

# Contents

Preface	v
<b>Part I</b>	
<b>Corpus design and text encoding</b>	<b>1</b>
The optimum corpus sample size? <i>Pieter de Haan</i>	3
Corpus sampling <i>Jeremy Clear</i>	21
International Corpus of English: Corpus design – problems and suggested solutions <i>Gerhard Leitner</i>	33
The Northern Ireland transcribed corpus of speech <i>John Kirk</i>	65
Problems in the compilation of a corpus of standard Caribbean English: A pilot study <i>Christian Mair</i>	75
The Text Encoding Initiative: A progress report <i>Lou Burnard</i>	97
<b>Part II</b>	
<b>Automated syntactic and semantic text analysis</b>	<b>109</b>
Pinpointing problematic tagging decisions <i>Nancy Belmore</i>	111
Inferences and lexical relations <i>Willem Meijs</i>	123

Tracing cohesive relations in corpora samples using a machine-readable dictionary <i>Sylvia Janssen</i>	143
Developing a scheme for annotating text to show anaphoric relations <i>Steve Fligelstone</i>	153
SUSANNE — a deeply analysed corpus of American English <i>Geoffrey Sampson</i>	171
Information retrieval and corpora <i>Stephen Sutton and Anthony McEnery</i>	191
<b>Part III</b>	
<b>Corpora in language description</b>	211
Relative infinitives in spoken and written English <i>Christer Geisler</i>	213
<i>Who(m)?</i> Case marking of wh-pronouns in written British and American English <i>Edgar W. Schneider</i>	231
Discourse category and text type classification: Procedural discourse in the Brown and the LOB corpora <i>Kay Wikberg</i>	247
Opaque and transparent features of Indian English <i>S. V. Shastri</i>	263
Computer analysis of spelling variants in Chaucer's <i>Canterbury Tales</i> <i>Geoff Barnbrook</i>	277
Pitch contours and tones in the Lancaster/IBM spoken English corpus <i>Gerry Knowles</i>	289

What do you think of that: A pilot study of the phraseology of the core words in English?	301
<i>Antoinette Renouf</i>	
Sequences of spatial and temporal adverbials in spoken and written English	319
<i>Hilde Hasselgård</i>	
Grammatical or native like?	329
<i>Göran Kjellmer</i>	
Collocation and bilingual text	345
<i>Jacques Noël</i>	
Keyword index	359