



American and British English

Divided by a Common Language?

PAUL BAKER



Is British English becoming more like American English? If so, why, and in what ways? This book compares examples of American and British language data from the 1930s, 1960s, 1990s and 2000s to track the most important ways that both varieties are changing over time and compares the extent to which they are following similar paths using a mixture of computer and human analysis. The analysis is carried out across several levels, including spelling differences (e.g. *colour* vs *color*), vocabulary (*truck* vs *lorry*) and a range of morphological, grammatical, semantic and pragmatic features. Baker explores the changing aspects of American and British society which help to explain the findings.

PAUL BAKER is Professor of English Language at Lancaster University. He has written fourteen books on a variety of topics, including language, corpus linguistics, discourse analysis and identities. These include *Using Corpora in Discourse Analysis* (2006) and *Discourse Analysis and Media Attitudes* (Cambridge, 2013). He is a commissioning editor of the journal *Corpora*.

'An engaging, in-depth look at British and American English. In addition, Baker demonstrates a range of methods for analyzing language at many levels, and for contextualizing the results.'

Randi Reppen, Professor of Applied Linguistics, Northern Arizona University

'*American and British English: Divided by a Common Language* provides a comprehensive, well-illustrated, and interesting description of how American and British English have changed from the 1930s through the 2000s, focusing on such topics as spelling differences, word frequency variations between the varieties, and the use of profanity and discourse markers.'

Charles Meyer, Professor of Applied Linguistics, University of Massachusetts, Boston

Cover illustration: © Sigal Suhler Moran

COVER DESIGNED BY HART McLEOD LTD

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-1-107-08886-3



Baker Ernest American and British English

CAMBRIDGE

American and British English

Divided by a Common Language?

Paul Baker

Lancaster University



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi - 110002, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107088863

DOI: 10.1017/9781316105313

© Paul Baker 2017

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 2017

Printed in the United Kingdom by Clays, St Ives plc

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

ISBN 978-1-107-08886-3 Hardback

ISBN 978-1-107-46088-1 Paperback

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.

American and British English

Is British English becoming more like American English? If so, why, and in what ways? This book compares examples of American and British language data from the 1930s, 1960s, 1990s and 2000s to track the most important ways that both varieties are changing over time and compares the extent to which they are following similar paths using a mixture of computer and human analysis. The analysis is carried out across several levels, including spelling differences (e.g. *colour* versus *color*), vocabulary (*truck* versus *lorry*) and a range of morphological, grammatical, semantic and pragmatic features. Baker explores the changing aspects of American and British society which help to explain the findings.

PAUL BAKER is Professor of English Language at Lancaster University. He has written fourteen books on a variety of topics, including language, corpus linguistics, discourse analysis and identities. These include *Using Corpora in Discourse Analysis* (2006) and *Discourse Analysis and Media Attitudes* (Cambridge, 2013). He is a commissioning editor of the journal *Corpora*.

Acknowledgements

I would like to thank Amanda Potts, Vaclav Brezina, Tony McEnery, Marianne Hundt, Geoffrey Leech, Nick Smith, Paul Rayson, Christian Mair, Matteo Di Christofaro, Mazura Muhammad, Wesam Ibrahim, Łukasz Pakuła, Andrew Hardie, Andrew Winnard and Helen Barton for their roles in making this book possible.

Contents

<i>List of Figures</i>	page viii
<i>List of Tables</i>	xi
<i>Acknowledgements</i>	xiii
1 Introduction	1
The Special Relationship	1
Corpus Linguistics	3
Meet the Brown Family	6
Comparing Corpora	11
Overview of the Book	24
2 Spelling Differences	27
Introduction	27
Finding Spelling Differences	29
-or or -our	31
-ize or -ise and Related Forms	33
-er or -re	36
-og or -ogue	37
-a/-o- or -ae/-oe-	39
-ce or -se	40
-ll-m or -lll-mme	43
Dropped <i>e</i>	45
-st	46
-ction or -xion	49
Toward or Towards	50
Gray or Grey	50
Conclusion	52
3 Letter Sequences and Affixation	57
Introduction	57
Morphemes and Affixation	58
The Coefficient of Variation	61
-ology	65
trans-	66
-techn-	67
-graph-	68
-cess	69

