Using Corpora for Language Research

Studies in the Honour of Geoffrey Leech

Edited by Jenny Thomas and Mick Short

Longman
London and New York
Contents

List of contributors vii
Preface ix

Part 1 Using corpora for language research

1. Corpora are becoming mainstream
   Jan Svartvik 3

2. From central embedding to corpus linguistics
   Geoffrey Sampson 14

3. Sipping a cocktail of corpora
   Randolph Quirk and Gabriele Stein 27

4. Corpora, databases and the organization of linguistic data
   Gerry Knowles 36

Part 2 Corpus-based language studies

5. Towards a unified account of the syntax and semantics of GET
   Stig Johansson and Signe Oksefjell 57

6. Complement clauses in English
   Sidney Greenbaum, Gerald Nelson and Michael Weitzman 76

7. Methodologies for studying a corpus of doctor-patient interaction
   Jenny Thomas and Andrew Wilson 92

8. Using a corpus for stylistics research: speech and thought presentation
   Mick Short, Elena Semino and Jonathan Culpeter 110
Part 3 Applications of corpus-based research to speech and language technology

9. A probabilistic LR parser of part-of-speech and punctuation labels
Ted Briscoe and John Carroll 135

10. Machine learning from corpus resources for speech and handwriting recognition
Eric Atwell 151

11. The robust tagging of unrestricted text: the BNC experience
Roger Garside 167

12. Template analysis: Bridging the gap between grammar and the lexicon
Steve Fligelstone, Paul Rayson and Nicholas Smith 181

Part 4 Wider applications of corpus-based research

13. Sentence and word alignment in the CRATER Project
Tony McEnery and Michael Oakes 211

14. English corpus linguistics and the foreign language teaching syllabus
Dieter Mindt 232

15. Do corpora have a role in language assessment?
Charles Alderson 248

16. Computer lexicography — the importance of representativeness in relation to frequency
Della Summers 260

Appendix: Geoffrey N. Leech — Academic Honours and publications 267
References 274
Index of Authors 297
Index of Subjects 300