Contents

Acknowledgements viii
List of Abbreviations ix

1 Early Corpus Linguistics and the Chomskyan Revolution 1
   1.1 Is Corpus Linguistics a Branch of Linguistics? 2
   1.2 Early Corpus Linguistics 2
   1.3 What Chomsky Said 5
   1.4 Why is this Book being Written? 13
   1.5 Corpus Linguistics from the 1950s to the early 1980s 20
   1.6 Conclusion 24
   1.7 Study Questions 25
   1.8 Further Reading 25
       Notes 26

2 What is a Corpus and What is in It? 29
   2.1 Corpora vs. Machine-readable Texts 29
   2.2 Text Encoding and Annotation 32
   2.3 Multilingual Corpora 69
   2.4 Chapter Summary 73
   2.5 Study Questions 73
   2.6 Further Reading 74
       Notes 74

3 Quantitative Data 75
   3.1 Introduction 75
   3.2 Qualitative vs. Quantitative Analysis 76
   3.3 Corpus Representativeness 77
   3.4 Approaching Quantitative Data 81
   3.5 Chapter Summary 98