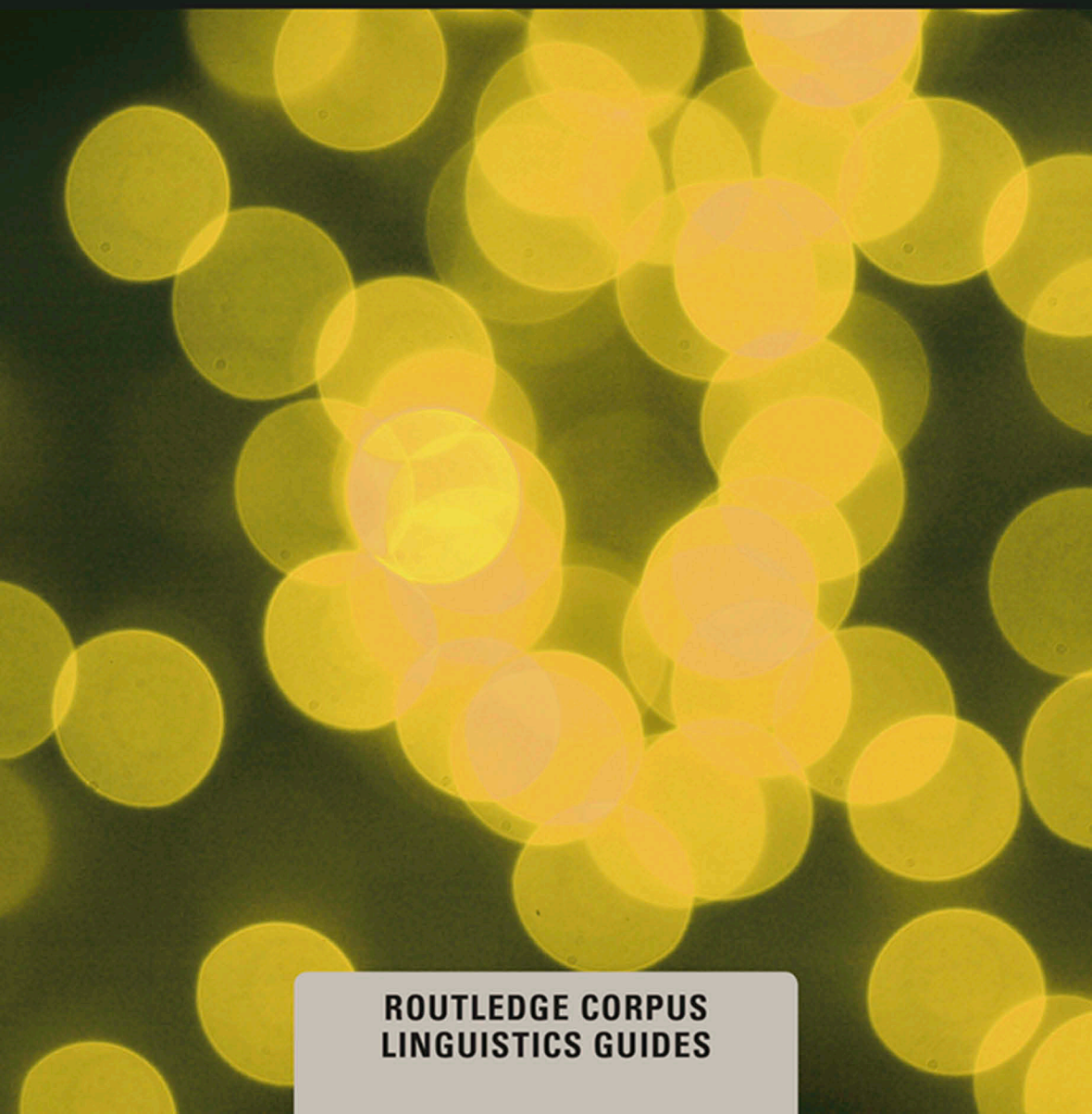
Claudia Lange
and Sven Leuckert

Corpus Linguistics for World Englishes

A Guide for Research



ROUTLEDGE CORPUS LINGUISTICS GUIDES

Corpus Linguistics for World Englishes

Corpus Linguistics for World Englishes offers a detailed account of how to analyse the many fascinating varieties of English around the world using corpus-linguistic methods. Employing case studies for illustration of relevant concepts and methods throughout, this book:

- introduces the theory and practice of analysing World Englishes
- illustrates the basics of corpus-linguistic methods and presents the vast World Englishes corpora
- links World Englishes to Learner Englishes and English as a Lingua Franca
- offers practical, hands-on exercises and questions for discussion in each chapter
- provides helpful overviews and course syllabi for students and instructors.