-ism	71
-ploy-	73
-cious-	75
-ment	77
-ence	79
-ness	80
-ual	81
-gress-	82
-fess-	83
-sion	84
Conclusion	85
4 Higher-Frequency Words	88
Introduction	88
Comparing the Varieties	89
Densification	95
Democratisation	101
Informalisation	106
Contrary Patterns: <i>Who</i> and <i>Say</i>	111
Two- and Three-Word Clusters	115
Conclusion	120
5 Lower-Frequency Words	122
Introduction	122
Identifying Candidate Items	122
Transport(ation)	127
Monarchy	131
Politics and Law	132
Business and Economics	136
People	140
Entertainment and Leisure	144
Conclusion	149
6 Part of Speech Categories	151
Introduction	151
Part of Speech Annotation	152
Key Comparisons and the Coefficient of Variation	153
Decreased Use of the Passive Voice	156
Decreased Modality	160
Decreased <i>Which</i> in Relative Clauses	166
Noun Compounds, Titles and Genitives	171
Conclusion	175
7 Semantic Categories	177
Introduction	177
Semantic Annotation	178
War and Law	184
Social Interactions and People	188
Technology and Media	192
Ideas and Concepts	197
Conclusion	203

Contents	vii
8 Swearing, Identity and Discourse Markers	205
Introduction	205
Swearing and Profanity	206
Language and Identity	211
Discourse Markers	222
Conclusion	234
9 Conclusion	236
Introduction	236
Americanisation	237
Densification	238
Democratisation	240
Informalisation and Colloquialisation	242
Grammaticalisation	244
Technologisation	245
Some Critical Considerations	249
<i>References</i>	254
<i>Index</i>	261

Figures

1.1	Frequencies of <i>so</i>	page 12
1.2	Frequencies of <i>who</i>	13
1.3	Frequencies of <i>didn't</i>	13
1.4	Mean word lengths	22
1.5	Mean sentence lengths	23
1.6	Standardised type token ratios	23
2.1	Proportion of cases of <i>-our</i> in British English and <i>-or</i> in American English	31
2.2	Proportion of cases of <i>-ise</i> in British English and <i>-ize</i> in American English	34
2.3	Proportion of cases of <i>-isation</i> in British English and <i>-ization</i> in American English	36
2.4	Proportion of cases of <i>-re</i> in British English and <i>-er</i> in American English	37
2.5	Proportion of cases of <i>-ogue</i> in British English and <i>-og</i> in American English	38
2.6	Proportion of cases of <i>-ae-</i> in British English and <i>-a-</i> in American English	39
2.7	Proportion of cases where <i>practise</i> is a verb in British English and <i>practice</i> is a verb in American English	41
2.8	Proportion of cases where <i>licence</i> is a noun in British English and <i>license</i> is a noun in American English	42
2.9	Proportion of cases of <i>defence</i> and <i>offence</i> in British English and <i>defense</i> and <i>offense</i> in American English	42
2.10	Proportion of cases of <i>-ll</i> in British English and <i>-l</i> in American English	44
2.11	Proportion of cases of <i>programme</i> in British English and <i>program</i> in American English	44
2.12	Proportion of cases of included <i>e</i> in British English and dropped <i>e</i> in American English	46
2.13	Proportion of cases of <i>amongst</i> in British English and <i>among</i> in American English	47

2.14	Proportion of cases of <i>whilst</i> in British English and <i>while</i> in American English	47
2.15	Proportion of cases of <i>connection(s)</i> as opposed to <i>connexion(s)</i> in British and American English	49
2.16	Proportion of cases of <i>towards</i> in British English and <i>toward</i> in American English	51
2.17	Proportion of cases of <i>grey</i> in British English and <i>gray</i> in American English	51
2.18	Summary of all relevant spelling differences between American and British English	55
3.1	The Coefficient of Variation	62
3.2	Frequencies of <i>-ology</i>	66
3.3	Frequencies of <i>trans-</i>	67
3.4	Frequencies of <i>-techn-</i>	68
3.5	Frequencies of <i>-graph-</i>	69
3.6	Frequencies of <i>-cess</i>	70
3.7	Frequencies of <i>-ism</i>	72
3.8	Frequencies of <i>-ploy-</i>	74
3.9	Frequencies of <i>-cious-</i>	76
3.10	Frequencies of <i>-ment</i>	77
3.11	Frequencies of <i>-ence</i>	80
3.12	Frequencies of <i>-ness-</i>	81
3.13	Frequencies of <i>-ual</i>	82
3.14	Frequencies of <i>-gress-</i>	83
3.15	Frequencies of <i>-fess-</i>	84
3.16	Frequencies of <i>-sion</i>	85
4.1	<i>UK</i> and <i>United Kingdom</i> in the British corpora	92
4.2	<i>US</i> and <i>United States</i> in the American corpora	93
4.3	Collocates of <i>upon</i> and <i>on</i> in British English, 1931	96
4.4	Collocates of <i>upon</i> and <i>on</i> in British English, 2006	97
4.5	Frequencies of <i>upon</i> and <i>on</i> in British English	98
4.6	Collocates of <i>around</i> and <i>round</i> in British English, 1931	99
4.7	Collocates of <i>around</i> and <i>round</i> in British English, 2006	100
4.8	Frequencies of <i>around</i> and <i>round</i> in British English	101
4.9	Frequencies of <i>will</i> as a modal verb	103
4.10	Frequencies of <i>I</i>	107
4.11	Frequencies of <i>me</i>	108
4.12	Frequencies of <i>who</i> and <i>whom</i> in American English	113
4.13	Frequencies of <i>say</i>	114
4.14	Frequencies of <i>to see</i>	118
6.1	Frequencies of <i>it</i> and <i>be</i> (infinitive)	157
6.2	Frequencies of VVN (past participle of lexical verb)	158

