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.
American and British English

Divided by a Common Language?

Paul Baker

Lancaster University
Contents

List of Figures page viii
List of Tables xi
Acknowledgements 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-i-o- or -ae-i-oe- 39
   -ce or -se 40
   -il- or -ill-mme 43
   Dropped e 45
   -st 46
   -tion 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-
   -techn-
   -graph-
   -cess 69

v
# Contents

<table>
<thead>
<tr>
<th>Part</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Higher-Frequency Words</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Comparing the Varieties</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Densification</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Democratisation</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Informalisation</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Contrary Patterns: Who and Say</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>Two- and Three-Word Clusters</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>Lower-Frequency Words</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Identifying Candidate Items</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Transport(ation)</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>Monarchy</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Politics and Law</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>Business and Economics</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>People</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>Entertainment and Leisure</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>149</td>
</tr>
<tr>
<td>6</td>
<td>Part of Speech Categories</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Part of Speech Annotation</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>Key Comparisons and the Coefficient of Variation</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Decreased Use of the Passive Voice</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>Decreased Modality</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>Decreased Which in Relative Clauses</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>Noun Compounds, Titles and Genitives</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>175</td>
</tr>
<tr>
<td>7</td>
<td>Semantic Categories</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>Semantic Annotation</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>War and Law</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>Social Interactions and People</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>Technology and Media</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>Ideas and Concepts</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>203</td>
</tr>
</tbody>
</table>
## Contents

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

References 254

Index 261
### Figures

1.1 Frequencies of *so*  
1.2 Frequencies of *who*  
1.3 Frequencies of *didn’t*  
1.4 Mean word lengths  
1.5 Mean sentence lengths  
1.6 Standardised type token ratios  
2.1 Proportion of cases of *-our* in British English and *-or* in American English  
2.2 Proportion of cases of *-ise* in British English and *-ize* in American English  
2.3 Proportion of cases of *-isation* in British English and *-ization* in American English  
2.4 Proportion of cases of *-re* in British English and *-er* in American English  
2.5 Proportion of cases of *-ogue* in British English and *-og* in American English  
2.6 Proportion of cases of *-ae* in British English and *-a* in American English  
2.7 Proportion of cases where *practise* is a verb in British English and *practice* is a verb in American English  
2.8 Proportion of cases where *licence* is a noun in British English and *license* is a noun in American English  
2.9 Proportion of cases of *defence* and *offence* in British English and *defense* and *offense* in American English  
2.10 Proportion of cases of *-ll* in British English and *-l* in American English  
2.11 Proportion of cases of *programme* in British English and *program* in American English  
2.12 Proportion of cases of included *e* in British English and dropped *e* in American English  
2.13 Proportion of cases of *amongst* in British English and *among* in American English
List of Figures

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

6.3 American English frequencies of PPIS1 (first person singular subjective personal pronoun: 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 (wh- determiners) 166
6.7 Frequencies of that and which 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 man (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  
1.2 The Brown family  
1.3 Frequencies of the word *today* in the eight Brown corpora  
1.4 Frequencies of the word *want* in the eight Brown corpora  
1.5 Frequencies of the word *who* in the eight Brown corpora  
1.6 Tools used for analysis  
1.7 Sample concordance lines of *bank* from 1931 American English  
2.1 List of spelling differences examined in this chapter  
2.2 Summary of spelling patterns  
3.1 American morphemes with the highest CVs  
3.2 American morphemes with the lowest CVs  
3.3 British morphemes with the highest CVs  
3.4 British morphemes with the lowest CVs  
4.1 Keywords in American English  
4.2 Keywords in British English  
4.3 Words showing strongest growth and decline in American and British English  
4.4 American two-word clusters with high CVs  
4.5 British two-word clusters with high CVs  
4.6 Three-word clusters which show the most change  
5.1 Frequencies for *holiday* and *vacation*  
5.2 References to major British and American cities  
5.3 Frequencies for *jazz*  
5.4 Frequencies for *courgette* and *zucchini*  
5.5 Transport words (preferences)  
5.6 *Plane* and related words  
5.7 Transport words (frequencies)  
5.8 Terms relating to the monarchy  
5.9 American politics and law keywords  
5.10 British politics and law keywords  
5.11 Terms relating to business and economics  
5.12 Terms relating to people
xii List of Tables

5.13 Frequencies for relations 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
Acknowledgements

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