Corpus Linguistics for World Englishes is key reading for advanced students of English as a World Language and Corpus Linguistics, as well as anyone keen to understand variation in World Englishes with the help of corpus linguistics.

Claudia Lange, Technische Universität Dresden, Germany.

Sven Leuckert, Technische Universität Dresden, Germany.

Routledge Corpus Linguistics Guides

Series consultant: Michael McCarthy

Michael McCarthy is Emeritus Professor of Applied Linguistics at the University of Nottingham, UK, Adjunct Professor of Applied Linguistics at the University of Limerick, Ireland and Visiting Professor in Applied Linguistics at Newcastle University, UK. He is co-editor of the Routledge Handbook of Corpus Linguistics, editor of the Routledge Domains of Discourse series and co-editor of the Routledge Applied Corpus Linguistics series.

Series consultant: Anne O'Keeffe

Anne O'Keeffe is Senior Lecturer in Applied Linguistics and Director of the Inter-Varietal Applied Corpus Studies (IVACS) Research Centre at Mary Immaculate College, University of Limerick, Ireland. She is co-editor of the Routledge Handbook of Corpus Linguistics and co-editor of the Routledge Applied Corpus Linguistics series.

Series co-founder: Ronald Carter

Ronald Carter (1947–2018) was Research Professor of Modern English Language in the School of English at the University of Nottingham, UK. He was also the co-editor of the Routledge Applied Corpus Linguistics series, Routledge Introductions to Applied Linguistics series and Routledge English Language Introductions series.

Routledge Corpus Linguistics Guides provide accessible and practical introductions to using corpus linguistic methods in key sub-fields within linguistics. Corpus linguistics is one of the most dynamic and rapidly developing areas in the field of language studies, and use of corpora is an important part of modern linguistic research. Books in this series provide the ideal guide for students and researchers using corpus data for research and study in a variety of subject areas.

Other titles in this series

Corpus Linguistics for Online Communication
Luke Curtis Collins

Corpus Linguistics for World Englishes
Claudia Lange and Sven Leuckert

More information about this series can be found at www.routledge.com/series/RCLG

Corpus Linguistics for World Englishes

A Guide for Research

Claudia Lange and Sven Leuckert

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2020
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2020 Claudia Lange and Sven Leuckert

The right of Claudia Lange and Sven Leuckert to be identified as authors of this work has been asserted by them in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Lange, Claudia, author. | Leuckert, Sven, 1989- author.
Title: Corpus linguistics for World Englishes : a guide for research / Claudia Lange and Sven Leuckert.
Description: New York : Taylor and Francis, 2020. |
Series: Routledge corpus linguistics guides | Includes bibliographical references and index.
Identifiers: LCCN 2019032148 (print) | LCCN 2019032149 (ebook)
Subjects: LCSH: English language–Globalization. | English language–Variation–English-speaking countries. | English language–Variation–Foreign countries. | Corpora (Linguistics)
Classification: LCC PE1073 .L358 2020 (print) | LCC PE1073 (ebook) | DDC 420.1/88–dc23
LC record available at https://lccn.loc.gov/20190321483
LC ebook record available at https://lccn.loc.gov/2019032149

ISBN: 978-1-138-59341-1 (hbk)
ISBN: 978-1-138-59345-9 (pbk)
ISBN: 978-0-429-48943-3 (ebk)

Typeset in Baskerville
by Swales & Willis, Exeter, Devon, UK

Visit the eResources: www.routledge.com/9781138593459

Contents

List of figures ix
List of tables xi
List of abbreviations xiii
Acknowledgements xv

1 Introduction 1

1.1 What is corpus linguistics? 1
1.2 What are World Englishes? 3
1.3 Who this book is for 3
1.4 How to use this book 4
1.5 Structure of the book 6

2 World Englishes 8

2.1 Introduction 8
2.2 How English became a world language 9
2.3 Conceptualising World Englishes 21
2.4 Making sense of variation 29
2.5 Summary 36
2.6 Exercises 36
2.7 Recommended reading 37
References 39

3 Corpus-linguistic approaches to language 41

3.1 Introduction 41
3.2 Getting started: the basics 42
3.3 Basic corpus design features 47
3.4 Accessing a corpus: concordances 55
3.5 Handling frequencies 57

