Globe: A Journal of Language, Culture and Communication, 2: 139-143 (2015)

The gentle art of number crunching in linguistic research

Sebastian M. Rasinger, *Quantitative Research in Linguistics: An Introduction*, 2nd ed. (Series: Research Methods in Linguistics). London: Bloomsbury, 2013, xii + 286 pp., ISBN 978-1-4725-6697-3.

Kim Ebensgaard Jensen, Aalborg University, Denmark.

1. Introduction

Linguistics is a member of the humanities family, so to speak, and the humanities are not normally associated with mathematics and statistics. However, there are several branches within linguistics where quantitative analysis is central, such as sociolinguistics, psycholinguistics, corpus linguistics, and usage-based linguistics.

With the recent empirical and quantitative turns in linguistics, quantitative analysis is only going to become more commonplace, and, consequently, knowledge of application of quantitative methods of analysis is going to be increasingly vital to students of language and linguistics in the years to come. While there are several excellent and informative volumes out there that serve to introduce linguists to statistical methods and tools, most of these have rather steep learning curves and may even have a deterrent effect on many students of linguistics who are not well-versed in mathematics.

The thing is that, although quantitative analysis is becoming increasingly important in linguistics, it is still the case that a majority of students have not received any training in statistics beyond whatever mathematics they learned in primary and secondary school. Thus, there is a real need for a very basic primer in statistics for students of linguistics who have little or no prior knowledge of statistics. Sebastian M. Rasinger's *Quantitative Research for Linguists: An Introduction* (henceforth, QRIL) was originally published in 2008 and re-published in a second enhanced edition in 2013 to fill this gap.

2. Synopsis

QRIL contains eleven chapters, including an introduction, which are distributed over three parts. Part one, which consists of chapters two to four, covers the basics of quantitative research in linguistics and introduces some elemental concepts and methodological issues, such as research design. Part two, which covers chapters five to nine, introduces the reader to a range of statistical methods and their applicability in linguistic research. Finally, part three includes chapters ten and eleven. The former deals with three advanced statistical methods, while the latter is a collection of appendices and solutions to exercises.

2.1. The basics

The first chapter is not included in part one, but serves as an introduction to the entire volume. In the introduction, taking his starting point in his own experience, Rasinger makes a good case for the importance of understanding quantitative methods as a linguist. He also specifies the three target reader types of the book: linguists and students who are simply afraid of using quantitative methods, linguists and students who know how do use quantitative methods but who do not know how to do it on a computer, and linguists and students who need to know about quantitative methods as part of their research.

Chapter 2 'Quantitative research – some basic issues' introduces the essential aspects of quantitative research in linguistics, contrasting it with qualitative research. Using /h/-dropping as an example, Rasinger introduces the reader to the notion of quantifiable variables, arguably the most

essential aspect of quantitative research in language studies. The quantitative-deductive approach is described in detail, and Rasinger also discusses which types of research call for quantitative analysis and which types call for qualitative research. The chapter goes into more detail with variables, operationalization, and measurement and also discusses reliability/validity as well as the relation between theories, hypotheses, and laws.

After this general introduction to quantitative research, Chapter 3 'Research design and sampling' takes a gentle step towards the more practical aspects of quantitative linguistic research, as it discusses longitudinal, cross-sectional, experimental, and quasi-experimental designs. Useful examples of these designs at work in actual research are provided throughout this chapter. In addition to discussing research designs and sampling, this chapter contains a section on research management – something which is often left out of books within the same genre as QRIL.

Chapter 4 'Questionnaire design and coding' introduces the reader to the craft of designing questionnaires that are suitable for linguistic research. A point which is made early on is that making questionnaires is not a simple procedure; it takes a lot of effort, because, as Rasinger rightly reminds us, questionnaires are a scientific tool and should be treated as such.

2.2. Getting to work

It is now time for the readers to get their hands dirty and work with data and quantitative analysis.

Starting Chapter 5 'A first glimpse at data' gently, the reader is introduced to simple addition (accompanied by an infobox describing how it is done in Excel) which is followed by a discussion of absolute and relative frequencies. This is where the reader gets their first taste of real mathematics in the form of calculations and descriptions of relative frequencies and ratios. Next are the slightly more abstract and complex notions of classes, width, and cumulative frequencies.[1] Lastly, the reader is introduced to the use of graphs in quantitative linguistic research, and a range of graphs and charts, such as bar charts, pie chars and line charts are discussed.

