

STUDIES IN ENGLISH LANGUAGE

Christian Mair

Twentieth-Century English

History, Variation and Standardization

CAMBRIDGE

Twentieth-Century English

History, Variation, and Standardization

CHRISTIAN MAIR



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press,
New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521832199

© Christian Mair 2006

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN-13 978-0-521-83219-9 hardback

ISBN-10 0-521-83219-5 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this publication,
and does not guarantee that any content on such websites is, or will remain, accurate
or appropriate.

The following authors and publishers have given permission to use extended
quotations from their work:

© Tom Leonard, from *Intimate voices: poems 1965–1983*, Etruscan Books Devon
2003 (for “i’ve not got a light”)

© India Knight/*The Sunday Times*, 11 November 2001 (for the extract from India
Knight, “Speak proper? Not likely”)

© Little, Brown Book Group (for the passage reproduced from Sarah Waters, *The
night watch*, London: Virago Press, 2006)

© OUP (for Appendix 4, “Motswana–mussy”)

Every effort has been made to secure necessary permissions to reproduce copyright
material in this work, though in some cases it has proved impossible to trace or
contact copyright holders. If any omissions are brought to our notice, we will be
happy to include appropriate acknowledgments on reprinting or in any subsequent
edition.

Acknowledgments

When I started working on the corpus-based investigation of change in progress in present-day English in the early 1990s, writing a book on this topic soon began to seem an attractive idea, though one which was bound to remain rather theoretical for a long time. Now that the book is about to be published, I would like to take the opportunity to thank a number of organizations and people without whose help it is unlikely that *Twentieth-Century English* would have seen the light of day a mere fifteen years after the idea for the book was first conceived.

To the *Deutsche Forschungsgemeinschaft* (DFG) I owe thanks for generously supporting two corpus-related research projects from 1994 to 1996 and from 2003 to 2006. Without their funding, F-LOB and Frown, two corpora providing important evidence for the present study, would have been completed much later (if at all), and various laborious but extremely useful annotation schemes and other enhancements would not even have been attempted. If the DFG gave the money, more members of my team at Freiburg than can be named here have given their expertise and dedication over the years. If I single out Marianne Hundt, Andrea Sand, Stefanie Rapp, Birgit Waibel, and Lars Hinrichs by name, I hope that many others involved in the projects for longer or shorter periods of time will not take this amiss.

At CUP, I would like to thank Kate Brett, who, after discussing the idea of a history of twentieth-century English with me at the Edinburgh Late Modern English conference in 1998, encouraged me to formalize it by submitting a proposal to the Press. Helen Barton, who eventually took over from Kate, was equally sympathetic and additionally showed welcome patience in the final stages of completing the manuscript. Valuable suggestions for improvements were made by Merja Kytö, one of the series editors, who carefully went through the first version of the completed manuscript. At the very end of the production process, working together with Nikky Twyman as a copy editor was a pleasant and humbling experience, pleasant because of her quiet and good-humored efficiency, and humbling because of the number of oversights she spotted in a manuscript which I thought I had proofread carefully.

xvi Acknowledgments

I hope that the book will convey to its readers some of my own fascination with the “living history” of English, its recent past, its rich and diversified present, and its future, and that it will encourage others to keep researching the many questions which I have had to leave unanswered.

Freiburg, February 2006 CM

Contents

<i>List of figures</i>	page ix
<i>List of tables</i>	xi
<i>Acknowledgments</i>	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 <i>going to</i> -future	95
4.5 Modality: <i>must</i> and <i>shall</i> – 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

viii Contents

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: <i>Motswana</i> to <i>mussy</i>	217
	<i>References</i>	231
	<i>Index</i>	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 <i>The Guardian (and Observer) on CD-ROM</i>	41
3.2	Frequency of use of selected military neologisms in <i>The Guardian (and Observer) on CD-ROM</i>	44
3.3	March 2003 OED updates for words containing the combining form <i>multi-</i>	55
3.4	March 2003 OED updates – out-of-sequence entries	56
3.5	Frequency of selected verbs of the <i>up/down</i> + V type in <i>The Guardian (and Observer) on CD-ROM</i>	65
3.6	Spread of three deverbal adjectives in <i>The Guardian (and Observer) on CD-ROM</i>	69
4.1	<i>Going to</i> and <i>gonna</i> 1600–2000 – frequency as n/10,000 citations	97
4.2	<i>Get</i> -passives according to age in the BNC	115
4.3	Nonfinite complements of <i>remember</i> in the OED quotation base by century – normalized frequency as n/10,000 quotations	127
4.4	<i>Help</i> + 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 <i>Dictionary of American regional English</i> (DARE) (Carver 1987) and the <i>Phonological atlas of North America</i> (Labov et al. 2006)	