3.6 Collocations 63
3.7 Regular expressions 64
3.8 Statistics and significance testing 66
3.9 Summary 69
3.10 Exercises 69
3.11 Recommended reading 70
References 72

4 Corpora and World Englishes 74

4.1 Introduction: does size matter? 74
4.2 The 'Brown family' of corpora 76
4.3 The International Corpus of English (ICE) 81
4.4 GloWbE and large (web-based) corpora 95
4.5 Creating your own corpus 98
4.6 Summary 103
4.7 Exercises 104
4.8 Recommended reading 105
List of corpora 106
References 107

5 Tracing variation and change in World Englishes 109

5.1 Introduction 109
5.2 From variation to language change 110
5.3 Variation, mistake, innovation 115
5.4 Contact and the feature pool 121
5.5 Spoken and written language 128
5.6 Summary 131
5.7 Exercises 132
5.8 Recommended reading 132
References 133

6 Interpreting variation and change in World Englishes 135

6.1 Introduction 135
6.2 Researching the verb phrase 136
6.3 Researching the noun phrase 142
6.4 Researching sentence structure 148
6.5 Researching the lexicon 155
6.6 Researching pragmatics 160
6.7 Summary 164
References 165

7 World Englishes, Learner Englishes, and English as a Lingua Franca 168

7.1 Introduction: World Englishes, Learner Englishes, and ELF 168
7.2 Learner corpus design and learner corpora 170
7.3 Comparing World Englishes, Learner Englishes, and ENL 176
7.4 Corpus linguistics and English as a Lingua Franca (ELF) 181
7.5 ELF corpora 182
7.6 Multilingual conversations: code-switching in ELF 187
7.7 Summary 190
7.8 Exercises 190
7.9 Recommended reading 191
List of corpora 192
References 192

8 The state of the art and the way ahead 194

8.1 New Englishes 194
8.2 New corpora 195
8.3 New models 197
8.4 New tools 198
8.5 More statistics 198
8.6 Limitations 199
8.7 Summary 201
References 201

Appendix

Appendix A: Answer key 203
Appendix B: Varieties of English and corpora 208
Appendix C: Suggestions for course uses 210

Index 217

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Figures

2.1 Anglo-Saxon England and its dialect areas (Irvine 2006: 36; reprinted by permission from Oxford University Press) 11
2.2 The spread of English (speakers) around the world (Schneider 2011: 49; reprinted by permission from Cambridge University Press) 15
2.3 Kachru's 'Three Circles' of English (Crystal 2003: 61; reprinted by permission from Cambridge University Press) 23
2.4 The levels of language and linguistic analysis 30
3.1 Screenshot of the *AntConc* interface 45
3.2 Screenshot of the *WordSmith* interface 45
3.3 Components in a register analysis (Biber & Conrad 2009: 6; reprinted by permission from Cambridge University Press) 50
3.4 Hiding tags in *AntConc* 52
3.5 Concordance lines for the word *help* in the conversation files of ICE-Hong Kong 56
3.6 Illustration of the *'Zap'* function in *WordSmith* 57
3.7 Frequency of hapax legomena in Herman Melville's *Moby Dick* (https://commons.wikimedia.org/wiki/File:Moby_Dick_Words.gif; public domain) 62
3.8 Example of a regular expression in *AntConc* 65
3.9 Creative word-formation in spoken and written first-language (L1) and second-language (L2) varieties (Biermeier 2014: 329; reprinted by permission from John Benjamins) 67
4.1 Functions of *Sketch Engine* 79
4.2 Some tokens of *yeah* produced by extra-corpus speakers in the direct conversations in ICE-Hong Kong 89
4.3 Search results for the word *furnitures* in GloWbE 96
5.1 Acquisition of English in multilingual Outer Circle contexts (based on Thomason 2001: 75) 118
5.2 Sources and processes leading to PCEs (Schneider 2007: 100; reprinted by permission from Cambridge University Press) 122

