The Linguistics of Speech

This insightful study proposes a unified theory of speech through which conflicting ideas about language might be understood. It is founded on a number of key points, such as the continuum of linguistic behavior, extensive variation in language features, the importance of regional and social proximity to shared linguistic production, and differential frequency as a key factor in linguistic production both in regional and social groups and in text corpora. The study shows how this new linguistics of speech does not reject rules in favor of language use, or reject language use in favor of rules; rather, it shows how rules can come from language as people use it. Written in a clear, engaging style and containing invaluably accessible introductions to complex theoretical concepts, this work will be of great interest to students and scholars of sociolinguistics, dialectology, and corpus linguistics.

William A. Kretzschmar, Jr. is Harry and Jane Willson Professor in Humanities at the University of Georgia.
The Linguistics of Speech

William A. Kretzschmar, Jr.
Contents

List of figures page vii
List of tables ix
Acknowledgments xi
The road not taken xii

Introduction 1

1 The contemporary marketplace of ideas about language 6
   The academic marketplace of ideas about language 6
   The popular marketplace for ideas about language: Writing and sounding correct 13
   The right language at home 18
   Ebonics: Correctness, rightness, and the marketplace of ideas about language in action 21

2 Saussure 31
   Saussure’s marketplace of ideas about language 33
   The aims of linguistics and linguistic structure 40
   The alternative to linguistic structure 44
   Foundations of the linguistics of speech 47
   Linguistic features in speech and structure 53
   Aggregation of evidence from speech 55
   A different linguistics 58
   Principles of the linguistics of speech 62

3 Evidence from linguistic survey research: basic description 64
   Findings from the LAMSAS survey: Boundaries and plots 66
   Findings from the LAMSAS survey: Lists and counts 82

4 Statistical evidence from linguistic survey research 104
   Findings from the LAMSAS survey: Density estimation (DE) 113
   Findings from the LAMSAS survey: Spatial autocorrelation 125
   Findings from the LAMSAS survey: Social categories 130

5 Evidence from corpus linguistics 146
   Firthian linguistics 148
   Corpus linguistics 151
Contents

Text types 158
Sampling text 162
Sampling documents 166
Conclusions 172

6 Speech as a complex system 174
  Complex systems 177
  State cycles and simulations 180
  Linguists and complexity 182
  Speech as a complex system 184
  Dimensionality 189
  Zipf and non-linear distributions in speech 190
  Scaling 198
  Speech and chaos 209
  Speech and evolution 211

7 Speech perception 218
  Prototypes and schemas 220
  Spatial perception 225
  Evidence from perceptual dialectology 228
  Perceptions of scaling 236
  Perceptions of the non-linear distribution of speech 242
  Perception and complex systems 247

8 Speech models and applications 251
  Towards a formal model of speech 252
  A less formal model 257
  Relationship with the linguistics of linguistic structure 260
  Speech and time 263
  Speech and public policy 271
  A last thought 277

References 278
Index 289
Figures

Figure 3.1  Kurath’s dialect map  page 67
Figure 3.2  Telsur dialect map  68
Figure 3.3  Northern diagnostic isoglosses  70
Figure 3.4  Kurath’s map of dragonfly variants  71
Figure 3.5  Northern isoglosses from Graddol, Leith, and Swann (1996: 271)  72
Figure 3.6  Pronunciations of the vowel in cow  73
Figure 3.7  Pronunciations of the vowel in find  74
Figure 3.8  darning needle variant for dragonfly  75
Figure 3.9  dragonfly variant for dragonfly  77
Figure 3.10  mosquito hawk variant for dragonfly  78
Figure 3.11  skeeter hawk variant for dragonfly  79
Figure 3.12  snake feeder variant for dragonfly  80
Figure 3.13  snake doctor variant for dragonfly  81
Figure 3.14  front room variant for parlor  82
Figure 3.15  parlor variant for parlor  83
Figure 3.16  living room variant for parlor  84
Figure 3.17  Pronunciations of fog containing [ɔ]  85
Figure 3.18  Pronunciations of fog containing [a]  86
Figure 3.19  cloudburst data (Table 3.1) charted by frequency of response  96
Figure 3.20  Adjusted cloudburst data (Table 3.2) charted by frequency of response  96
Figure 3.21  fog, plot of vowel 1a height (Table 3.3)  97
Figure 3.22  thunderstorm data, with chart by frequency  98
Figure 3.23  mantel data, with chart by frequency  99
Figure 3.24  fifth data, with chart by frequency (vowel 1a only)  100
Figure 3.25  first data, with chart by frequency (vowel 1a only)  101
Figure 3.26  night data, with chart by frequency (vowel 1a only)  102
viii List of figures