x List of figures

	(source: http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html)	178
5.3	Merger of the vowels in <i>cot</i> and <i>caught</i> (source: William Labov, http://www.ling.upenn.edu/phono_atlas/ICSLP4.html#Heading2)	180
A3.1	Textual growth in <i>The Guardian (and Observer)</i> on CD-ROM	215
A3.2	<i>Deep breath</i> 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 <i>different</i> in four corpora	26
2.2	Prepositions following <i>different</i> in regionally stratified Web material (Google, 30 May 2004)	27
2.3	Proportion of <i>on/upon</i> in four corpora	28
2.4	Proportion of <i>on/upon</i> 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 <i>wanna</i> in the BNC per age group (x/1,000,000 words)	32
3.1	<i>Wicked</i> – frequency in the spoken-demographic BNC per million words by age group	48
3.2	<i>Massive</i> – 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

xii List of tables

3.7	Prefixed verbs in <i>up-</i> and <i>down-</i> in the OED (twentieth century) and the BNC (compiled from information in Scheible 2005)	64
3.8	<i>Up/down</i> + <i>V</i> in four corpora – types (tokens)	64
3.9	<i>Ongoing</i> in four corpora	66
3.10	<i>On/off/in/out/up/down</i> + <i>V</i> + <i>ing</i> in four corpora – survey	67
3.11	<i>On/off/in/out/up/down</i> + <i>V</i> + <i>ing</i> in four corpora – type/token ratios	67
3.12	Discourse frequency of <i>virtue(s)</i> , <i>value(s)</i> , etc. in selected twentieth-century corpora	78
4.1	Progressive forms in the press sections (A–C) of four reference corpora	90
4.2	<i>Been being</i> and <i>be being</i> on the English-language Web (Google, 23 July 2003)	92
4.3	<i>Going to</i> -futures in four corpora (examples from direct speech in brackets)	96
4.4	<i>Going to</i> – manually post-edited output for four quarter-centuries	98
4.5	<i>Will/shall-</i> and <i>going to</i> -futures in four spoken corpora (percentages, adapted from Szmrecsanyi 2003: 303)	98
4.6	<i>Going to-</i> and <i>will-</i> 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	<i>Shall</i> -futures in four corpora	102
4.9	<i>Have (got) to</i> in four corpora (press texts, section A–C only)	104
4.10	<i>Have got to</i> in four corpora	104
4.11	<i>Have to</i> 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	<i>Must</i> and <i>have to</i> by function in ICE-GB (spoken), adapted from Depraetere and Verhulst (forthcoming)	105
4.14	<i>Get</i> -passives in four corpora (examples from direct speech in brackets)	113
4.15	Frequency indices for <i>get</i> -passives in the BNC	116
4.16	<i>For</i> + NP + <i>to</i> -infinitival clauses in four corpora	124
4.17	<i>For</i> + NP + <i>to</i> -infinitival clauses in three OED Baseline Corpora	124
4.18	<i>For</i> + NP + <i>to</i> -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 <i>remember</i> 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 <i>remember</i> in the OED quotation base	128
4.21	Proportion of infinitival and gerundial complements after <i>begin</i> in four corpora	128
4.22	Infinitive vs. gerund complements with <i>begin</i> in selected British databases	129
4.23	Infinitive vs. gerund complements with <i>begin</i> in selected American databases	130
4.24	<i>To</i> -infinitive: <i>V-ing</i> after <i>start</i> in four corpora	130
4.25	Nonfinite complements of <i>prevent</i> in the OED Baseline Corpus	131
4.26	Ratio of <i>prevent</i> NP <i>from V-ing</i> vs. <i>prevent</i> NP <i>V-ing</i> in four corpora	132
4.27	Ratio of <i>stop</i> NP <i>from V-ing</i> vs. <i>stop</i> NP <i>V-ing</i> in four corpora	132
4.28	<i>Save</i> NP <i>V-ing</i> in selected top-level Web domains	135
4.29	<i>To</i> - vs. bare infinitives with <i>help</i> in four corpora	136
4.30	Complementation of <i>help</i> in the “spoken-demographic” BNC	136
4.31	<i>Whom</i> 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 <i>of</i> -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	<i>Fewer</i> vs. <i>less</i> 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 <i>say</i> 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 (<i>not</i> -contractions) in journalistic and academic prose	190
6.4	Decline in frequency of use of the <i>be</i> -passive in the four reference corpora (from Leech and Smith 2005)	190

xiv List of tables

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

1 Setting the scene

... ask yourself whether our language is complete; – whether it was so before the symbolism of chemistry and the notation of the infinitesimal calculus were incorporated in it; for these are, so to speak, suburbs of our language. (And how many houses or streets does it take before a town begins to be a town?) Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses.