5.3 Realisation of the variable (r) first in casual and then in emphatic speech by shop assistants in *Saks* (S), *Macy's* (M), and *Klein* (K) (Labov 2006: 48; reprinted by permission from Cambridge University Press) 132
6.1 *Madam & Eve* (from https://twitter.com/madamevecartoon/status/426232112824328192; reprinted by permission from the creators) 155
7.1 Countries and regions included in ICNALE (map created with mapchart.net; CC Attribution-SA 4.0 International Licence) 175
7.2 Frequency comparison of PVs with *up* between NEs, LEs, and ENL (Gilquin 2015: 105; reprinted by permission from John Benjamins) 178
7.3 Frequency comparison of PVs with *up* in spoken and written NEs, LEs, and ENL (Gilquin 2015: 106; reprinted by permission from John Benjamins) 179
7.4 VOICE Online with POS-tagging (www.univie.ac.at/voice/images/help/10_lemma_popup.jpg) 183
7.5 ASEAN+3 countries in the Asian Corpus of English (ACE) (https://de.wikipedia.org/wiki/ASEAN_Plus_Three#/media/Datei:ASEAN_Plus_Three_members.png; CC BY-SA 3.0 Licence) 185
A.1 Parsing of 'Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo' (from https://commons.wikimedia.org/wiki/File:Buffalo_sentence_1_parse_tree.svg; public domain) 205

Tables

2.1	The productive inflectional endings of English	33
2.2	Classifying varieties of English	36
2.3	Postcolonial Englishes and their position in the Dynamic Model (Schneider 2007)	38
3.1	Annotation examples from the *International Corpus of English*	51
3.2	Frequency of *dicit* and *dixit* in the Gospel of Matthew and the Gospel of John (www.lancaster.ac.uk/fss/courses/ling/corpus/Corpus3/3SIG.HTM)	68
3.3	Data for adverbs formed with and without *-ly*	70
4.1	Overview of corpora included in the 'Brown family'	77
4.2	Text types included in the Brown corpus (adapted from www.helsinki.fi/varieng/CoRD/corpora/BROWN/basic.html)	78
4.3	Structure of the ICE corpora (adapted from www.ice-corpora.uzh.ch/en/design.html)	82
4.4	Available ICE components and ICE components currently being compiled (as of June 2019)	84
4.5	Mark-up symbols used in the spoken sections of the ICE corpora (adapted from the ICE manual available at www.ice-corpora.uzh.ch/en/manuals.html)	86
4.6	Additional mark-up symbols used in the written sections of the ICE corpora (adapted from the ICE manual available at www.ice-corpora.uzh.ch/en/manuals.html)	92
4.7	Countries included in GloWbE and abbreviation and word count for each country (adapted from Davies & Fuchs 2015: 6; Loureiro-Porto 2017: 450)	96
4.8	Major steps in corpus compilation and potential issues	99
6.1	Frequencies of PVs normalised to 1 million words and rounded (Schneider 2004: 236; reprinted by permission from Wiley)	139
6.2	Range of meanings of selected PVs (meaning types/tokens) (Schneider 2004: 236; reprinted by permission from Wiley)	140

6.3 Definite articles across all text types and corpora normalised per 1,000 words (adapted from Sand 2004: 287; reprinted by permission from Wiley) 146
6.4 Inversion in main clause interrogatives in ICE-India (Hilbert 2008: 273; reprinted by permission from John Benjamins) 151
6.5 Inversion in indirect questions with constituent interrogatives in ICE-India and ICE-Singapore, relative frequencies (in %) (Hilbert 2008: 276; reprinted by permission from John Benjamins) 151
6.6 Verb types in inverted embedded constituent interrogatives (Hilbert 2011: 133; reprinted by permission from John Benjamins) 153
6.7 Verb types in inverted embedded polar interrogatives (Hilbert 2011: 134; reprinted by permission from John Benjamins) 154
6.8 Varieties covered in Biermeier's (2014) study on prefixation and compounding in World Englishes 158
6.9 Corpora, word count, and corresponding circle of the included varieties in Aijmer (2018) (Aijmer 2018: 109; reprinted by permission from Brill Rodopi) 162
6.10 Frequencies of *very*, *really*, and *so* per one million words in AmE, BrE, NZE, and SinE (adapted from Aijmer 2018: 112) 163
7.1 Learner-related and task-related variables in learner corpus design (see Granger 2008: 264) 171
7.2 Learner and task variables included in ICLE and ICLEv2 (adapted from https://cdn.uclouvain.be/public/Exports%20reddot/cecl/documents/LEARNER_PROFILE.txt) 172
7.3 Structure of ICNALE (adapted from http://language.sakura.ne.jp/icnale/index.html#6) 174
7.4 Corpora used in Gilquin's (2015) study of phrasal verbs with *up* (adapted from Gilquin 2015: 104; reprinted by permission from John Benjamins) 177
7.5 Error annotation task 191
A.1 Classifying varieties of English (solutions) 204
A.2 Data for adverbs formed with and without *-ly* 205
A.3 Error annotation task with solutions 207
B.1 Availability of corpora for different World Englishes 208
C.1 Syllabus using *Corpus Linguistics for World Englishes* 211

