

University of Groningen

Quantitative assessment of English-American speech relationships

Shackleton Jr., Robert George; Shackleton, R.G.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Shackleton Jr., R. G., & Shackleton, R. G. (2010). Quantitative assessment of English-American speech relationships Groningen: s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Quantitative Assessment of English-American Speech Relationships

Robert G. Shackleton, Jr.



Groningen Dissertations in Linguistics 81
ISSN: 0928-0030
ISBN: 978-90-367-4318-1
ISBN: 978-90-367-4317-4 (electronic version)

©2010, Robert G. Shackleton, Jr.

Cover image: Peter Kleiweg
Printed by Wöhrmann Print Service, Zutphen, The Netherlands
Document prepared with \TeX nicCenter, \LaTeX 2 ϵ and typeset in pdf \TeX .

RIJKSUNIVERSITEIT GRONINGEN

Quantitative Assessment of
English-American
Speech Relationships

Proefschrift

ter verkrijging van het doctoraat in de
Letteren
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
donderdag 17 juni 2010
om 13.15 uur

door

Robert George Shackleton

geboren op 15 augustus 1956
te Gümligen bei Bern, Zwitserland

Promotores:

Prof. dr. ir. J. Nerbonne
Prof. dr. W. Kretzschmar, Jr.

Beoordelingscommissie:

Prof. dr. R. van Hout
Prof. dr. A. Lenz
Prof. dr. S. Mufwene

Acknowledgements

Many people living and dead have contributed to the research published in this dissertation and it would be impossible to acknowledge and thank them all. Nevertheless it is important to appreciate the ways in which research builds on the efforts of others and the full extent to which one's own accomplishments could only have been attained because of them.

My mother, Gilberte Ranvier Shackleton, inspired me to learn about languages in general, while my father Robert George Shackleton instilled in me a deep appreciation for history. Together they planted the seed of curiosity about dialects with stories of the late American dialectologist Henry Lee Smith, Jr., founder of the Foreign Service Language Institute (where he was my mother's supervisor), and his ability to determine American speakers' origins on the basis of subtle differences in their speech.

Though he passed from the scene long before I appeared, Guy Lowman painstakingly collected much of the data that was indispensable to my research. Many other fieldworkers contributed to the corpora on which I have drawn but Lowman played a pivotal role by collecting data in both the United States and England. I hope that he would be pleased that his work yielded the insights that it has. Peter M. Anderson produced *A Structural Atlas of the English Dialects* as his dissertation under Stanley Ellis at the University of Leeds and then disappeared entirely from the dialectology community. I cannot overstate the extent to which his work helped me make sense of the *Survey of English Dialects* corpus.

Professor Clopper Almon of the University of Maryland, College Park, my dissertation advisor in economics, taught me much of the mathematics I know, laying the basis for a great deal of the work published here, and inspired my curiosity about the origins of his northern Alabama planter's speech. Professor William Kretzschmar of the University of Georgia welcomed and encouraged my amateur's curiosity in American dialect origins, pointing me in the right direction toward basic linguistic materials and data sources and inviting me into the community of American dialectologists. Professor Salikoko Mufwene of the University of Chicago further encouraged me, inviting me to present early versions of this research and emphasizing the importance of language contact and founder effects. Bill and Sali also stongly encouraged me to publish some of this research at a time when I had concluded that it was relatively unimportant. Professors Alexandra Lenz and Roeland van Hout kindly served on my committee. Professor William Labov of the University of Pennsylvania owes my thanks for asking, at the LAVIS III conference in Tuscaloosa, Alabama in 2004, a key question that led to my development of feature-based principal

components analysis.

Professors Charles F. Meyer of the University of Massachusetts, Boston and Anne Curzan and Robin Queen of the University of Michigan gave me useful advice while editing papers for the *Journal of English Linguistics*, and Professors Michael D. Picone and Catherine Evans Davies of the University of Alabama did the same while editing a paper for the forthcoming *LAVIS III: Language Variety in the South: Historical and Contemporary Perspectives*.

