The Cambridge Handbook of English Corpus Linguistics

The Cambridge Handbook of English Corpus Linguistics (*CHECL*) surveys the breadth of corpus-based linguistic research on English, including chapters on collocations, phraseology, grammatical variation, historical change, and the description of registers and dialects. The most innovative aspects of the *CHECL* are its emphasis on critical discussion, its explicit evaluation of the state of the art in each subdiscipline, and the inclusion of empirical case studies. While each chapter includes a broad survey of previous research, the primary focus is on a detailed description of the most important corpus-based studies in this area, with discussion of what those studies found, and why they are important. Each chapter also includes a critical discussion of the corpus-based methods employed for research in this area, as well as an explicit summary of new findings and discoveries.

DOUGLAS BIBER is Regents' Professor of Applied Linguistics in the English Department at Northern Arizona University.

RANDI REPPEN is Professor of Applied Linguistics in the English Department at Northern Arizona University.

The Cambridge Handbook of English Corpus Linguistics

Edited by
Douglas Biber
and
Randi Reppen
Northern Arizona University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107037380

First published 2015

Printing in the United Kingdom by TJ International Ltd. Padstow Cornwall

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

Library of Congress Cataloguing in Publication data
The Cambridge handbook of English corpus linguistics / edited by Douglas Biber and Randi Reppen
Northern Arizona University.
p. cm. – (Cambridge handbooks in language and linguistics)
Includes bibliographical references and index.
ISBN 978-1-107-03738-0 (hardback)
1. Corpora (Linguistics) 2. Linguistic analysis (Linguistics) 3. Computational linguistics. I. Biber, Douglas, editor. II. Reppen, Randi, editor.
P128.C68C46 2015
420.1′88–dc23

2014035299

ISBN 978-1-107-03738-0 Hardback

We dedicate this handbook to the memory of Geoffrey Leech, one of the true pioneers of English corpus linguistics. Geoff contributed seminal research to all aspects of corpus linguistics, from corpus construction to corpus tagging and annotation, to the evaluation of corpus representativeness. But from our point of view, Geoff's most important contributions involved the application of corpus-based analysis to investigate linguistic research questions about language variation and use. Geoff contributed to nearly all major subdisciplines of English linguistics with such analyses, including pragmatics, stylistics, descriptive grammar, dialect differences, and historical change. For us personally, Geoff was a model, mentor, co-author, and friend. He will be sorely missed.

Contents

Figures

Tables

Major corpora cited in the handbook

ARCHER	*A Representative Corpus of Historical English Registers*
BNC	*British National Corpus*
Brown	*Brown Corpus*
CED	*Corpus of English Dialogues*
CEEC	*Corpus of Early English Correspondence*
CIC	*Cambridge International Corpus*
COCA	*Corpus of Contemporary American English*
COHA	*Corpus of Historical American English*
COLT	*Bergen Corpus of London Teenager Language*
ELFA	*Corpus of English as a Lingua Franca in Academic Settings*
EMEMT	*Early Modern English Medical Texts*
FLOB	*Freiburg–LOB Corpus*
FRED	*Freiburg English Dialect Corpus*
FROWN	*Freiburg–Brown Corpus*
GloWbE	*Corpus of Global Web-based English*
HKCSE	*Hong Kong Corpus of Spoken English*
ICE	*International Corpus of English*
ICLE	*International Corpus of Learner English*
LINDSEI	*Louvain International Database of Spoken English Interlanguage*
LLC	*London–Lund Corpus*
LOB	*Lancaster–Oslo–Bergen Corpus*
LOCNEC	*Louvain Corpus of Native English Conversation*
LSWC	*Longman Spoken and Written English Corpus*
LWP	*Language in the Workplace Project*
MICASE	*Michigan Corpus of Spoken Academic English*
MICUSP	*Michigan Corpus of Upper-level Student Papers*

NHCC	*Nottingham Health Communication Corpus*
T2KSWAL	*TOEFL 2000 Spoken and Written Academic Language Corpus*
Time	*Time Magazine Corpus of American English*
VOICE	*Vienna–Oxford International Corpus of English*
WrELFA	*The Written Corpus of English as a Lingua Franca in Academic Settings*

Contributors

Silvia Bernardini, Associate Professor of English Language and Translation, University of Bologna
Douglas Biber, Regents' Professor, Applied Linguistics, Northern Arizona University
Alex Boulton, Professor, English and Applied Linguistics, Université de Lorraine
Ray Carey, Researcher, ELFA Project, University of Helsinki
Winnie Cheng, Professor, English language, Hong Kong Polytechnic University
Brian Clancy, Researcher, Mary Immaculate College, University of Limerick
Thomas Cobb, Associate Professor, Applied Linguistics, Université du Québec à Montréal, Canada
Susan Conrad, Professor, Applied Linguistics, Portland State University
Jonathan Culpeper, Professor, Linguistics and English Language, Lancaster University
Mark Davies, Professor, Linguistics, Brigham Young University
Jane Demmen, Research Fellow, Linguistics and Modern Languages, University of Huddersfield
Gaëtanelle Gilquin, Research Associate, FNRS, Centre for English Corpus Linguistics, Université catholique de Louvain
Sylviane Granger, Professor, Centre for English Corpus Linguistics, Université catholique de Louvain
Bethany Gray, Assistant Professor, Applied Linguistics and Technology, Iowa State University
Stefan Th. Gries, Professor, Linguistics, University of California, Santa Barbara
Jack Grieve, Lecturer, Forensic Linguistics, Aston University
Martin Hilpert, Assistant Professor, English Linguistics, University of Neuchâtel

Marianne Hundt, Professor and Chair, English Language and Linguistics, University of Zurich
Susan Hunston, Professor, English Language and Applied Linguistics, University of Birmingham
Ken Hyland, Chair, Professor of Applied Linguistics and Director of the Center for Applied English Studies, University of Hong Kong
Daniela Kolbe-Hanna, Assistant Professor, English Studies, University of Trier
Merja Kytö, Professor, English Language, Uppsala University
Geoffrey Leech, Professor Emeritus, Linguistics and English Language, Lancaster University
Michaela Mahlberg, Professor, English Language and Linguistics, University of Nottingham
Christian Mair, Professor and Chair, English Linguistics, University of Freiburg
Anna Marchi, Linguistics and English Language, Lancaster University
Ron Martinez, Assistant Professor of English, Federal University of Paraná
Anna Mauranen, Professor of English Philology and Vice Rector, University of Helsinki
Fanny Meunier, Professor, Centre for English Corpus Linguistics, Université catholique de Louvain,
Anne O'Keeffe, Senior Lecturer, Applied Linguistics and English Language Teaching, Mary Immaculate College, University of Limerick
Magali Paquot, Postdoctoral Researcher, FNRS, Centre for English Corpus Linguistics, Université catholique de Louvain
Alan Partington, Associate Professor, Linguistics, University of Bologna
Elina Ranta, Researcher, ELFA Project University of Helsinki, University of Tampere
Paul Rayson, Director, University Centre for Computer Corpus Research on Language (UCREL), Lancaster University
Randi Reppen, Professor, Applied Linguistics and TESL, Northern Arizona University
Norbert Schmitt, Professor, Applied Linguistics, University of Nottingham
Erik Smitterberg, Researcher, English Language, Uppsala University
Shelley Staples, Assistant Professor, English, Purdue University
Benedikt Szmrecsanyi, Odysseus Research Professor, Linguistics, KU Leuven
Irma Taavitsainen, Professor, English Philology, University of Helsinki
Richard Xiao, Reader, Linguistics and English Language, Lancaster University