Abbreviations

ACE	*Asian Corpus of English/Australian Corpus of English*
AdvP	adverb phrase
AmE	American English
ANC	African National Congress
AP	adjective phrase
ASEAN	*Association of Southeast Asian Nations*
AusE	Australian English
BNC	*British National Corpus*
BrE	British English
CECL	Centre for English Corpus Linguistics
CED	*Collins English Dictionary*
CEFR	*Common European Framework of Reference for Languages*
CMC	Computer-Mediated Communication
COCA	*Corpus of Contemporary American English*
COHA	*Corpus of Historical American English*
CoRD	*Corpus Resource Database*
EFL	English as a Foreign Language
EIC	East India Company
EIF	Extra- and Intraterritorial Forces model
ELF	English as a Lingua Franca
EModE	Early Modern English
ENL	English as a Native Language
ESL	English as a Second Language
eWAVE	*electronic World Atlas of Varieties of English*
F-LOB	*Freiburg-LOB Corpus of British English*
FROWN	*Freiburg-Brown Corpus of American English*
GA	General American English
GloWbE	*Corpus of Global Web-based English*
HiCE Ghana	*Historical Corpus of English in Ghana*
ICE	*International Corpus of English*
ICLE	*International Corpus of Learner English*
ICNALE	*International Corpus Network of Asian Learners of English*

IDG	indigenous strand
IndE	Indian English
IPA	International Phonetic Alphabet
IrE	Irish English
KWIC	Keyword in Context
L1	first language
L2	second language
LD	left-dislocation
LEs	Learner Englishes
LINDSEI	*Louvain International Database of Spoken English Interlanguage*
LOB	*Lancaster-Oslo/Bergen Corpus*
LOCNEC	*Louvain Corpus of Native English Conversation*
ME	Middle English
NEs	New Englishes
NP	noun phrase
NZE	New Zealand English
ODE	*Oxford Dictionary of English*
OE	Old English
OED	*Oxford English Dictionary*
PCE	Postcolonial English
PDE	Present-Day English
POS	parts-of-speech
PP	prepositional phrase
PV	particle verb/phrasal verb
RD	right-dislocation
RegEx	regular expression
RP	Received Pronunciation
SAfE	South African English
SAR	Special administrative region
SAVE	*South Asian Varieties of English Corpus*
SBCAE	*Santa Barbara Corpus of American English*
SinE	Singapore English
SLA	Second Language Acquisition
SLE	Sri Lankan English
STL	settler strand
TOEFL	*Test of English as a Foreign Language*
ViMELF	*Corpus of Video-Mediated English as a Lingua Franca Conversations*
VOICE	*Vienna-Oxford International Corpus of English*
VP	verb phrase
WWI	World War One
WWII	World War Two

Acknowledgements

Celebrating and investigating the diversity of World Englishes is always an inspiring endeavour for us, which is why we are particularly happy to be the authors of a book that does exactly that. Since no book is ever written and published without the help of others, we would like to thank everybody who has contributed to this book in one way or another.

In particular, we would like to thank Julia Kruse for reading some chapters in detail, and all of our students who commented on parts of this book. We also thank Marie-Louise Brunner, Stefan Diemer, Beke Hansen, and Louise Mycock for their contributions. Furthermore, we would like to express our gratitude to Anne O'Keeffe and Mike McCarthy for giving us the opportunity to contribute to the *Routledge Corpus Linguistics Guides* series. We also thank Lizzie Cox, Ruth Berry, Isabella Ritchie, and everybody at Routledge for assisting us in the creation of this book.

Thanks also go to our friends and families who gave us the time to write this book.

Finally, we would like to thank the following publishers and copyright holders for allowing us to reprint their material:

Figure from Irvine (2006), Chapter 2; reprinted by permission from Oxford University Press.

Figures from Schneider (2011), Chapter 2; Biber & Conrad (2009), Chapter 3; Labov (2006), Chapter 4; Schneider (2007), Chapter 5; all reprinted by permission from Cambridge University Press.

Figures and tables from Biermeier (2014), Chapter 3; Hilbert (2008) and Hilbert (2011), Chapter 6; Gilquin (2015), Chapter 7; all reprinted by permission from John Benjamins.

Tables from Sand (2004) and Schneider (2004), Chapter 6; reprinted by permission from Wiley.

Table from Aijmer (2018), Chapter 6; reprinted by permission from Brill Rodopi.

Cartoon "Madam & Eve", Chapter 6; reprinted by kind permission from the creators.

