The first study of its kind, *Regional Variation in Written American English* takes a corpus-based approach to map over 100 grammatical alternation variables across the United States. A multivariate spatial analysis of these maps shows that grammatical alternation variables follow a relatively small number of common regional patterns in American English, which can be explained based on both linguistic and extra-linguistic factors. Through this rigorous analysis of extensive data, Grieve identifies five primary modern American dialect regions, demonstrating that regional variation is far more pervasive and complex in natural language than is generally assumed. The wealth of maps and data and the ground-breaking implications of this volume make it essential reading for students and researchers in linguistics, English language, geography, computer science, sociology, and communication studies.

- Identifies and maps regional linguistic variation in written Standard English for the first time.
- Introduces a corpus-based approach to dialectology.
- Presents a statistical method for identifying individual and common patterns of regional variation.

**Jack Grieve** is Senior Lecturer in Forensic Linguistics in the School of Languages and Social Sciences at Aston University in Birmingham, England. He holds a Ph.D. in Applied Linguistics from Northern Arizona University, where he studied quantitative corpus linguistics under the supervision of Professor Douglas Biber. He was also a postdoctoral research fellow in Professor Dirk Geeraerts’s Quantitative Lexicology and Variational Linguistics research unit at the University of Leuven in Belgium.
The aim of this series is to provide a framework for original studies of English, both present-day and past. All books are based securely on empirical research, and represent theoretical and descriptive contributions to our knowledge of national and international varieties of English, both written and spoken. The series covers a broad range of topics and approaches, including syntax, phonology, grammar, vocabulary, discourse, pragmatics and sociolinguistics, and is aimed at an international readership.

Already published in this series:

Irma Taavitsainen and Päivi Pahta (eds.): Medical Writing in Early Modern English
Colette Moore: Quoting Speech in Early English
David Denison, Ricardo Bermúdez-Otero, Chris McCully and Emma Moore (eds.): Analysing Older English
Jim Feist: Premodifiers in English: Their Structure and Significance
Steven Jones, M. Lynne Murphy, Carita Paradis and Caroline Willners: Antonyms in English: Construals, Constructions and Canonicity
Christiane Meierkord: Interactions across Englishes: Linguistic Choices in Local and International Contact Situations
Haruko Momma: From Philology to English Studies: Language and Culture in the Nineteenth Century
Raymond Hickey (ed.): Standards of English: Codified Varieties Around the World
Benedikt Szмrecsanyi: Grammatical Variation in British English Dialects: A Study in Corpus-Based Dialectometry
Daniel Schreier and Marianne Hundt (eds.): English as a Contact Language
Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis (eds.): The Verb Phrase in English: Investigating Recent Language Change with Corpora
Martin Hilpert: Constructional Change in English: Developments in allomorphy, word formation, and syntax
Jakob R. E. Leimgruber: Singapore English: Structure, Variation and Usage
Christoph Rühlemann: Narrative in English Conversation
Dagmar Deuber: English in the Caribbean: Variation, Style and Standards in Jamaica and Trinidad
Eva Berlage: Noun Phrase Complexity in English
Nicole Dehé: Parentheticals in Spoken English: The Syntax-Prosody Relation
Jock Onn Wong: English in Singapore: A Cultural Analysis
Anita Auer, Daniel Schreier and Richard J. Watts: Letter Writing and Language Change
Marianne Hundt: Late Modern English Syntax
Irma Taavitsainen, Merja Kyto, Claudia Claridge, and Jeremy Smith: Developments in English: Expanding Electronic Evidence
Arne Lohmann: English Co-ordinate Constructions: A Processing Perspective on Constituent Order
John Flowerdew and Richard W. Forest: Signalling Nouns in English: A corpus-based discourse approach
Jeffrey P. Williams, Edgar W. Schneider, Peter Trudgill, and Daniel Schreier: Further Studies in the Lesser-Known Varieties of English
Nuria Yáñez-Bouza: Grammar, Rhetoric and Usage in English: Preposition Placement 1500–1900
Jack Grieve: Regional Variation in Written American English

Earlier titles not listed are also available
REGIONAL VARIATION IN WRITTEN AMERICAN ENGLISH

