Analyzing sociolinguistic variation

By Sali A. Tagliamonte


Reviewed by Terry Malone

Over the course of thirty years as a field linguist, I have noticed that a majority of linguists enter their field with background and training chiefly in the humanities. Unlike their colleagues who enter the field after working in the physical or biological sciences, most are not well-trained in the process of hypothesis formulation, experiment design, data collection, data recording, analysis of results, interpretation of analyses, and presentation of a research project in conferences or as publishable papers. The author (a professor of sociolinguistics at the University of Toronto) has written this book to guide her students in conducting analyses of sociolinguistic variation in a way that includes all these activities.

The book is part of the series “Key Topics in Sociolinguistics”. The advertising blurb on the inside page of the paperback and on Cambridge University Press’ internet site claims that “this is the first comprehensive, ‘how-to’ guide to the formal analysis of sociolinguistic variation. It shows step-by-step how the analysis is carried out, leading the reader through every stage of a research project from start to finish.” The book is indeed “comprehensive”, and this is perhaps its most outstanding characteristic. Further, each stage of the research process is carefully broken down into individual steps—another major strength of the book.

The author begins with a general introduction situating the field of variation analysis within sociolinguistics, and more generally within the field of linguistics (pp. 1-16). She briefly mentions the process of hypothesis formation: one might formulate a hypothesis before beginning the research project, based on available literature¹, or one’s observations in the course of collecting, processing, and/or analyzing data might lead to an interesting, testable hypothesis.

Chapters 2 (“Data collection”) and 3 (“The sociolinguistic interview”) concern themselves with data collection. The author recommends using “social networks” and “ethnographic methods” to contact informants, stating that the resulting data tends to be more natural than that collected using random sampling. The discussion of interview techniques and the interview schedule (Appendix B) provide valuable tips to beginners who just don’t have a clue. Indeed, all through the book one senses that the author has had rich experience in guiding raw beginners.

Chapter 4 (“Data, data, and more data”) is devoted to data processing, with special emphasis on data transcription and the development of a transcription protocol (see Appendix C). Chapter 5
(“The linguistic variable”) discusses properties of linguistic variables and how to recognize them. Basically, these are alternate forms such as “went slow” vs. “went slowly” or “goin[n]” vs. “goi[n], which seem to substitute for each other with little or no change in function or meaning, i.e. a “variable rule”—one that sometimes applies and sometimes doesn’t, for no apparent reason. In these cases one can suspect that sociolinguistic factors influence the variants, and this, of course, calls for the statistical methods presented in this book. “The prerequisites for variable rule analysis are : 1) choice, 2) unpredictability and 3) recurrence” (p. 130). The author recommends using linguistic literature (where possible) as an aid in identifying sociolinguistically significant variables.

Chapter 6 (“Formulating hypotheses/operationalising claims”) considers how one proceeds once a potential variable has been identified. The first step is to extract data (i.e. examples of the variable). In order to formulate a hypothesis, one must analyze “factors”, i.e. “some aspect of the context (either internal linguistic or external social) which affects whether or not a variant occurs” (p. 104). The analysis of factors is only as good as one’s coding system, so the author spends a good part of this chapter discussing how to code the data (i.e. code the context in which the data occurs so that a computer program can extract examples of the contexts in which the variants occur).

Chapter 7 (“The variable rule program: theory and practice”) introduces the “Goldvarb series of programs” which the author and her students use to analyze sociolinguistic variation. On p. 128 the reader finds the web addresses for downloading free versions of the program. According to Tagliamonte, the chief advantage of the Goldvarb program lies in its ability to account for “the extreme distributional imbalances” so typical of natural human language (p. 133). Essentially, one cannot guarantee the presence of contexts in which the variable optionally occurs, with the result that there will be gaps in the data; unless taken into account, these gaps skew the results. The author describes the advantages of variation analysis quite nicely:

The apparent randomness of the choice process makes it appear that the variation has no structure and has many more exceptions than it really does. Statistical inference by its very nature extracts regularities and tendencies from data presumed to have a random component. In order to accomplish this, the inference procedures must be applied to some sample containing the outcomes of the choice repeated many times (your token file/data) usually in a variety of contexts, each context being defined as a specific configuration of conditioning factors (your coding schema) (p. 131).

The program makes provision for “binomial one-step” (all factors co-occurring with a variable are analyzed at once) and “binomial step-up/step-down” analysis (contextual factors are either added one by one or subtracted one by one as the analysis is repeated, allowing the researcher to determine which factors are more significant in influencing the variable). Supposedly, “a series of diagrams on the companion website schematises the operation of the variable rule program” (p. 140); this would have been really helpful for following the discussion in this and the next three chapters, but unfortunately, when I went to the companion website I couldn’t find the diagrams.
In chapters 8 (“The how-to’s of a variationist analysis”), 9 (“Distributional analysis”), and 10 (“Multivariate analysis”) Tagliamonte walks her reader through the various steps involved in using the program to analyze a variant, using examples from some of her data. Most of these chapters are taken up with details involving the manipulation of the Goldvarb program, including setting up input files (“token” files), modifying condition files (specify “precisely how the data in the token file are to be configured”, p. 160), reading “cell files” (output of the condition files and input into the Goldvarb program, p. 162), and interpreting results files. The analyst must write the condition files in LISP (a programming language), probably not attractive to some researchers; however, the exposition demonstrates that this allows experimenting with contextual factors and reveals hidden correspondences in ways that might not otherwise be possible.4

