Salient Linguistic Features 134
6.2.2 Patterning of Texts 135
6.2.3 Sample Texts 138
6.3 Colloquial Markedness 140
6.3.1 Salient Linguistic Features 140
6.3.2 Patterning of Texts 142
6.3.3 Sample Texts 145
Table of Contents

7 Dimensions with Other Patterns of Distribution 147
  7.1 Explicit Stance-Marking 147
     7.1.1 Salient Linguistic Features 148
     7.1.2 Patterning of Texts and Sample Analyses 149
     7.1.3 Inter-Varietal Differentiation 151
  7.2 Future-Oriented Discourse 153
     7.2.1 Salient Linguistic Features 153
     7.2.2 Patterning of Texts 154
     7.2.3 Sample Texts 158
  7.3 Assertion of Factual Validity 159
     7.3.1 Salient Linguistic Features 160
     7.3.2 Patterning of Texts 161
     7.3.3 Sample Texts 164
  7.4 Addressee-Orientation 166
     7.4.1 Salient Linguistic Features 167
     7.4.2 Patterning of Texts 167
     7.4.3 Sample Texts 171

8 Discussion: Feature Space and Geographical Space 173
  8.1 The Space of Variation as a Whole 173
  8.2 The Respective Influence of Register and Variety 180
     8.2.1 Individual Patterns and Their Relations to WE Models 182
     8.2.2 On the Irrelevance of Variety 187

9 Conclusion 191
  9.1 Summary of Results 191
  9.2 An Evaluation of Theoretical Models 193
  9.3 The Differential Influence of Register and Variety 194

Appendix I Descriptive Statistics for the 236 Extracted Variables 197
Appendix II Structure Coefficients and Communality (h²) Values for the 236 Features 207
Appendix III Pattern Coefficients for the 236 Features 217
References 227
Index 249
Figures

2.1 Kachru’s Circles Model. As represented by Mesthrie (2008: 31).


4.1 Four hypothetical distributions of observations in a three-dimensional variable space. (Adapted from Baayen 2008: 128).

5.1 Distribution of dimension scores across modalities: INVOLVED VERSUS INFORMATIONAL PRODUCTION.

5.2 Distribution of dimension scores across corpus text categories: INVOLVED VERSUS INFORMATIONAL PRODUCTION.

5.3 Distribution of dimension scores across varieties and modalities: INVOLVED VERSUS INFORMATIONAL PRODUCTION.

5.4 Distribution of dimension scores across corpus text categories: COLLABORATIVE COMMUNICATIVE ORIENTATION.

5.5 Dimension scores for all corpus texts individually: COLLABORATIVE COMMUNICATIVE ORIENTATION.

5.6 Distribution of dimension scores across varieties and spoken registers: COLLABORATIVE COMMUNICATIVE ORIENTATION.

5.7 Distribution of dimension scores across corpus text categories: CONCEPTUAL VERSUS CONCRETE INFORMATIONAL FOCUS.

5.8 Distribution of dimension scores across varieties and modalities: CONCEPTUAL VERSUS CONCRETE INFORMATIONAL FOCUS.

5.9 Differences in dimension scores between spoken and written corpus material, plotted for each country: CONCEPTUAL VERSUS CONCRETE INFORMATIONAL FOCUS.

6.1 Distribution of dimension scores across corpus text categories: CANONICAL NARRATIVE FOCUS.
List of Figures

6.2 Distribution of dimension scores across varieties and modalities (fiction writing excluded): canonical narrative focus. 129
6.3 Distributions of dimension scores for countries in different phases according to the Dynamic Model: canonical narrative focus. 130
6.4 Distributions of dimension scores for countries in different phases according to the Dynamic Model; TwICE data only: canonical narrative focus. 131
6.5 Distributions of dimension scores across corpus text categories: situational anchoring of reference. 136
6.6 Distributions of dimension scores across varieties and modalities: canonical narrative focus. 138
6.7 Distributions of dimension scores across corpus text categories: colloquial markedness. 142
6.8 Distributions of dimension scores across varieties and modalities: colloquial markedness. 144
7.1 Distributions of dimension scores across corpus text categories: explicit stance-marking. 150
7.2 Distributions of dimension scores across varieties and modalities: explicit stance-marking. 152
7.3 Distributions of dimension scores across corpus text categories: future-oriented discourse. 155
7.4 Distributions of dimension scores across varieties: future-oriented discourse. 156
7.5 Distributions of dimension scores across varieties and modalities: future-oriented discourse. 157
7.6 Distributions of dimension scores across varieties in spoken registers: future-oriented discourse. 158
7.7 Distributions of dimension scores across corpus text categories: assertion of factual validity. 162
7.8 Distributions of dimension scores across varieties and modalities: assertion of factual validity. 164
7.9 Distributions of dimension scores across corpus text categories: addressee-orientation. 169
7.10 Distributions of dimension scores across varieties and modalities: addressee-orientation. 170
8.1 Hierarchical cluster analysis of countries in the ICE/TwICE data. 181
List of Figures

8.2 Distribution of dimension scores across the Inner and Outer Circle in different modalities: explicit stance-marking. 183
8.3 Distribution of dimension scores across different developmental phases and modalities: explicit stance-marking. 184
Tables

2.1 Linguistic features associated with the first six dimensions of variation in Biber (1988); adapted from table 1 in Biber (2014: 11).