JACK GRIEVE

Aston University
For Emily
Contents

List of figures  page xii
List of tables  xv
Acknowledgements  xvii

1 Introduction  1
   1.1 American dialectology  1
   1.2 Outline  13

2 Corpus  16
   2.1 Corpus-based dialectology  16
   2.2 Corpus design  19
      2.2.1 Register selection  19
      2.2.2 City selection  23
   2.3 Corpus compilation  27
   2.4 Corpus dimensions  29

3 Grammatical analysis  36
   3.1 Alternation variables  36
   3.2 Alternation variable selection and measurement  39
   3.3 Alternation variable mapping  43
   3.4 Feature set  47
      3.4.1 Pronouns  50
      3.4.2 Relative pronouns  52
      3.4.3 Determiners  52
      3.4.4 Adjectives  54
      3.4.5 Nouns  57
      3.4.6 Prepositions  58
      3.4.7 Particles  70
      3.4.8 Subordinators  72
      3.4.9 Coordinators  74
      3.4.10 Verbs  76
      3.4.11 Modals  78
      3.4.12 Infinitives  81
Contents

3.4.13 Not 82
3.4.14 Adverbs 83
3.4.15 Adverbials 87

4 Spatial analysis 98
  4.1 Standard approaches 98
  4.2 Regional linguistic data matrix 102
  4.3 Introduction to spatial analysis 103
  4.4 Local spatial autocorrelation analysis 107
  4.5 Spatial weights matrix 114
  4.6 Local spatial autocorrelation maps 122
  4.7 Discussion of results 122

5 Multivariate analysis 144
  5.1 Dialectometry 144
  5.2 Multivariate spatial analysis 147
  5.3 Spatially autocorrelated linguistic data matrix 150
    5.3.1 Alternation variable selection 150
    5.3.2 Missing data 154
  5.4 Factor analysis 156
    5.4.1 Introduction to factor analysis 156
    5.4.2 Preliminary analyses 161
    5.4.3 Results 166
  5.5 Fuzzy cluster analysis 175
    5.5.1 A note on dialect regions 175
    5.5.2 Introduction to fuzzy cluster analysis 176
    5.5.3 Results 178

6 Sources of regional linguistic variation 185
  6.1 Internal explanations 185
    6.1.1 Factor 1 185
    6.1.2 Factor 2 189
    6.1.3 Factor 3 193
    6.1.4 Regional variation in formality 195
  6.2 External explanations 198
  6.3 Comparison of internal and external explanations 207
  6.4 Comparison to previous American dialect surveys 209
  6.5 Future American dialect regions 216

7 Conclusion 219

Appendices
Appendix A: Corpus dimensions 227
Appendix B: Grammatical alternation variables: search patterns 255
Contents

Appendix C: Grammatical alternation variables: descriptive statistics 288
Appendix D: Spatial analysis results 299
Appendix E: Multivariate results 311
References 320
Index 330
Figures