Throughout Chapters 7, 8, 9, and 10 Tagliamonte consistently ties the manipulation of the program to linguistic principles and considerations; this is a major strength of the book. The text is rich (here and throughout) in caveats and tips (the latter enclosed in gray boxes). Chapters 10 (“Interpreting the results”) and 11 (“Finding the story”) aid the student or reader in interpreting and presenting the results to a professional audience—the author even includes a model outline of a research paper as the last exercise (p. 262).

Certain small details detract from the overall positive nature of this work. For instance, chapter subsections are not listed in the book’s table of contents.5 Although chapter headings are reproduced at the head of each page, the chapter number is not included as part of the heading. The author often refers to other chapters by number in the course of her discussion; consequently, the lack of numbers in the headings results in quite a bit of page-flipping to check out any given reference. The author does provide a brief statement of each chapter’s purpose at its beginning and a summary with a concluding exercise at the end, all of which is helpful.

I was amazed to find almost no typos in this book.6 Instead of being printed at the end of the last chapter, the appendix appears on-line. I looked through the entire book without finding the appendix’s web address—and then after I’d finished reading it I found it on the back cover (paperback version).7 No subtitles occur in the on-line list of appendices, so I reproduce them below for the convenience of the reader.

1. Appendix A: Informational letter and informed consent form
2. Appendix B: Interview schedule: Guideline questions
3. Appendix C: Transcription protocol
4. Appendix D: Coding verbs
5. Appendix E: Variable (t,d), internal and external factors together
6. Scattergram

No reason is given for putting the appendix on-line, but I assume it allows modification of material still under development in a way that permits readers to access it without waiting for a second edition—not a bad idea. A glossary of terms occurs at the end of the book (pp. 262-266), also a good idea, especially since the author does not always clearly define her terminology.8 As I read chapters 7, 8, 9, and 10, I found myself wishing for a solid background in statistical analysis (or a good statistics text at hand). I didn’t have access to the references listed in the introduction of chapter 7 (pp. 128-129), which might have helped to remedy this lack.
Tagliamonte repeatedly refers to the Goldvarb documentation (Rand and Sankoff 1990); if I read the book again, or use it as a guide for a variation analysis, I would print out the program documentation (the web address for the Macintosh version is listed in the book’s bibliography under Rand and Sankoff 1990), in order to follow the exposition in these four chapters with more ease.

The faults of this book are minor, especially when compared to its virtues. I cannot speak authoritatively for the field of sociolinguistics, but I have seen little comparable elsewhere in linguistics wrapped up into one single book—conducting students through an entire research project from beginning to end and relating individual steps in the use of a statistical/mathematical tool to concrete linguistic phenomena and linguistic theory. This book will be a must in advanced sociolinguistic courses, and certainly a must for any sociolinguist studying variation. Further, any graduate student in linguistics or field linguist wanting to know how to design, conduct, and present a research project will profit by studying this book.

Notes

1 Throughout the author emphasizes the importance of knowing the literature: “You must be prepared to interpret any result vis-à-vis the observations and claims in the literature” (p. 125).

2 Some examples of asymmetric distribution of variable rules are given on p. 181.

3 The author cites Sankoff 1988 here.

4 One does not need to be a sophisticated LISP user, if one follows the models in the book.

5 The table of contents is available on the internet at http://www.cambridge.org/9780521771153 (click “Look Inside”, then “View frontmatter as PDF”). Subsections are not listed on-line either.

6 P. 131 (“that” instead of “than”—corrected in the quote above), p. 170 (“would usually never”—not a typo, but the combination looks strange), p. 229 (the period in the second line should be a colon); p. 236 (the abbreviation in the last line should be NPE instead of NPR).

7 Shouldn’t this address (www.cambridge.org/9780521778183, then click “Resources”) be somewhere in the body of the book, preferably in the position normally reserved for an appendix? I can’t help wondering where I would find it in the hardback version (only the paperback was available to me); hopefully it is not just on the dust jacket.

8 The term “binomial step-up/step-down” was one example (see p. 140). I eventually figured out what it meant by studying examples in later chapters.

9 On p. 160 this is referred to as “the Goldvarb 2.0 user’s manual” and on p. 175 as “the Goldvarb 2.0 manual”. The reference is listed on p. 128 as a place to download the Goldvarb program, with no mention made of the documentation. (Documentation for Windows and DOS implementations in past could be found at www.unh.edu/linguistics/lab/goldvarb.html,
along with the manual for the Macintosh implementation, also software for all three implementations, however, that URL is apparently no longer accessible). The version for Windows is referred to as Goldvarb 2001; Goldvarb 2.0 and 2.1 are Macintosh versions. The program was originally written for Macintosh under the name VARBRUL.

References
