List of figures *page* ix
List of tables xi
Acknowledgments xv

1 Setting the scene 1

2 Ongoing language change: problems of detection and verification 12
2.1 "Visible" and "invisible" changes 12
2.2 The pitfalls of anecdotal observation 15
2.3 Documenting change 21
2.4 Outlook: a plea for methodological pluralism 33

3 Lexical change in twentieth-century English 36
3.1 Introduction 36
3.2 Case studies 38
3.3 Major trends 53
3.4 Neologizing in its social context 69

4 Grammatical changes in twentieth-century English 82
4.1 Introduction 82
4.2 Review of the literature 84
4.3 Aspect: twentieth-century changes in the structure and use of the progressive 88
4.4 The *going to*-future 95
4.5 Modality: *must* and *shall* – two modals on the way out, and possible replacements 100
4.6 Further developments in tense, aspect, modality: a synopsis of current research 108
4.7 Current changes in the English voice system 111
4.8 Nonfinite verb forms: some twentieth-century developments in the field of clausal complementation 119

4.9 Nouns, pronouns, adjectives 140
4.10 Conclusion 154

5 Pronunciation 156
5.1 Introduction 156
5.2 A history of RP in the twentieth century 162
5.3 "General American": myth or reality? 173

6 Language change in context: changing communicative and discourse norms in twentieth-century English 181
6.1 Introduction 181
6.2 The colloquialization of written English in the twentieth century 183
6.3 Americanization? 193
6.4 Analysis of selected sample texts 195

Conclusion 200

Appendix 1 Brief survey of the corpora used for the present study 206

Appendix 2 The OED Baseline Corpora 210

Appendix 3 Estimating text size in the newspaper archives and the World Wide Web 213

Appendix 4 A quarterly update of the OED Online (New Edition) – 13 March 2003: *Motswana* to *mussy* 217

References 231
Index 242

Figures

1.1 Languages of publication in five natural sciences (1879–1980), (Tsunoda 1983) *page* 9
2.1 Four matching one-million-word corpora of written English 24
3.1 Frequency of use of selected computer neologisms in *The Guardian (and Observer) on CD-ROM* 41
3.2 Frequency of use of selected military neologisms in *The Guardian (and Observer) on CD-ROM* 44
3.3 March 2003 OED updates for words containing the combining form *multi-* 55
3.4 March 2003 OED updates – out-of-sequence entries 56
3.5 Frequency of selected verbs of the *up/down* + V type in *The Guardian (and Observer) on CD-ROM* 65
3.6 Spread of three deverbal adjectives in *The Guardian (and Observer) on CD-ROM* 69
4.1 *Going to* and *gonna* 1600–2000 – frequency as n/10,000 citations 97
4.2 *Get*-passives according to age in the BNC 115
4.3 Nonfinite complements of *remember* in the OED quotation base by century – normalized frequency as n/10,000 quotations 127
4.4 *Help* + infinitive 1600–2000 – frequency as n/10,000 quotations 138
4.5 Analytical and synthetic comparison for four classes of adjectives (Kytö and Romaine 1997: 344) 150
5.1 The decline of /ʊə/ (John Wells, source: http://www.phon.ucl.ac.uk/home/wells/) 166
5.2 Major dialect areas in the US based on the *Dictionary of American regional English* (DARE) (Carver 1987) and the *Phonological atlas of North America* (Labov et al. 2006)

(source: http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html) 178
5.3 Merger of the vowels in *cot* and *caught* (source: William Labov, http://www.ling.upenn.edu/phono_atlas/ICSLP4.html#Heading2) 180
A3.1 Textual growth in *The Guardian (and Observer) on CD-ROM* 215
A3.2 *Deep breath* in top-level Web domains, December 2002 to March 2003 216

Tables

1.1 Population of major urban centers in the English-using world *page* 8
1.2 Percentage of languages in natural science publications, 1980 to 1996 10
1.3 Percentage of languages used in publications in the humanities, 1974 to 1995 (adapted from a graph in Ammon 1998a: 167) 10
2.1 Prepositions following *different* in four corpora 26
2.2 Prepositions following *different* in regionally stratified Web material (Google, 30 May 2004) 27
2.3 Proportion of *on/upon* in four corpora 28
2.4 Proportion of *on/upon* in three samples from the OED quotation base 29
2.5 Lexical items most characteristic of four groups of speakers in a corpus of spoken British English (compiled from Rayson et al. 1997) 30
2.6 Frequency of *wanna* in the BNC per age group (x/1,000,000 words) 32
3.1 *Wicked* – frequency in the spoken-demographic BNC per million words by age group 48
3.2 *Massive* – frequency in the spoken-demographic BNC per million words by age group 49
3.3 OED Online – new words first attested in the twentieth century 58
3.4 Nineteenth- and twentieth-century borrowings from Yiddish in the OED 60
3.5 Twentieth-century borrowings from Spanish first attested in the OED 62
3.6 Proper nouns consisting entirely of capital letters: comparison of frequency in LOB and F-LOB 62