Map of countries included in ICNALE, Chapter 7, created with mapchart.net; reprinted under the Creative Commons Attribution-ShareAlike 4.0 International Licence – we thank the creator of this website for this great tool.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 1

Introduction

As a citizen of the 21st century, you are not only aware of the existence of computers and of the spread of English around the world, you probably take both for granted (for better or worse). Wherever we live and whatever our first language happens to be, a lot of our online interactions will happen in English – online shopping, gaming, social media, and so on. Yet English and the computer are more than mere tools for international communication: we can use the considerable processing power and storage capacity of an average personal computer to help us investigate global English. In other words, we can do corpus-linguistic research on World Englishes. This chapter offers a first glimpse at the topics covered in this book. It gives an insight into the different kinds of English, what it means to do corpus linguistics, and how we can study English using so-called 'corpora' – that is, collections of spoken and written texts. In addition, the chapter introduces you to the structure as well as to the key features of the book.

1.1 What is corpus linguistics?

You might be familiar with linguistics, the study of the forms and functions of language. Linguists are interested in the unity and diversity of language in general, or of a particular language. They typically do not pass value judgments: You know you have run into a linguist when you say something like, 'Ain't nobody home' or 'She home already', and their reaction is, 'Interesting, where are you from?' What might be considered 'slang', or 'bad English', is food for thought for the linguist because it represents *variation*: different ways of saying the same thing. We will have more to say about variation later, but even at this stage you are likely to agree that variation in language is very often not random, but shows specific patterns. A *biscuit* in the United Kingdom is a *cookie* in the United States, to take just one simple example of regional variation. Linguists are then interested in finding patterns, classifying them, and ideally also explaining them.

Let us stick with this example of lexical differences between British and American English, the two most prominent 'standard' varieties of English (more about standard languages in Chapter 5). You might wonder whether

it is really true that the *Harry Potter* books were slightly edited for the American market to make life easier for readers who are familiar with *cookies*, but not necessarily with *crumpets*. How are you going to find out for yourself (rather than turning to the internet, where fans have already compiled lists)? You could take the printed books, place the British and American editions parallel on your desk, and start reading and taking notes of the differences. If you intend to do a thorough job and tackle all the books it will take you some time. If, however, you have the electronic editions of the books, you can make the computer do the work for you. You can use any editing program to create word lists, and with a dedicated 'concordance program' (more about that in Chapter 3.1) you have an astounding range of tools at your disposal to analyse your *Harry Potter* corpus. A corpus is basically a collection of texts stored in an electronic, machine-readable format. Chapter 4 tells you more about corpora in the field of World Englishes and deals specifically with issues of size and content. In our example, we are working with a relatively small, single-genre, purpose-built corpus designed to answer a specific research question. To be precise, we are working not only with a single-genre, but also a single-author corpus, quite unlike the corpus projects which are the topic of Chapter 4.

Notions like genre, text type, or register play a big role in corpus linguistics. Some scholars in the field use these terms more or less interchangeably, while others ascribe nuanced differences to them. Right now, we just need to be aware that the corpus-linguistic idea of a text differs from the everyday understanding of the term in that a text need not be something printed, or even written; it can also be a transcript of a conversation. A second point to note, which is related to the concept of text type, is that of the representativeness of corpora, a tricky issue for corpus compilers. If you want to move beyond your *Harry Potter* corpus and work with a corpus that represents your variety of English faithfully, what kind of text types should be included? Think about the kinds of texts that you actively produce, and those that you passively consume, over the course of a week or month. You probably have lots of conversations, but in different degrees of formality; you will write some e-mails, post some comments on social media, draft a term paper. You might listen to song lyrics and the news on TV, you might read a novel, read and sign a contract. Your friends' and your parents' production and reception of text types are probably quite different from your linguistic experience, which already indicates the difficulty in coming up with a representative corpus for, say, Australian English – and we have not even talked about how to lay our hands on all those text types. Still, linguists have found some fairly solid solutions to these problems, as we will see in Chapter 4.

You now have a working knowledge of what a corpus is and what it might contain. Corpus linguistics obviously has to start with corpora, but whether corpus linguistics is more like a toolbox or a separate discipline is already a matter of debate – perhaps you might want to return to this question once you have worked your way through this book.

1.2 What are World Englishes?

In the beginning of this chapter, we alluded to the fact that English is today's most important global language. Billions of people interact with and in English on a daily basis, but to different degrees. You might have grown up with English as your only language or as one of the multiple languages spoken in your home; you might have acquired English as a second language at school in a country where it has an official status; you might have learnt English at school or later in life, but you only need it in international communication and largely with others for whom it is likewise a foreign language. Many people around the world also simply pick up English as they go, in order to function in specific contexts such as tourism.