Figure 4.1 Plot of cloudburst from the LAMSAS cloudburst item, ‘heavy rain’ 114
Figure 4.2 DE plot of cloudburst from the LAMSAS cloudburst item, ‘heavy rain’ 115
Figure 4.3 DE plot of blinds from the LAMSAS blinds item, ‘roller window coverings’ 117
Figure 4.4 DE plot of lightwood from the LAMSAS kindling item 118
Figure 4.5 Unconstricted [r] in Thursday among LAMSAS speakers 119
Figure 4.6 Locations for 1162 speakers in LAMSAS 121
Figure 4.7 Kernel method DE plot for pail from the LAMSAS pail/bucket item (four probability levels) 122
Figure 4.8 Kernel method DE plot for pail from the LAMSAS pail/bucket item (two probability levels) 123
Figure 4.9 Nearest-neighbors method DE plot for pail from the LAMSAS pail/bucket item 124
Figure 4.10 Spatial autocorrelation (join-count) map for gully from the LAMSAS item for ‘washed out place in a field’ 127
Figure 4.11 LAMSAS speaker locations, Maryland 135
Figure 6.1 Plot (logarithmic) of Zipf’s Law in comparison to Mandelbrot’s improvement 191
Figure 6.2 Illustration of frequency of frequencies 193
Figure 6.3 mantel (adjusted data set), frequency of frequencies 196
Figure 6.4 Frequency of frequencies, “prediction” words 197
Figure 7.1 Two examples of Preston’s Draw-A-Map perceptual maps 230
Figure 7.2 Location of the Southern Dialect 231
Figure 7.3 Tamasis map of one respondent’s cognitive classification of perceptual dialects 232
Figure 7.4 Cluster analysis of New Jersey data (Tamasi 2003: 64) 234
Figure 8.1 Phase transition 265
Figure 8.2 A-curves at different moments in time (hypothetical) 266
Figure 8.3 A-curves at different moments in time, with associated S-curves (hypothetical) 268
Tables

| Table 3.1 | LAMSAS tally for cloudburst ‘a brief heavy rain’ | page 89 |
| Table 3.2 | List of types for cloudburst ‘a brief heavy rain’ | 91 |
| Table 3.3 | LAMSAS tally of vowel characteristics of fog | 93 |
| Table 3.4 | Atlanta Survey African American F1/F2 mean frequencies for [i] | 94 |
| Table 4.1 | Pooled within-groups correlation matrix for LAMSAS variables | 112 |
| Table 4.2 | Results of join-count analysis for 11 LAMSAS lexical features (60 Variants) | 128 |
| Table 4.3 | LAMSAS speaker characteristics, Maryland | 133 |
| Table 4.4 | Number of LAMSAS speakers classified into social variables | 136 |
| Table 4.5 | Number of LAMSAS speakers classified into geographical quadrats | 137 |
| Table 4.6 | Variable results for clearing from the LAMSAS clearing up item | 137 |
| Table 4.7 | Variable results for clearing off from the LAMSAS clearing up item | 139 |
| Table 4.8 | Variable results for clearing up from the LAMSAS clearing up item | 139 |
| Table 4.9 | Variable results for tokens containing fair from the LAMSAS clearing up item | 140 |
| Table 4.10 | Variable results for tokens containing break from the LAMSAS clearing up item | 141 |
| Table 4.11 | Variable results for tokens containing off from the LAMSAS clearing up item | 141 |
| Table 4.12 | Variable results for tokens containing no verbal particle (‘zero’ particle) from the LAMSAS clearing up item | 142 |
| Table 4.13 | Percentage of feature variants found significant for regional and social variables | 144 |

ix
x List of tables

Table 5.1 Frequency of top collocates 153
Table 5.2 undergo with its top twenty collocates 154
Table 5.3 Twenty tokens of undergo from the Brown and Frown corpora 156
Table 5.4 Frequency of coordinators in a corpus of speech and writing 164
Table 5.5 Univariate statistics for coordination corpus across 20 selections 164
Table 5.6 Correlation matrix for coordination corpus 165
Table 5.7 Distribution of documents by classification categories 167
Table 5.8 Words more frequent in tobacco documents 169
Table 6.1 chest of drawers data from three studies 199
Table 6.2 bureau, New York 202
Table 6.3 bureau, South Carolina/Georgia/Florida 203
Table 6.4 bureau, women speakers 204
Table 6.5 LAMSAS pronunciation of Baltimore (1034 responses) 205
Table 7.1 LAMSAS pronunciation of Asheville (594 responses) 238
Table 7.2 Rates of /l/ vocalization 240
Acknowledgments

I am grateful to my students over the years who have worked through these ideas with me, often as members of my language variation seminar. In particular, from among many excellent thinkers about language, I can mention Clai Rice, Matt Zimmerman, Allison Burkette, Susan Tamasi, and Joe Kuhl, as students who engaged seriously with this line of thought and, sometimes in collaborative articles and sometimes on their own, contributed to the development of the ideas presented here. I am also grateful to my colleagues, particularly Edgar Schneider, Chuck Meyer, and Lee Pederson (who collaborated with me on works and projects influential in the preparation of this book), as well as Salikoko Mufwene and John Nerbonne (whose views generally differ from my own), with whom I have carried on running conversations over many years about the ideas offered here. I must also thank Laura Wright, who has been most generous with her time and ideas for improvement of the work, and several anonymous readers engaged by Cambridge University Press whose comments have led to many real improvements in the text. Any mistakes in this book, of course, are my own problem, not theirs.

I cannot express enough my gratitude for her consistent support over many years to my wife Claudia, who has always had a keen interest in language as it is used in the real world. I am grateful to my son Russell, who continues to show me things that I did not know about language accommodation. And I am grateful to my son Brendan, himself a trained linguist, for the question to begin this book:

“What makes Ferdinand so sure?”
The road not taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I –
I took the one less traveled by,
And that has made all the difference.

Robert Frost, from Mountain Interval (1916)