6.3	American English frequencies of PPIS1 (first person singular subjective personal pronoun: <i>I</i>) by text category	159
6.4	Frequencies of RG (adverb of degree)	162
6.5	Frequencies of VM (modal verbs)	163
6.6	Frequencies of DDQ (<i>wh</i> - determiners)	166
6.7	Frequencies of <i>that</i> and <i>which</i> relative clauses	168
6.8	Frequencies of zero relative clauses	169
6.9	Frequencies of NN2 (plural common nouns)	172
6.10	Frequencies of GE (genitive marker)	175
7.1	Frequencies of G3 (warfare, defence and army; weapons)	184
7.2	Frequencies of G2.1 (law and order)	186
7.3	Frequencies of male and female terms in American English	191
7.4	Frequencies of media tags in the British corpora	196
7.5	Frequencies per million words of 'danger' words in American English, by text category over time	198
8.1	Frequencies of sexual and excretory swear words	207
8.2	Frequencies of profane uses of language	209
8.3	Percentage of profane uses of religious words	209
8.4	Percentage of gendered terms that are female as opposed to all gendered terms	212
8.5	Percentage of gendered pronouns that are female as opposed to all gendered pronouns	213
8.6	Percentage of times a male term comes before an equivalent female term	214
8.7	Generic cases of <i>man</i> (from random samples of 100)	215
8.8	Frequencies of sexual identity words	217
8.9	Frequencies of racial categories	219
8.10	Percentage of uses of a race-related term as a noun (as opposed to an adjective)	220
8.11	Frequencies of collective plural nouns for racial terms	223
8.12	Frequencies of Z4 (discourse bin)	224

Tables

1.1	Text categories in the Brown family	page 7
1.2	The Brown family	8
1.3	Frequencies of the word <i>today</i> in the eight Brown corpora	15
1.4	Frequencies of the word <i>want</i> in the eight Brown corpora	17
1.5	Frequencies of the word <i>who</i> in the eight Brown corpora	18
1.6	Tools used for analysis	19
1.7	Sample concordance lines of <i>bank</i> from 1931 American English	20
2.1	List of spelling differences examined in this chapter	30
2.2	Summary of spelling patterns	54
3.1	American morphemes with the highest CVs	63
3.2	American morphemes with the lowest CVs	64
3.3	British morphemes with the highest CVs	64
3.4	British morphemes with the lowest CVs	64
4.1	Keywords in American English	89
4.2	Keywords in British English	91
4.3	Words showing strongest growth and decline in American and British English	94
4.4	American two-word clusters with high CVs	116
4.5	British two-word clusters with high CVs	116
4.6	Three-word clusters which show the most change	119
5.1	Frequencies for <i>holiday</i> and <i>vacation</i>	123
5.2	References to major British and American cities	125
5.3	Frequencies for <i>jazz</i>	126
5.4	Frequencies for <i>courgette</i> and <i>zucchini</i>	126
5.5	Transport words (preferences)	127
5.6	<i>Plane</i> and related words	129
5.7	Transport words (frequencies)	129
5.8	Terms relating to the monarchy	131
5.9	American politics and law keywords	133
5.10	British politics and law keywords	133
5.11	Terms relating to business and economics	137
5.12	Terms relating to people	141

5.13	Frequencies for <i>relations</i> and related words	142
5.14	Terms relating to entertainment and leisure	145
5.15	Variety-specific words which show the strongest frequency differences between 1931 and 2006	150
6.1	Most key American parts of speech, compared to British English	154
6.2	Most key British parts of speech, compared to American English	154
6.3	American part of speech tags with the highest CVs	155
6.4	British part of speech tags with the highest CVs	156
6.5	Excerpt of Internet meme illustrating perceptions of British politeness	165
6.6	1–3 tag sequences containing nouns	172
6.7	4–5 tag sequences containing nouns	174
7.1	The higher-order tags in the USAS tagset	179
7.2	Most key American semantic categories, compared to British English	180
7.3	Most key British semantic categories, compared to American English	181
7.4	American semantic tags with the highest CVs	182
7.5	British semantic tags with the highest CVs	183
8.1	Frequencies of items tagged as Z4 (discourse bin)	225