(Ludwig Wittgenstein, *Philosophische Untersuchungen/Philosophical investigations*, translated by G. E. M. Anscombe. Oxford: Blackwell, 1967: 18)

Anyone proposing to write a history of the English language in the twentieth century begs a number of questions, which it is necessary to answer at the very outset of what might seem an excessively ambitious project.

Isn't the topic too vast and complex for a single author to tackle? If one bears in mind that in contrast to historians of Old and Middle English, who in general suffer from a poverty of evidence, the historian of recent and contemporary English is deluged with data and, in principle, needs to write separate histories of several richly documented standard and nonstandard varieties, and a history of contact and influence among them, the answer to this question is an obvious "yes." The only justification that the present writer is able to offer for undertaking the project against the odds is that he has narrowed the focus from the very start to one highly codified variety, namely the written standard which – in the twentieth century – was in use throughout the English-speaking world with minor local differences in spelling, lexicon, idiom, and grammar. The spoken usage of educated speakers in formal situations, which can be considered the oral correlate of this written standard, will be considered where relevant. While this restriction is problematical for many reasons, it is justifiable because of the social prominence of the standard in the present, and also because most histories of English covering developments from the late Middle English period onwards have – explicitly or implicitly – been histories of the standard, too.

2 Twentieth-century English

What about the observer's paradox? In a history of contemporary English, this paradox takes two forms. First, it might be impossible for us to identify and document recent and ongoing linguistic changes against the background noise of synchronic regional, social, or stylistic variation that surrounds us and in which these diachronic developments are embedded. Second, assuming that we can identify ongoing language change, we will still have to ask the question whether we can free ourselves from the social prejudices which have normally caused ongoing changes to be viewed negatively – as instances of erroneous or illogical usage or even as signs of decay or degeneration. As for the first manifestation of the paradox (our ability or inability to even perceive ongoing change), there is a long tradition of skepticism – exemplified, for example, in a much-quoted statement in Bloomfield's *Language*.¹ The optimistic tradition, by contrast, is a much younger one, going back to William Labov's 1960s work on extrapolating diachronic trends from synchronic variation, and is still largely confined to sociolinguistic circles. As a descriptive contribution to the history of English from around 1900 to the present, the current study will not be able to settle the dispute between the optimists and the pessimists in a principled way; rather, it has opted for a practical compromise by not concentrating on all aspects of linguistic change to the same degree. Little emphasis will be placed on the often futile search for the first authentic and/or unambiguous recorded instance of an innovation, or on speculations about possible reanalyses, rule reorderings, or other adjustments in speaker competence or the abstract system underlying the recorded data. Rather, the focus will be on the spread of innovations through varieties, textual genres, and styles, or on provable shifts in frequency of use in a defined period. In other words, the present study aims to exploit the full potential of the corpus-linguistic working environment that has become available to the student of English in recent decades – an environment which, in addition to corpora in the narrow sense (that is, machine-readable collections of authentic texts or natural discourse which have been compiled expressly for the use of linguists), now includes important electronic dictionaries such as the continuously updated online version of the *Oxford English dictionary* (OED) and a vast mass of digitized textual material not originally compiled for the purposes of linguistic study.²

¹ "The process of linguistic change has never been directly observed; we shall see that such observation, with our present facilities, is inconceivable" (Bloomfield 1933: 347). In Chapter 2 we shall see that Bloomfield's position – categorically negative in this passage – is modified elsewhere in his work and, more importantly, that there has been considerable improvement in "our present facilities."

² The corpora consulted for the present study and the methods used for their analysis will be discussed in the appropriate places, with a summary of the relevant information in the Appendix. Readers interested in a more general introduction to the thriving field of English corpus-linguistics are referred to introductory handbooks such as Biber et al. (1998) or Meyer (2002).

