Corpora in Applied Linguistics
THE CAMBRIDGE APPLIED LINGUISTICS SERIES

Series editors: Michael H. Long and Jack C. Richards

This series presents the findings of recent work in applied linguistics which are of direct relevance to language teaching and learning and of particular interest to applied linguists, researchers, language teachers, and teacher trainers:

In this series:
The Development of Second Language Proficiency edited by Birgit Harley, Patrick Allen, Jim Cammins and Merrill Swain
Cognition in Second Language Instruction edited by Peter Robinson
Corpora in Applied Linguistics by Susan Hunston
Computers in Second Language Acquisition by Carol A. Chapelle
Contrastive Rhetoric by Ulla Conner
Evaluating Second Language Education edited by J. Charles Alderson and Alan Beretta
Exploring the Second Language Mental Lexicon by David Singleton
Focus on Form in Classroom Second Language Acquisition edited by Catherine Doughty and Jessica Williams
Genre Analysis by John M. Swales
Immersion Education: International Perspectives edited by Robert Keith Johnson and Merrill Swain
Interactive Approaches to Second Language Reading edited by Patricia L. Carrel, Joanne Devine and David E. Eskey
Interfaces between Second Language Acquisition and Language Testing Research edited by Lyle F. Bachman and Andrew D. Cohen
Language Learning and Deafness edited by Michael Strong
Language Program Evaluation by Brian K. Lynch
Language Transfer by Terence Odlin
The Learner-Centred Curriculum by David Nunan
Learning Strategies in Second Language Acquisition by J. Michael O’Malley and Unna Ulh Chamot
Learning Vocabulary in Another Language by I.S.P. Nation
Linguistic Perspective on Second Language Acquisition edited by Susan M. Gass and Jaquelyn Schachter
Network-based Language Teaching edited by Mark Warschauer and Richard Kern
Perspectives on Pedagogical Grammar edited by Terence Odlin
Research Perspectives on English for Academic Purposes edited by John Flowerdew and Matthew Peacock
Researching and Applying Metaphor edited by Lynne Cameron and Graham Low
Second Language Vocabulary Acquisition edited by James Coady and Thomas Huckin
Second Language Writing edited by Barbara Kroll
Sociolinguistics and Language Teaching edited by Sandra Lee McKay and Nancy H. Hornberger
Teacher Cognition in Language Teaching by Devon Woods
Text, Role and Context by Ann M. Johns
Corpora in Applied Linguistics

Susan Hunston

University of Birmingham
For Doreen and Peter Hunston
Contents

Series editors’ preface ix
Acknowledgements xi

1 Introduction to a corpus in use
What this book is about 1
What a corpus can do 3
What corpora are used for 13
Types of corpora 14
Some key terms 16
Why corpora? Why not? 20
Conclusion 23
Note on sources of examples 24

2 The corpus as object: Design and purpose
Issues in corpus design 25
Corpus, text and language 32
Conclusion 37

3 Methods in corpus linguistics: Interpreting concordance lines
Introduction 38
Searches, concordance lines and their presentation 39
What is observable from concordance lines? 42
Coping with a lot of data: using phraseology 52
Using a wider context: observing hidden meaning 56
Using probes 62
Issues in accessing and interpreting concordance lines 65

4 Methods in corpus linguistics: Beyond the concordance line
Frequency and key-word lists 67
Collocation 68
Tagging and parsing 79
Other kinds of corpus annotation 86
Competing methods 92
## Contents

5 *Applications of corpora in applied linguistics*  
  - Dictionaries and grammars  
  - Studying ideology and culture  
  - Translation  
  - Stylistics  
  - Forensic linguistics  
  - Help for writers  
  - Conclusion  
  - 96

6 *Corpora and language teaching: Issues of language description*  
  - Language as phraseology  
  - Language variation  
  - Conclusion  
  - 137

7 *Corpora and language teaching: General applications*  
  - Data-driven learning  
  - Reciprocal learning and parallel concordances  
  - Corpora and language teaching methodology  
  - Corpora and syllabus design  
  - Challenges to the use of corpora in language teaching  
  - Conclusion  
  - 170

8 *Corpora and language teaching: Specific applications*  
  - Corpora and EAP  
  - Corpora and language testing  
  - The evidence of learner corpora  
  - Conclusion  
  - 198

9 *An applied linguist looks at corpora*  
  - List of relevant web-sites  
  - References  
  - Index  
  - 213

- 217
- 218
- 233
Series editors’ preface

The use and uses of linguistic corpora in applied linguistics have expanded rapidly over the past 20 years, in part because of the advent of improved, more accessible systems of electronic storage and analysis, and also because of an ever-growing appreciation of the huge potential of corpus work. Among the pioneers in this effort has been the team in the English Department at the University of Birmingham, led for many years by Professor John Sinclair, and known among other things for its prolific publications surrounding the COBUILD Project. Dr Susan Hunston has long been a core member of that team, including several years as a Senior Grammarians with COBUILD, and has steadily become one of corpus analysis’ most widely recognised experts. In Corpora in Applied Linguistics, she provides an original and authoritative introduction to modern uses of corpus studies, and does so in a clear fashion that will appeal to a wide range of readers.

After a lucid introduction to the scope and limitations of modern corpus work, including research on frequency, phraseology, and collocation, Dr Hunston defines some of the basic issues surrounding the construction and evaluation of corpora in chapter 2. These include questions concerning the size, content, representativeness, and stability of corpora over time, and such matters as the relative merits of corpora designed as collections of texts and as collections of language samples, in turn reflecting how and for what purposes corpora are investigated. Subsequent chapters offer detailed coverage of methods of analysis in corpus linguistics, and of applications of corpus-based research in a variety of domains – most centrally in the design of syllabi and materials for language teaching, in English for specific purposes, and language testing, but also in dictionary work, the study of ideology and culture, translation, stylistics, forensic linguistics, and the provision of on-line assistance for writers in well-defined technical domains. Key terms and concepts are clearly explained, copious examples of corpus research are presented
succinctly, and statistical analyses are handled in such a way as to be easily comprehensible to expert and non-expert alike.

Corpora in Applied Linguistics should be of great value to language teachers, syllabus designers, materials writers, and language testers, as well as to applied linguists involved in the kinds of work mentioned above. Offering an overview and searching analysis of the current state-of-the-art in applied corpus work, and based on years of first-hand experience in language teaching and corpus research, Dr Hunston’s book is a most welcome and valuable addition to the Cambridge Applied Linguistics Series.

Michael H. Long
Jack C. Richards
Acknowledgements

Most of the corpus examples in this book are taken from the Bank of English. I would like to thank Collins Dictionaries and the University of Birmingham for permission to use and quote from this corpus. I would also like to acknowledge the contribution of colleagues and students at the University of Birmingham, whose insights into the uses of corpora have informed my own understanding.