1.1 American dialect regions: *Linguistic Atlas of United States and Canada*  
1.2 American dialect regions: *Dictionary of American English*  
1.3 American dialect regions: *Atlas of North American English*  
2.1 City sub-corpora  
2.2 Most common year of publication  
2.3 Words per sub-corpus  
2.4 Letters per sub-corpus  
2.5 Authors per sub-corpus  
3.1 *Be Not* contraction  
3.2 *Be Not* full  
3.3 *Although* alternation  
3.4 *Though* alternation  
3.5 *Even though* alternation  
3.6 *Be Not* contraction simplified maps  
3.7 *Although, Though, and Even Though* simplified maps  
3.8 Pronoun alternation maps  
3.9 Relative pronoun alternation maps  
3.10 Determiner alternation maps (Set 1)  
3.11 Determiner alternation maps (Set 2)  
3.12 Adjective alternation maps  
3.13 Noun alternation maps  
3.14 Preposition alteration maps (Set 1)  
3.15 Preposition alteration maps (Set 2)  
3.16 Preposition alteration maps (Set 3)  
3.17 Preposition alteration maps (Set 4)  
3.18 Particle alternation maps  
3.19 Subordinator alternation maps  
3.20 Coordinator alternation maps  
3.21 Verb alternation maps
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.22</td>
<td>Modal alternation maps</td>
<td>80</td>
</tr>
<tr>
<td>3.23</td>
<td>Infinitive alternation maps</td>
<td>82</td>
</tr>
<tr>
<td>3.24</td>
<td>Not alternation maps</td>
<td>84</td>
</tr>
<tr>
<td>3.25</td>
<td>Adverb alternation maps (Set 1)</td>
<td>86</td>
</tr>
<tr>
<td>3.26</td>
<td>Adverb alternation maps (Set 2)</td>
<td>88</td>
</tr>
<tr>
<td>3.27</td>
<td>Adverbial alternation maps (Set 1)</td>
<td>90</td>
</tr>
<tr>
<td>3.28</td>
<td>Adverbial alternation maps (Set 2)</td>
<td>93</td>
</tr>
<tr>
<td>3.29</td>
<td>Adverbial alternation maps (Set 3)</td>
<td>95</td>
</tr>
<tr>
<td>3.30</td>
<td>Adverbial position alternation maps</td>
<td>97</td>
</tr>
<tr>
<td>4.1</td>
<td>Histograms for the variants of selected alternations</td>
<td>104</td>
</tr>
<tr>
<td>4.2</td>
<td>*Anyone/<em>Anybody</em> map</td>
<td>111</td>
</tr>
<tr>
<td>4.3</td>
<td>*Anyone/<em>Anybody</em> local spatial autocorrelation map</td>
<td>112</td>
</tr>
<tr>
<td>4.4</td>
<td><em>Anyone</em> and <em>Anybody</em> simplified local spatial autocorrelation map</td>
<td>112</td>
</tr>
<tr>
<td>4.5</td>
<td>Comparison of spatial weights matrices for *Anyone/*Anybody</td>
<td>119</td>
</tr>
<tr>
<td>4.6</td>
<td>Comparison of spatial weights matrices for <em>Do Not</em> contraction</td>
<td>120</td>
</tr>
<tr>
<td>4.7</td>
<td>Local spatial autocorrelation maps</td>
<td>123</td>
</tr>
<tr>
<td>4.8</td>
<td><em>MAG</em> score by word class box plots</td>
<td>143</td>
</tr>
<tr>
<td>5.1</td>
<td>Interpolated local spatial autocorrelation map for <em>Rarely/Seldom</em></td>
<td>155</td>
</tr>
<tr>
<td>5.2</td>
<td>Additional amount of variance explained chart and scree plot</td>
<td>163</td>
</tr>
<tr>
<td>5.3</td>
<td>Factor 1 map</td>
<td>170</td>
</tr>
<tr>
<td>5.4</td>
<td>Factor 2 map</td>
<td>170</td>
</tr>
<tr>
<td>5.5</td>
<td>Factor 3 map</td>
<td>171</td>
</tr>
<tr>
<td>5.6</td>
<td>Factor 1 vs. Factor 2 cities</td>
<td>171</td>
</tr>
<tr>
<td>5.7</td>
<td>Factor 1 vs. Factor 2 states</td>
<td>172</td>
</tr>
<tr>
<td>5.8</td>
<td>Factor 1 vs. Factor 3 cities</td>
<td>172</td>
</tr>
<tr>
<td>5.9</td>
<td>Factor 1 vs. Factor 3 states</td>
<td>173</td>
</tr>
<tr>
<td>5.10</td>
<td>Factor 2 vs. Factor 3 cities</td>
<td>173</td>
</tr>
<tr>
<td>5.11</td>
<td>Factor 2 vs. Factor 3 states</td>
<td>174</td>
</tr>
<tr>
<td>5.12</td>
<td>Northeastern dialect region cluster</td>
<td>179</td>
</tr>
<tr>
<td>5.13</td>
<td>Midwestern dialect region cluster</td>
<td>180</td>
</tr>
<tr>
<td>5.14</td>
<td>Southeastern dialect region cluster</td>
<td>180</td>
</tr>
<tr>
<td>5.15</td>
<td>South Central dialect region cluster</td>
<td>181</td>
</tr>
<tr>
<td>5.16</td>
<td>Western dialect region fuzzy cluster</td>
<td>181</td>
</tr>
<tr>
<td>5.17</td>
<td>Overall dialect regions</td>
<td>182</td>
</tr>
<tr>
<td>5.18</td>
<td>Factor 1 vs. Factor 2 dialect regions</td>
<td>183</td>
</tr>
<tr>
<td>5.19</td>
<td>Factor 1 vs. Factor 3 dialect regions</td>
<td>183</td>
</tr>
<tr>
<td>5.20</td>
<td>Factor 2 vs. Factor 3 dialect regions</td>
<td>184</td>
</tr>
</tbody>
</table>
List of figures