As the title indicates, Chapter 6 'Describing data properly', QRIL now moves into the territory of deeper treatment of quantitative data. More specifically, this chapter focuses on descriptive statistics techniques. Under the heading of measures of central tendency, Rasinger discusses means (arithmetic and trimmed arithmetic means), median, and mode. This is followed by a brief discussion of quartiles, quintiles and percentiles. The following section on measures of dispersion takes the reader through range, variance, standard deviation, and *z*-scores. Finally, fairly detailed discussions of normal distribution and the standard error are offered.

Chapter 7 'Analysing Data – a few steps further' focuses on probability, with probability theory being addressed in some detail, and introduces the reader to various more advanced measures such as the chi-squared test, Pearson correlation, and R^2 as well as regression. This chapter also contains a brief and to-the-point discussion of significance and one on the notion of causality (the latter in relation to R^2).

Under the heading of 'Testing hypotheses', Chapter 8 discusses the linkage between hypotheses and causality and introduces a range of tests, focusing on *t*-tests. However, the *F*-test and ANOVAs are also introduced, and the chi-squared test is revisited.

Lastly, in Chapter 9 'Analysing non-parametric data: When things are not quite normal', Rasinger turns to non-parametric tests Spearman Rank correlation test, Kendall's Tau, Wilcoxon Signed Rank test and, briefly, the Mann-Whitney *U* test.

1 Be advised that, if you use other spreadsheet software than Excel (I used Calc), there are slight differences in the operation of the formula bar. For instance, in automatically constructing a frequency table using the FREQUENCY function, I had to use a semicolon to separate the arguments of the function rather than a comma as described in QRIL. Moreover, readers should be informed that QRIL is written for the English setup of Excel, and the functions are language specific. Thus, the FREQUENCY function is, FREKVENS in the Danish version, FREQUNTIE in the Dutch version, and TAAJUUS in the Finnish version.

2.3. A peek ahead

Part three consists of two chapters only. With Chapter 11 being a list of solutions, there is in effect only one very short chapter in part three – namely, Chapter 10. Entitled 'Beyond the basics: Other methods, other tools', this chapter briefly introduces MANOVAs and meta-analysis. While the section devoted to the former is rather short, the section devoted to the latter provides a slightly more detailed, step-wise description of meta-analysis. Chapter 10 concludes with a quick summary of the advantages of using spreadsheets and also offers short commentaries on SPSS and R. Short though it is, this chapter peeks ahead at the tools and challenges available to those readers who should choose to venture into advanced quantitative linguistics, using more advanced digital tools and resources.

3.4. Whistles and bells

QRIL comes with a companion website,[2] which contains the following supplementary material:

- Voiced-over flash video clips demonstrating use of Excel spreadsheets. For instance, three flash demos accompany chapter five. One guides the reader/viewer through the basic functions of addition, subtraction, multiplication and division in Excel. Another one demonstrates how to make a frequency table, and the third one demonstrates how to calculate relative frequencies. Seeing that most of the clips draw on data in the chapters in the book that they accompany, the reader is advised to watch them after reading the chapter in question. Moreover, since doing quantitative analysis is a matter of practical work, readers are also advised to open Excel themselves and replicate what is demonstrated in the clips.
- PowerPoint presentations covering a range of central topics in the book, such as questionnaires, dispersion, research designs, and correlation. The readers can download these from the website and click through the presentations at their own pace.
- Crib sheets for chapters 4-9, offering concise reference-friendly information on topics and operations treated in the book. These are very useful for reference, and I can imagine that the best way to use these crib sheets is to print them out and keep them in a handy binder that one can consult whenever needed.
- Excel templates for "descriptives", chi-squared tests, partial correlations, t-tests, and meta-analysis calculations. The reader can enter their own values into these templates, and Excel will perform the calculations in question. While these are really helpful, it is recommended that the readers do all the steps required themselves so as to get the practical experience and, for lack of a better term, feel. The templates are actually useful as a means to this end, because the reader can click the different active cells in the spreadsheets and see the underlying operations.