The label 'World Englishes' applies to all of these contexts. Some varieties of English are older, more prestigious, more widespread, and more stable than others, but all display the high degree of creativity that goes into people's communicative interactions. A traditional and well-known subdivision is the one between English as a Native Language (ENL), English as a Second Language (ESL), and English as a Foreign Language (EFL). We will see how this threefold distinction shows up in different ways of categorising Englishes around the world. It is also important to realise that monolingual native speakers of English are a minority within the English-speaking world – many former British colonies in Asia and Africa retained English as one of their official languages after independence. In these countries, English serves both an intranational and an international function. Speakers in these postcolonial countries are typically multilingual and may use English as a link language amongst themselves. While language is constantly developing, the interaction of English with the local languages has become a particularly interesting and dynamic field to study.

The influence of English as a global language does not stop there. English is the preferred foreign language for millions of people worldwide. International institutions and companies use it as their working language. The term 'English as a Foreign Language' (EFL) stresses the learner perspective. The label 'English as a Lingua Franca' (ELF) focuses on how people who do not share a mother tongue use English as a resource for effective communication. Both approaches are taken up again in Chapter 7.

1.3 Who this book is for

If you cherish variation in English and want to get to the bottom of it, this book is for you. It will help you gain access to a vibrant field in linguistics and will equip you with the basic knowledge and the tools to embark on your own projects. We believe this book can be used in many different ways. In this section, we focus on who this book is written for rather than who it is *not* written for and provide brief ideas for usage scenarios.

Undergraduate and graduate students

If you are interested in how corpus linguistics can fruitfully be applied to World Englishes, this book is the right choice for you: we introduce all important terms and illustrate new concepts and methods using examples, figures, and exercises. The only things we suggest you bring along is an interest in the English language and its many forms as well as a computer with access to the internet. The theory and methods taught by the book can be applied to term papers and final theses, which means that undergraduate and graduate students alike can benefit from the book. Furthermore, this book can be used in conjunction with or in addition to other titles published in the *Routledge Corpus Linguistics Guides* series.

University instructors

This book is designed in such a way that the chapters can be read one after the other on a weekly basis; exercises follow right at the end of each chapter. Thus, individual sections or entire chapters can be read in class or as part of the students' homework. The same goes for the exercises, which can be set as tasks in the classroom or just discussed after they have been assigned as homework. As a special component of this book, we designed a course syllabus including a description of weekly activities on the basis of the chapters and exercises. A template of the syllabus is also available on the book's accompanying eResources and can easily be modified to fit different needs. The example syllabus can be found in Appendix C.

Researchers

For a researcher in linguistics or adjacent fields, this book can serve both as a reference guide to fill individual gaps in corpus linguistics and/or World Englishes or as an introduction if you are not familiar with either of the two fields. The book's chapters can easily be complemented and built upon by referring to the literature suggested in the 'Recommended reading' sections.

1.4 How to use this book

In this section, we shed light on some key components of this book: *Key terms; In-text exercises and questions; Chapter exercises; Open exercises and discussion questions; and Recommended reading*. Please note that not every chapter features each of these components. Chapter 6, for instance, contains many exercises spread across the text, which is why there are no additional exercises at the end.

Key terms

Each chapter from 2 to 7 begins with a list of key terms. These terms are not only of particular relevance for their respective chapter (and, in some cases, the rest of the book) but they also represent terms and concepts that you should know and understand if you want to succeed in your corpus-linguistic endeavours. In order to make finding these terms in the chapter easier, they are printed in bold in the running text either when they are first mentioned or when we provide a definition for them.

In-text exercises and questions

Throughout this book, we included in-text exercises and questions. You can recognise them by paying attention to the following symbols:

⇒ Symbol for an in-text exercise or question,
◊ Symbol for the corresponding answer or comment.

These sections give you an opportunity to actively try out what you just read and an opportunity to think about issues related to the chapter you are working on. Almost every in-text exercise or question is immediately followed by an answer or a comment, so that you can compare your thoughts and ideas to our suggested answer. We do not include an answer if the question merely serves to familiarise you with or to let you explore a topic or an idea. Most of the exercises in this format are relatively short and do not take up too much of your time; others are a bit more complex – how you want to approach them depends on how much time you are willing to invest.

