Book Notice
Phonetica 2012;69:274–275
DOI: 10.1159/000349926

Keith Johnson
Quantitative Methods in Linguistics
Blackwell, Oxford 2008
296 pp.; Pb GBP 26.99/Hb GBP 70.00

The increasing tendency towards quantitative empirical research in many modern linguistic areas (dialectology, corpus linguistics, and construction grammar, to name but a few) heighten the need for solid, dedicated introductions as well as a sound general quantitative methodology that unifies those manifold new areas.

While it can be argued that there is already a sufficient number of books on statistics available, the selection of texts specifically aimed at linguists is still rather sparse; additionally, most of them specialize in specific areas of linguistics or provide rather short instructions for the application of selected techniques within the framework of conducting a complete empirical study [Di Paolo and Yaeger-Dror, 2010]. Keith Johnson’s Quantitative Methods in Linguistics, however, fills the gap between these elementary linguistic introductions on the one hand and extensive handbooks on statistics (that aim at a more general audience) on the other hand.

The strategy of the book is twofold, basically making it two texts at once: one is a practically oriented introduction to statistical methods for linguists, the other is a concise workbook for the free statistics software environment/programming language R, allowing the reader to reproduce the techniques instantly. Although both parts are meant to be used in conjunction with each other, the grey shaded R passages (which make up about one fifth of the book) can easily be skipped without disrupting the flow of the main text.

The book is structured into seven chapters. From the third chapter onwards, each chapter covers a different sub-discipline of linguistics (namely phonetics, psycholinguistics, sociolinguistics, historical linguistics, and syntax): every technique is introduced using examples taken from the discipline which is most common for the actual method. However, Johnson has no intention of hiding that he is a phonetician at heart, as he frequently chooses data for his examples that are more or less directly related to phonetics. Each chapter builds upon the preceding ones, at least to a certain degree, and closes with exercises.

The introductory chapter provides a concise rundown of the ‘hows’ and ‘whys’ of quantitative analyses as well as basic types of data distributions and interpretations. As with the rest of the book, novices at statistical methods might need to take a second read: this chapter already explains interpretation techniques such as quantile-quantile plots, subjecting data to the arcsine transformation or standardizing data to z-scores via the standard deviation.

The second chapter on ‘patterns and tests’ explains the principles of sampling, hypothesis testing and correlation as well as regression. As an example, authentic data on VOT and F1 values are used.

The third chapter focuses on phonetics; first, a t-test is applied to two sets of VOT data to check different hypotheses concerning their relation. The chapter then introduces multiple regression for predicting the position of the tongue root from measurements taken of the front of the tongue, before extracting factors for tongue position using principal components analysis.

With a focus on sociolinguistics, chapter five extends the view to nominal data, starting with Pearson’s $\chi^2$ test and moving on to maximum likelihood estimation and logistic regression. The closing pages of this chapter provide a worthwhile comparison between the Varbrul software package (commonly used for logistic regression in a lot of sociolinguistic research) and R.

The book is available at www.karger.com/pho
The following chapter on historical linguistics reaches way back to the examination of cognates in Indo-European languages to introduce basic hierarchical clustering, string edit distance [nowadays primarily used in dialectology/dialectometry; see Nerbonne and Heeringa, 2010], and multidimensional scaling, basically making this chapter the beginner’s toolbox for dialectometry as well.

The final chapter (syntax) makes use of magnitude estimation for measuring sentence acceptability as well as linear mixed-effects modeling (fixed and random effects) to predict syntactic phenomena from large text corpora. The book closes with an appendix on how to conduct a magnitude estimation experiment.

Meanwhile, the R portions of the book proceed like this: right from the start, for a lot of the figures in the book the R code that was used to generate it is provided. The first chapter introduces the basic plotting of graphs and functions and the concept of data vectors, the second shows how to read external data and functions from files, program basic loops and create new commands. Afterwards, we learn how to choose subsets of the data (chapter 3), preprocess external data for use in R and incorporate external packages (chapter 4). The following chapters show how to deal with matrices and dwell upon the use of tests and the refining of graphs. In the final chapter, the author shows how to create tables, split data, and perform iterations on the data.

In summary, Johnson’s hope that the reader would be able to ‘move from this introduction – oriented explicitly around linguistic data – to more general statistics reference books’ (xi) has good chances of succeeding: not only does his book offer a versatile and inspiring view on both basic and not-so-basic quantitative methods, it is also both technically sound and faithfully consistent. The author calls a spade a spade instead of merely demonstrating applications, thus providing a good basis for a unified quantitative methodology in linguistics that agrees with other disciplines as well. Concerning style I find it noteworthy that Keith Johnson knows how to write decent scientific prose; he knows when to be entertaining and when not. This turns a dense book on statistics into surprisingly good reading. Moreover, the overall number of typing errors is remarkably low (although, on a slight downside, some of them cumulate in the R code in the last third of the book).

The blurb claims that the book is ‘an ideal text for students’. This addressed audience might as well be extended: Keith Johnson’s Quantitative Methods in Linguistics is a missing link between introductions and handbooks, not only for students with a basic knowledge of statistics, but also for advanced researchers who wish to broaden their knowledge on both detail and diversity in quantitative studies.

Simon Pröll, Augsburg, Germany

References