Martijn Wieling of the University of Groningen served as my paranimf, helped me through the bureaucratic requirements of the University, painstakingly proofread an earlier draft of the thesis, and provided me with a Dutch translation of the summary. The process would have been infinitely more complex without his generous hand. Wilbert Heeringa also served as my paranimf, and his thesis (especially the sections on measuring segment distances acoustically) proved invaluable as a guide in my own research. Peter Kleiweg of the University of Groningen kindly made available the Rug/L04 software, helped me compose this dissertation in L^AT_EX, and designed the cover. Joseph Felsenstein made the PHYLIP software phylogeny inference package available, while Franz Manni, Etienne Guérard, and Evelyne Heyer made the Barrier software available. Carol Frost helped me use SAS/STAT to perform some of the principal component analyses in Chapter 2.

Last but certainly not least, my advisor and promotor Professor John Nerbonne has given me a great deal of extremely helpful advice over the years and ultimately offered me the most flattering recognition imaginable by inviting me to present this dissertation. To him I offer my grateful thanks for this unexpected, delightful honor.

Finally, it is incumbent on me to emphasize that the analysis and conclusions expressed in this dissertation are mine alone and should not be interpreted as those of my employer, the U.S. Congressional Budget Office.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Quantitative Tools and Phonetic Data	5
1.2.1	Quantifying Phonetic Variation	6
1.2.2	Distribution of Variants and Features	7
1.2.3	Distance Measures among Speakers	8
1.2.4	Cluster Analysis to Delineate Dialect Regions	8
1.2.5	Multidimensional Scaling to Map Dialect Regions	9
1.2.6	Principal Component Analysis	9
1.2.7	Multiple Regression and Monmonier Analysis	10
1.2.8	Discussion	11
2	The Traditional English Dialects	13
2.1	Introduction	13
2.2	Data	16
2.3	Variants and Features	20
2.3.1	Feature-Based Approach: Quantification	21
2.3.2	Feature-Based Approach: Distribution and Correlations	22
2.3.3	Variant-Based Approach: Quantification	23
2.3.4	Variant-Based Approach: Distribution and Correlations	25
2.4	Calculating Linguistic Distances	28
2.4.1	Segments	29
2.4.2	Aggregation	30
2.4.3	Results: Feature-Based Distances	32
2.4.4	Results: Variant-Based Distances	34
2.4.5	Middle English, Received Pronunciation, Western Reserve, and the <i>Survey of English Dialects</i>	35
2.5	Cluster Analysis and MDS	37
2.5.1	Multidimensional Scaling	38
2.5.2	Results	39

2.5.3	Phylogenetic Inference	49
2.6	Multiple Regression and Barrier Analysis	49
2.6.1	Multiple Regression Analysis	49
2.6.2	Distance Regressions	50
2.6.3	Variant-Area Regressions	55
2.6.4	Barrier Analysis	56
2.7	Principal Component Analysis	57
2.7.1	Feature-Based Results	59
2.7.2	Variant-Based Results	70
2.7.3	Variant-based Principal Components and Variant Frequency	78
2.8	Conclusions	79
3	English-American Speech Relationships	83
3.1	Introduction	84
3.2	Data	88
3.3	Choice of Quantitative Techniques	92
3.4	Cluster Analysis	94
3.5	The Distribution of Variants	98
3.6	Shared Variants	101
3.7	Linguistic Distances	108
3.8	Principal Component Analysis	112
3.9	Multiple Regression	119
3.10	Conclusions	126
4	Southern American English	131
4.1	Introduction	132
4.2	Data	135
4.3	Quantitative Techniques	137
4.4	Linguistic Distances	137
4.5	Cluster Analysis	141
4.6	Principal Component Analysis	145
4.7	Conclusion	152
5	Summary	155
5.1	The Usefulness of Quantitative Tools	155
5.2	The Traditional English Dialects	157
5.2.1	Linguistic Distances	158
5.2.2	Principal Components as Dialect Structures	160
5.2.3	Clusters as Dialect Regions	162
5.2.4	Regression Analysis and Dialect Geography	167
5.3	English-American Speech Relationships	168