6.1 Average formality factor scores (based on Factors 1, 2, 3) 198
6.2 Elevation 200
6.3 Population density (2011) 201
6.4 Per capita income by county (1999) 202
6.5 Percentage Republican presidential vote (2008) 203
6.6 Percentage German ancestry 204
6.7 Percentage African American (2010) 204
6.8 Census Bureau regions 205
6.9 Modern American dialect regions 210
6.10 Phonetic Factor 1 (Atlas of North American English reanalysis) 212
6.15 Future American dialect regions 218
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td></td>
<td>Letter to the editor examples</td>
<td>21</td>
</tr>
<tr>
<td>2.2</td>
<td></td>
<td>City sub-corpora</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td></td>
<td>The complete set of 135 grammatical alternations and their 295 variants</td>
<td>48</td>
</tr>
<tr>
<td>3.2</td>
<td></td>
<td>Pronoun alterations</td>
<td>50</td>
</tr>
<tr>
<td>3.3</td>
<td></td>
<td>English relative pronoun system</td>
<td>52</td>
</tr>
<tr>
<td>3.4</td>
<td></td>
<td>Relative pronoun alterations</td>
<td>53</td>
</tr>
<tr>
<td>3.5</td>
<td></td>
<td>Determiner alterations (Set 1)</td>
<td>55</td>
</tr>
<tr>
<td>3.6</td>
<td></td>
<td>Determiner alterations (Set 2)</td>
<td>57</td>
</tr>
<tr>
<td>3.7</td>
<td></td>
<td>Adjective alterations</td>
<td>59</td>
</tr>
<tr>
<td>3.8</td>
<td></td>
<td>Noun alterations</td>
<td>61</td>
</tr>
<tr>
<td>3.9</td>
<td></td>
<td>Preposition alterations (Set 1)</td>
<td>63</td>
</tr>
<tr>
<td>3.10</td>
<td></td>
<td>Preposition alterations (Set 2)</td>
<td>64</td>
</tr>
<tr>
<td>3.11</td>
<td></td>
<td>Preposition alterations (Set 3)</td>
<td>66</td>
</tr>
<tr>
<td>3.12</td>
<td></td>
<td>Preposition alterations (Set 4)</td>
<td>68</td>
</tr>
<tr>
<td>3.13</td>
<td></td>
<td>Particle alterations</td>
<td>70</td>
</tr>
<tr>
<td>3.14</td>
<td></td>
<td>Subordinator alternations</td>
<td>72</td>
</tr>
<tr>
<td>3.15</td>
<td></td>
<td>Coordinator alternations</td>
<td>74</td>
</tr>
<tr>
<td>3.16</td>
<td></td>
<td>Verb alternations</td>
<td>76</td>
</tr>
<tr>
<td>3.17</td>
<td></td>
<td>Modal alternations</td>
<td>79</td>
</tr>
<tr>
<td>3.18</td>
<td></td>
<td>Infinitive alternations</td>
<td>81</td>
</tr>
<tr>
<td>3.19</td>
<td></td>
<td><em>Not</em> alternations</td>
<td>83</td>
</tr>
<tr>
<td>3.20</td>
<td></td>
<td>Adverbs alternations (Set 1)</td>
<td>85</td>
</tr>
<tr>
<td>3.21</td>
<td></td>
<td>Adverbs alternations (Set 2)</td>
<td>87</td>
</tr>
<tr>
<td>3.22</td>
<td></td>
<td>Adverbial alternations (Set 1)</td>
<td>89</td>
</tr>
<tr>
<td>3.23</td>
<td></td>
<td>Adverbial alternations (Set 2)</td>
<td>92</td>
</tr>
<tr>
<td>3.24</td>
<td></td>
<td>Adverbial alternations (Set 3)</td>
<td>94</td>
</tr>
<tr>
<td>3.25</td>
<td></td>
<td>Adverbial position alternations</td>
<td>96</td>
</tr>
<tr>
<td>4.1</td>
<td></td>
<td>Structure of the regional linguistic data matrix</td>
<td>103</td>
</tr>
<tr>
<td>4.2</td>
<td></td>
<td>Top 25 alternations/variants by $MAG_i$ score</td>
<td>142</td>
</tr>
</tbody>
</table>
List of tables