4.1 Sampling framework of the ICE corpus project (based on Greenbaum & Nelson 1996).
4.2 The ten national standard varieties represented in the study.
4.3 Countries represented in the TwICE corpus.
4.4 Overview of the ten dimensions of variation.
4.5 Inter-factor correlations among the ten dimensions.
8.1 Explanatory utility of World Englishes models along the ten dimensions.
8.2 Coefficients of determination for linear models predicting dimension scores from variety and register.
Preface and Acknowledgments

This book is a condensed version of my Ph.D. dissertation (The University of Texas at Austin, 2017). During the research stages I received generous institutional, academic, and financial support from the Department of English at UT Austin. While revising the manuscript for publication, my current place of employment at the University of Freiburg provided an ideally stimulating and productive working environment.

I am grateful to the following people, all of whom have contributed in unique ways to provide me with strength, criticism, and relief while writing this book: my dissertation committee members Lars Hinrichs, Mary Blockley, Barbara Bullock, Katrin Erk, and Coleman Hutchison, who have been encouraging, stimulating and – most importantly – patient throughout the process; my long-term graduate adviser, the late Wayne Lesser, whom I will always remember as one of the most supportive and warm-hearted people in academia; Patricia Schaub, my graduate program coordinator, who got me out of more than one administrative pickle; my former teacher and current supervisor Christian Mair; my Freiburg office mates Mirka Honkanen and Fatlum Sadiku, and all my other colleagues at the University of Freiburg; Danae Perez for careful proofreading of the manuscript at various stages and general emotional support; Jocelyne Daems, Karlien Franco, Laura Rosseel, and Melanie Röthlisberger for inviting me to present the results of my research at their lectometry panel in Málaga in 2017; and Michel Westphal, a good friend long before either of us started thinking about syntax or vowel spaces. Most of all, I am thankful for my wife Nancy Mirza, who has stuck it out with me through years of a transatlantic long-distance relationship and, more recently, book-writing induced mood swings and sleepless nights.

Any merits this book might have are in no small part thanks to the people listed above. Any incoherencies, flaws, faults, far-fetched interpretations, spurious correlations and other issues on the following pages are entirely my own.

xvii
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAVE</td>
<td>African American Vernacular English</td>
</tr>
<tr>
<td>AmE</td>
<td>American English</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>AusE</td>
<td>Australian English</td>
</tr>
<tr>
<td>CanE</td>
<td>Canadian English</td>
</tr>
<tr>
<td>CMD</td>
<td>computer-mediated discourse</td>
</tr>
<tr>
<td>EFA</td>
<td>exploratory factor analysis</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a foreign language</td>
</tr>
<tr>
<td>ENL</td>
<td>English as a native language</td>
</tr>
<tr>
<td>ESL</td>
<td>English as a second language</td>
</tr>
<tr>
<td>GloWbE</td>
<td>The Corpus of Global Web-based English</td>
</tr>
<tr>
<td>HKE</td>
<td>Hong Kong English</td>
</tr>
<tr>
<td>ICE</td>
<td>The International Corpus of English</td>
</tr>
<tr>
<td>IDG</td>
<td>indigenous strand</td>
</tr>
<tr>
<td>IndE</td>
<td>Indian English</td>
</tr>
<tr>
<td>IrE</td>
<td>Irish English</td>
</tr>
<tr>
<td>JamE</td>
<td>Jamaican English</td>
</tr>
<tr>
<td>L₁</td>
<td>native language</td>
</tr>
<tr>
<td>L₂</td>
<td>second or foreign language</td>
</tr>
<tr>
<td>LVC</td>
<td>language variation and change</td>
</tr>
<tr>
<td>MD (analysis)</td>
<td>multidimensional analysis</td>
</tr>
<tr>
<td>NZE</td>
<td>New Zealand English</td>
</tr>
<tr>
<td>PCA</td>
<td>principal components analysis</td>
</tr>
<tr>
<td>PhilE</td>
<td>Philippine English</td>
</tr>
<tr>
<td>POS</td>
<td>part of speech</td>
</tr>
<tr>
<td>SgE</td>
<td>Singapore English</td>
</tr>
<tr>
<td>STL</td>
<td>settler strand</td>
</tr>
<tr>
<td>TwlICE</td>
<td>The global English Twitter corpus</td>
</tr>
</tbody>
</table>