5.3.1	Linguistic Distances	169
5.3.2	Clusters as Dialect Regions	169
5.3.3	Regression Analysis and Dialect Geography	170
5.3.4	Principal Components as Dialect Structures	170
5.3.5	Discussion	171
5.4	Future Directions	172
A	English Data for Chapter 2	175
B	English-American Data for Chapter 3	187
C	English-American Data for Chapter 4	205
	Bibliography	209
	Short Summary	215
	Samenvatting	221
	Groningen Dissertations in Linguistics (GRODIL)	227

List of Tables

2.1	Traditional Counties in the <i>Survey of English Dialects</i>	19
2.2	Correlations between Front and Back Long Vowels	27
2.3	Calculation of Linguistic Distance	31
2.4	Distance Regressions with Regional Dummy Variables	50
2.5	Difference between Regressions with and without Dummy Variables	54
2.6	Loadings: First Feature-based Unrotated Principal Component	60
2.7	Loadings: First Feature-based Rotated Principal Component	65
2.8	Loadings: First Variant-based Rotated Principal Component	72
3.1	Percentage of Total Variants Shared Between Regions	99
3.2	Percentage of Regions' Variants Shared Between Regions	99
3.3	Mean Shared Variants Between Speakers	103
3.4	Shared Variants Between Typical Speakers	103
3.5	Mean Linguistic Distance Between Speakers in Regions	110
3.6	Linguistic Distance Between Typical Speakers	111
3.7	Standard Deviation of Linguistic Distances	111
3.8	Loadings for the First Principal Component	115
3.9	Component Scores for the First Principal Component	116
3.10	Loadings for the Second Principal Component	118
3.11	Component Scores for the Second Principal Component	120
3.12	Regressions of Linguistic Similarity Region and Distance	123
4.1	Mean Linguistic Distance between Speakers in Regions	139
4.2	Linguistic Distance between Typical Speakers in Regions	139
4.3	Clustering of Southern American English Informants	146
4.4	First Principal Component for Southern American English Informants	150
5.1	English Regions and Their Distinctive Features	166

A.1	Words and Features in the Feature-Based Analysis	176
A.2	Variants Used in the Variant-Based Analysis	180
B.1	Variants and Their Vector Characterization	188
B.2	Speaker Characteristics	200
C.1	Speaker Characteristics	205

List of Figures

2.1	Localities of the <i>Survey of English Dialects</i>	18
2.2	Map from the <i>Structural Atlas of English Dialects</i>	24
2.3	Distribution of Linguistic Distances	33
2.4	Feature-based and Variant-based English Dialect Regions	40
2.5	Variant-based Linguistic Distances	43
2.6	Traditional Dialect Regions from Trudgill (1999)	48
2.7	Feature-Based Unrotated Principal Components	64
2.8	Feature-Based Rotated Principal Components	69
2.9	Variant-based Rotated Principal Components	71
3.1	Location of English and American Informants and Regions	89
3.2	Percentage of Informants Using Each Variant	102
3.3	Percentage of Shared Variants: Massachusetts	105
3.4	Percentage of Shared Variants: Eastern Virginia	106
3.5	Percentage of Shared Variants: Western Virginia	107
3.6	Linguistic Distance versus Shared Variants	109
3.7	Principal Components for English and American Dialect Features	121
4.1	Location of English and American Informants and Regions	134
4.2	Clusters Using Complete Linkage	142
4.3	Southern American Speakers: First and Second Principal Components	148
5.1	The Traditional English Dialect Regions	163