5.1 Preliminary factor analyses 162
5.2 Factor loadings and variable communalities 167
6.1 Factor 1 variant loadings 186
6.2 Factor 2 variant loadings 190
6.3 Relative frequency of selected alternations PMW in COCA (Factor 2) 191
6.4 Factor 3 variant loadings 193
I began working on the research reported in this book ten years ago as part of my Ph.D. dissertation. I had finished my M.A. at Simon Fraser University in Vancouver with Paul McFetridge and Maite Taboada, and moved to Northern Arizona University in Flagstaff to work with Doug Biber. I was incredibly lucky to be supervised by Doug, who more than anyone else is responsible for where I am today as an academic. His influence can be seen throughout this book, both in the methods I have applied and the manner in which I have interpreted my results. I am also grateful for the support and guidance I received from the members of my doctoral committee, Bill Crawford, Ray Huang, and Randi Reppen.

Following my Ph.D., I became a researcher at the Quantitative Lexicology and Variational Linguistics (QLVL) research unit at the University of Leuven, headed by Dirk Geeraerts and Dirk Speelman, whose theoretical and methodological outlook on linguistics has also shaped this book. While at QLVL, I worked closely with Costanza Asnaghi and Tom Ruette. Many of the ideas and methods described in this book were developed during conversations with them, as well as with other members of QLVL, including Alena Anishchanka, Dirk De Hertog, Kris Heylen, Natalia Levshina, Pedro Pulquéro Vieira, Eline Zenner, and Weiwei Zhang.

After Leuven, I took up a lectureship at Aston University in Birmingham as a member of the Centre for Forensic Linguistics (CFL), which is where this book was written. I am particularly grateful to Tim Grant, the head of CFL, for his constant support, as well as my other colleagues at CFL, including Kate Haworth, Krzysz Kredens, and Andrea Nini, as well as Malcolm Coulthard, who founded the unit. I would also like to thank all my colleagues in the School of Languages and Social Sciences, especially Judith Baxter, Urszula Clark, Simon Green, Pam Moores, Garry Plappert, and Gertrud Reershemius.

The research reported in this book has also benefited from discussions with a number of linguists and social scientists from around the world,
especially the dialectometrists Wilbert Heeringa, John Nerbonne, Benedikt Szmrecsanyi, and Martijn Weeling, as well as with Federica Barbieri, Eric Friginal, Diansheng Guo, Roland Van Hout, Dan Johnson, Alice Kasakoff, Yuri Thierein, and Bert Vaux. I would also like to thank Mark Davies and Bill Labov for sharing their data with me, as well as their thoughts on this project.  

This book would also not have been possible without the support of Cambridge University Press, especially Helen Barton and Merja Kytö. I am grateful for the confidence they have had in this project and their feedback throughout this process has been invaluable. I would also like to thank Helena Dowson, Bethany Gaunt, Christina Sarigiannidou, and Martin Barr for their help preparing this book for publication.  

Finally, I would like to thank my family, especially my sister, Meg, my father, Tom, and my mother, Paula, who passed away while I was studying for my Ph.D. I would also like thank my father’s partner Linda and my parents-in-law Bob and Kathleen for their love and support. Most of all, I must thank my wife, Emily, to whom I dedicate this book. I met Emily in Flagstaff and we were married in Leuven. She has read every line of this book, most several times, and we have discussed every part of this research together. This book would not have been possible without her.