3.7 Prefixed verbs in *up-* and *down-* in the OED (twentieth century) and the BNC (compiled from information in Scheible 2005) 64
3.8 *Up/down* + *V* in four corpora – types (tokens) 64
3.9 *Ongoing* in four corpora 66
3.10 *On/off/in/out/up/down* + *V* + *ing* in four corpora – survey 67
3.11 *On/off/in/out/up/down* + *V* + *ing* in four corpora – type/token ratios 67
3.12 Discourse frequency of *virtue(s)*, *value(s)*, etc. in selected twentieth-century corpora 78
4.1 Progressive forms in the press sections (A–C) of four reference corpora 90
4.2 *Been being* and *be being* on the English-language Web (Google, 23 July 2003) 92
4.3 *Going to*-futures in four corpora (examples from direct speech in brackets) 96
4.4 *Going to* – manually post-edited output for four quarter-centuries 98
4.5 *Will/shall-* and *going to*-futures in four spoken corpora (percentages, adapted from Szmrecsanyi 2003: 303) 98
4.6 *Going to-* and *will*-futures in two age groups in the spoken-demographic BNC (normalized frequencies/words per million) 99
4.7 Use of the modal auxiliaries in four corpora 101
4.8 *Shall*-futures in four corpora 102
4.9 *Have (got) to* in four corpora (press texts, section A–C only) 104
4.10 *Have got to* in four corpora 104
4.11 *Have to* in four corpora 104
4.12 Obligation and necessity in the Santa Barbara Corpus of Spoken American English and the conversation component of ICE-GB 105
4.13 *Must* and *have to* by function in ICE-GB (spoken), adapted from Depraetere and Verhulst (forthcoming) 105
4.14 *Get*-passives in four corpora (examples from direct speech in brackets) 113
4.15 Frequency indices for *get*-passives in the BNC 116
4.16 *For* + NP + *to*-infinitival clauses in four corpora 124
4.17 *For* + NP + *to*-infinitival clauses in three OED Baseline Corpora 124
4.18 *For* + NP + *to*-infinitival clauses in three OED Baseline Corpora and F-LOB and Frown (normalized, as instances per million words) 125

4.19 Gerunds and infinitives after *remember* in the OED quotation base – normalized frequencies ("n/10,000 quotations," rounded to the first decimal, absolute frequencies in brackets) 127
4.20 Notional subjects in gerundial constructions after *remember* in the OED quotation base 128
4.21 Proportion of infinitival and gerundial complements after *begin* in four corpora 128
4.22 Infinitive vs. gerund complements with *begin* in selected British databases 129
4.23 Infinitive vs. gerund complements with *begin* in selected American databases 130
4.24 *To*-infinitive: V-*ing* after *start* in four corpora 130
4.25 Nonfinite complements of *prevent* in the OED Baseline Corpus 131
4.26 Ratio of *prevent* NP *from* V-*ing* vs. *prevent* NP V-*ing* in four corpora 132
4.27 Ratio of *stop* NP *from* V-*ing* vs. *stop* NP V-*ing* in four corpora 132
4.28 *Save* NP V-*ing* in selected top-level Web domains 135
4.29 *To*- vs. bare infinitives with *help* in four corpora 136
4.30 Complementation of *help* in the "spoken-demographic" BNC 136
4.31 *Whom* in four matching corpora 142
4.32 Nominative vs. objective case for pronouns in specific syntactic contexts (BNC) 144
4.33 *'S*-genitives in four tagged corpora 146
4.34 Use of genitive and *of*-phrase in the press sections (A–C) of two corpora (compiled from Raab-Fischer 1995) 147
4.35 Normalized frequencies (occurrences per million words) for selected genitives in spoken and written text types from the BNC 148
4.36 *Fewer* vs. *less* with count nouns in selected corpora 151
5.1 Variably pronounced words in contemporary RP (John Wells, Longman Pronunciation Survey) 168
6.1 Frequencies of *say* in selected genre categories of four corpora 189
6.2 Verb and negative contractions in the four corpora (from Leech and Smith 2005) 189
6.3 Contraction ratios (*not*-contractions) in journalistic and academic prose 190
6.4 Decline in frequency of use of the *be*-passive in the four reference corpora (from Leech and Smith 2005) 190

6.5 Frequency of noun+common-noun combinations in four corpora 192
A2.1 Composition of the OED Baseline Corpora 210
A2.2 The five most common words in Baseline, Brown, LOB, Frown and F-LOB 211
A2.3 "Absolute frequencies (rank)" for selected function words in five corpora 212
A3.1 Frequency of selected collocations in two corpora and eight newspaper archival discs 214
A3.2 Frequencies of ten collocations in the BNC and selected top-level Web domains 215