Chapter exercises

More complex tasks are featured at the end of most chapters. They give you the opportunity to apply your knowledge and work on exercises that, at times, require several steps to be solved. An answer key with our comments on how to approach the exercises as well as a suggestion for a solution is given at the end of the book. Since not all exercises feature answers (see the next section on open exercises and discussion questions), those that do are marked with an asterisk. The following is an example of a chapter exercise with a corresponding comment at the end of the book:

*Use the *Google n-gram viewer* (available at https://books.google.com/ngrams) to trace word pairs that are associated with BrE and AmE respectively over time, such as

- *cookie – biscuit*
- *sidewalk – pavement*
- or a grammatical example: *have gotten – have got*

Change the setting from English/general to BrE and AmE separately and take note of any spikes or overlaps in usage.

Open exercises and discussion questions

In addition to chapter exercises that come with answers at the end of the book, there are also some open exercises and questions. These are designed to let you explore topics discussed in a chapter but do not necessarily have a fixed answer, which is why you will not find any commentary in the appendix. These questions can also be used as a basis for discussion in a seminar context.

Recommended reading

The last section of each chapter contains 'Recommended reading'. Rather than giving you a list of titles, these sections provide you with essential reading related to the topic of the chapter as well as comments on the included titles. This should be helpful, for instance, to decide whether a book is the right fit for you, since some books are geared more towards beginners and others more towards advanced researchers. The readings reflect our personal choices, but also our experience in research and teaching from the undergraduate to the postgraduate level.

1.5 Structure of the book

Out of the many ways of researching World Englishes, this book focuses on corpus-linguistic approaches. One overwhelming advantage is that everyone with access to a computer can do it – with so much free software available and so many accessible corpora around, you have a world of resources at your fingertips. This book proceeds by first introducing World Englishes and corpus linguistics individually and then gradually joining the two fields.

Chapter 2 opens up the incredibly fascinating field of World Englishes by answering some important questions, such as: How did English become a, perhaps *the*, major global language? How can we meaningfully group different forms of English world-wide? And, finally, how can we make sense of and study variation in World Englishes? The chapter begins with a historical perspective and gradually moves into the present.

Chapter 3 introduces corpus linguistics. After an introduction to some technical basics of doing corpus-linguistic research, the important terms and concepts of the field are introduced. In addition, general concerns relevant to conducting linguistic research are addressed: we stress the importance of the research question and ask you to consider the quality of any data source. Throughout the chapter, hands-on exercises give you the opportunity to apply what you have just read.

Bringing together what you have learned in the previous two chapters, **Chapter 4** provides an overview of available corpora of World Englishes and the means of accessing them, the way they are structured, the kinds of text that went into them, and the additional information that might come with them. 'Knowing your corpus' is a basic prerequisite for evaluating your own and others' corpus findings. Since creating your own corpus without even leaving your screen is much easier than you might think, the chapter will also give you an idea of how to do it.

Once you have worked your way through Chapters 3 and 4, you are familiar with a wide range of corpora and know how to retrieve data from them. So how do you interpret your findings? **Chapter 5** tells you more about variation and change with special reference to World Englishes. Languages and language use change all the time, and linguists have come up with different approaches towards explaining how variation turns into change. We consider how variation in language can acquire social meaning, for example as a marker of identity. You will also learn more about the role of language contact in shaping the Englishes spoken in multilingual environments.

Chapter 6 offers descriptions of corpus-linguistic case studies and has two main purposes: It serves to introduce you to the manifold possibilities of researching World Englishes using corpus-linguistic methods and, in addition, invites you to consider how previous studies can be used as a basis for your own research projects.

Chapter 7 moves beyond traditional approaches to World Englishes by considering 'Learner Englishes' and 'English as a Lingua Franca'. English as produced by learners has long been treated as separate from World Englishes. Over the last decade, attempts at integrating and comparing these two types of English have experienced a revival. The chapter introduces you to important aspects of designing a learner corpus, provides an in-depth look at a case study, and also considers corpora and studies involving English as a Lingua Franca, that is, English used as a link language for people who do not share another language.

Chapter 8, the final chapter of the book, provides an overview of the state of the art in this field and the way ahead. In particular, the chapter offers a perspective on many exciting current developments: with the development of new corpora, new tools, new models, and the emergence of new Englishes, further research directions, and more complex statistics, corpus linguistics and World Englishes remain at the forefront of English linguistics.

We have thoroughly enjoyed sharing our experience of World Englishes as teachers and researchers with you, and we hope you will also enjoy your computer-assisted journey around the English-speaking world!