Chapter One: Corpus linguistics

1.1 Introduction

1.2 Mode of communication

1.3 Corpus-based versus corpus-driven linguistics

1.4 Data collection regimes

1.4.1 Monitor corpora

1.4.2 The Web as Corpus

1.4.3 The balanced corpus approach

1.4.4 Balance, representativeness and comparability

1.4.5 ‘Opportunistic’ corpora and minority and endangered languages

1.5 Annotated versus unannotated corpora

1.6 Total accountability versus data selection

1.6.1 Total accountability, falsifiability, and replicability

1.6.2 Data selection – not (necessarily) a bad thing

1.7 Monolingual versus multilingual corpora

1.8 Conclusion

Chapter Two: Accessing and analysing corpus data

2.1 Introduction

2.2 Are corpora the answer to all research questions in linguistics?

2.3 Corpus annotation

2.3.1 Metadata, markup, and annotation

2.3.2 Consistency of annotation

2.4 Introducing concordances

2.5 A historical overview of corpus analysis tools
2.5.1 First-generation concordancers

2.5.2 Second-generation concordancers

2.5.3 Third-generation concordancers

2.5.4 Fourth-generation concordancers

2.6 Statistics in corpus linguistics

2.6.1 Descriptive statistics

2.6.2 Beyond descriptive statistics

2.7 Conclusion

Chapter Three: The web, laws and ethics

3.1 Introduction

3.2 The web and legal issues

3.3 Ethical issues

3.3.1 Ethics and respondents

3.3.2 Ethics and corpus builders

3.3.3 Ethics and corpus distributors

3.3.4 Ethics and corpus users

3.3.5 Some cases of ethically problematic research

3.4 Conclusion

Chapter Four: English Corpus Linguistics

4.1 Introduction

4.2 University College London (UCL)

4.3 Lancaster University

4.4 University of Birmingham
4.5 Université Catholique de Louvain

4.6 University of Nottingham

4.7 Northern Arizona University and the USA

4.8 Conclusion

Chapter Five: Corpus-based studies of synchronic and diachronic variation

5.1 Introduction

5.2 Diachronic change from Old English to Modern English

5.2.1 Some notable general historical corpora

5.2.2 Specialised historical corpora

5.3 Diachronic variation in contemporary Modern English

5.3.1 The Brown Family of corpora

5.3.2 Results from the analysis of the Brown Family

5.4 The multi-dimensional approach to variation

5.4.1 An overview of the MD method

5.4.2 Applications of the MD approach

5.4.3 Criticisms of Biber's MD methodology

5.4.4 The MD approach: a summary

5.5 Corpora and variationist sociolinguistics

5.6 Conclusion

Chapter Six: Neo-Firthian corpus linguistics

6.1 Introduction

6.2 Collocation

6.3 Discourse
6.4 Semantic prosody and semantic preference

6.5 Lexis and grammar

6.5.1 The Idiom Principle

6.5.2 Pattern Grammar

6.5.3 Lexical Priming

6.6 Corpus-as-theory versus corpus-as-method

6.6.1 Data and theory

6.6.2 Text sampling

6.6.3 Corpus annotation

6.6.4 Corpus linguistics and pre-existing theory

6.7 Conclusion: Sinclair’s contribution to corpus linguistics

Chapter 7: Corpus methods and functionalist linguistics

7.1 Introduction

7.2 Functionalism in linguistics: a brief overview

7.3 Corpus-based research from a functionalist perspective

7.4 Corpora and typology

7.5 Corpora and cognitive approaches to linguistics

7.5.1 Cognitive approaches to syntax

7.5.2 From syntax to semantics

7.6 Corpora in the analysis of metaphor

7.7 Conclusion

Chapter Eight: The convergence of corpus linguistics, psycholinguistics, and functionalist linguistics
8.1 Introduction

8.2 Corpus methods and psycholinguistics

8.2.1 Corpus data in experiments on language processing

8.2.2 Language acquisition and child language corpora

8.2.3 Corpus data and computational modelling of language acquisition

8.2.4 Formulaic language

8.2.5 Corpora in psycholinguistics: an afterword

8.3 The convergence of neo-Firthian corpus linguistics and functionalist linguistics

8.4 Conclusion

Chapter Nine: Conclusion

9.1 Introduction

9.2 The story of corpus linguistics, from past to future

9.3 Revisiting old friends: computational linguistics

9.4 The textually-mediated world: the humanities and social sciences

9.5 The challenge ahead: integrating corpora with new methods in linguistics

9.6 The final word