As hinted at above, the second manifestation of the observer's paradox in the study of ongoing linguistic change is the possible distorting influence of the prescriptive tradition. This is a serious problem which needs to be acknowledged. Of course, it is unlikely that professional linguists will repeat the often exaggerated and irrational value judgments on linguistic usage propagated by this tradition. The effect the prescriptive tradition exerts on research on current change is more subtle and indirect; it introduces a hidden bias into the study of ongoing change by setting the agenda of topics worth the researcher's attention. In this way, relatively minor points of usage and variation receive an amount of attention completely out of proportion to their actual significance (even if the linguist's intention may merely be to refute prescriptive prejudice), while much more important and comprehensive changes go unnoticed. To give a few examples, the literature on grammatical change in present-day English is rife with comment on the allegedly imminent disappearance of *whom* (a development for which there is very little documentary evidence – see Chapter 4) or the use of *hopefully* as a sentence adverb (which at least is a genuine twentieth-century innovation on the basis of the OED evidence, with a first attestation for the year 1932). This is so because these two points of usage have a high profile as linguistic markers in the community and are much discussed by prescriptivists. Measured against the sum total of ongoing changes in present-day English, however, both are mere trivia. Comprehensive and far-reaching developments, on the other hand, which affect the very grammatical core of Modern English, such as the spread of gerunds into functions previously reserved for infinitives, tend to go unnoticed because these changes proceed below the level of conscious speaker awareness and hence do not arouse prescriptive concerns. Again, the remedy here is the use of corpora. Corpora make it possible to describe the spread of individual innovations against the background of the always far greater and more comprehensive continuity in usage, and corpus-based studies of linguistic change in progress are therefore likely to correct more alarmist perceptions based on the unsystematic collection of examples or impressionistic observation, which are inevitably biased towards the strange, bizarre, and unusual.

Is there sufficient previous work on the recent history of English to write a survey such as the present one?

A mere twenty years ago, the answer to this question would have been in the negative. Throughout the twentieth century there was never a dearth of "state of the language" books aimed at the general educated public. Brander Matthews, the American man of letters, published his *Essays on English* in 1921. J. Hubert Jagger's *English in the future*, which – in contrast to what the title suggests – is mostly about English in the present, appeared in 1940. More recently, two collections of essays on the *State of the language* were edited by Leonard Michaels and Christopher Ricks (Michaels and Ricks 1980, Ricks 1991). Most such works cover ongoing changes (whether perceived or real), but they tend to do so only very superficially. A more reliable source of in-depth

4 Twentieth-century English

information on current change would thus seem to be the major scholarly histories of the language. However, until recently these tended to peter out at some point around 1800, leaving the history of English in the nineteenth and twentieth centuries as largely uncharted territory.³

Over the last twenty years, however, the situation has definitely improved. There has been a surge of interest in research on the recent history of English, which has also resulted in several landmark publications offering at least partial surveys. The recent history of English, with a strong (and, in the first two cases, exclusive) emphasis on the nineteenth century, is dealt with in two book-length studies (Bailey 1996, Görlach 1999), and volume IV (“1776–1997”) of the *Cambridge history of the English language*. In a broad sense, the present book is a chronological continuation of Bailey’s and Görlach’s monographs – albeit with slightly different priorities. In comparison to Bailey (1996), it will aim for a fuller coverage of the structural history of the language (particularly the grammar), whereas in comparison to Görlach the two major differences are that the treatment is not restricted to England exclusively and that, in compensation for the widening of the geographical scope, less emphasis will be placed on the didactic presentation and annotation of source texts. The most important point of reference for most chapters, though, will be volume IV (“1776–1997”) of the *Cambridge history*. As will become clear, this work’s treatment of nineteenth-century developments is admirable and provides a good foundation for the present study. Its coverage of the twentieth century, on the other hand, is less complete and will be expanded here.

More problematical sources than these scholarly linguistic works are the many popular works on the recent history of English and the state of the language. For one thing, the number of such publications is vast – from books written by non-linguists for lay audiences (e.g., Michaels and Ricks 1980, Ricks 1991, or Howard 1984) to works such as Barber (1964) or Potter (1969 [1975]), which are valuable as provisional surveys of the field by experts. Many of these “state of the language” books are informed by a spirit of traditional prescriptivism and/or cultural pessimism or more concerned with the ideological and political aspects of language standardization than the linguistic facts themselves. But even a work such as Barber’s (1964) excellent survey of “linguistic change in present-day English” needs to be treated with some caution. The insights and claims it contains are generally based on the author’s anecdotal observations and unsystematic collection of examples, which – as will be shown in Chapter 2 – is a notoriously unreliable methodology in the documentation of ongoing changes.

³ This is partly a matter of author interest, which gave priority to earlier developments, and partly a result of publication date, as classic works such as Jespersen (1909–1949) have not really been challenged or even equaled in comprehensiveness of coverage and authoritativeness until recently.