While this is supplemental material, and one can technically benefit from QRIL without making use of the companion website, using the material available here is very likely to considerably facilitate the learning process which is the intended outcome of the book. In particular, readers who are not well versed in the use of spreadsheets are recommended to access Rasinger's flash demos at the very least.

2 Available here http://bloomsbury.com/cw/quantitative-research-in-linguistics/. Be advised that it is necessary to sign up for a free Bloomsbury user account to access the companion website.

3. Evaluation and discussion

Unlike many other introductions to quantitative research in language studies that I have come across, Rasinger's book maintains a high level of pedagogy throughout such that the learning curve never becomes too steep.

For instance, the quantitative-deductive approach is carefully described in the second chapter, as the reader is guided from the hypothesis statement through methodology development and data collection to verification or falsification of the individual hypothesis. This process sounds simple enough, but, to many students within the humanities, where the qualitative approach is more widespread, this way of doing research may be completely alien. Thus, it is very important that Rasinger establishes the difference between qualitative and quantitative research, and gently describes the latter, as early in his book as he does. In QRIL, Rasinger manages to maintain the difficult balance between detail and conciseness in his descriptions, which is another pedagogical strength. As an example of this, consider his description of the notion of research design in the introduction to Chapter 3:

> *Research design* is best described as the actual *structure* according to which our study is organized. As such, together with the theoretical grounding the design forms an important part of the overall methodological-analytical framework which we use to answer our research questions, and prove or disprove our hypotheses. Research design does not, though, refer to the actual instruments we use in our investigation (such as questionnaires or interviews), although the relevance of good interplay between hypotheses, existing theory, methods and design cannot be emphasized enough; and there is a particularly strong connection between design and instruments chosen. (p. 35)

Here, Rasinger describes clearly and concisely what 'research design' means how it relates to other aspects of a quantitative linguistic study. Appearing in the introduction to the chapter on research design and sampling, this key definition gives the reader a foundation to build as the reader progresses through the chapter. This passage also illustrates the clear and accessible language that Rasinger uses; any freshman with just a semester's worth of exposure to typical academic discourse in English should be able to easily decode. Still, while the learning curve remains relatively flat, the two last chapters in Part 2 may strike some readers as particularly abstract, and the help of an instructor might be required to appreciate them.

Throughout the book, the contents are supplemented by helpful graphs, illustrations, and tables. Chapter 3, in particular, contains extremely helpful illustrations to assist the reader understand the principles of sampling. The book is also rich in illustrative tables. For example, at table in Chapter 3 provide overviews of the pros and cons of research design types and, at the end of Chapter 4, the reader is taken through a fictional questionnaire designed to collect feedback on QRIL itself, pointing out its strengths and some deliberately inbuilt weaknesses as well.

The accompanying flash demos are particularly valuable because they not only show how to perform calculations and make graphs in Excel, but they also show that Excel is actually considerably simple to use. I think this is very important for students in the humanities to see, because – at least in my experience – students in the humanities are generally not very experienced in the use of spreadsheets. In a more long-term perspective, seeing that spreadsheet skills are useful not only in quantitative linguistics, but in a range of possible future careers of humanities students, the clips and QRIL can help the reader to acquire very useful transferable skills. The PowerPoint presentations available on the companion website summarize many of the topics in the book in the form of key points. While this is definitely useful in a learning perspective, an even more valuable addition to the companion website could have been eLectures based on the PowerPoint presentations, in which the points in the PowerPoint presentations were accompanied by actual

explanations by Rasinger in the form of either voice-over (like the Excel flash demos) or perhaps even his talking head.

It should be pointed out that, while QRIL introduces a wide range of statistical analytical techniques, it is not exhaustive and will be more useful to students of sociolinguistics, psycholinguistics, and language acquisition, for example, than students of corpus linguistics, as most of the techniques preferred by corpus linguists are left out. Therefore, I would recommend that instructors in corpus linguistics seek out other coursebooks. Chapter 2, however, would be suitable reading material for the first session in any course in quantitative linguistics – even corpus linguistics.

Overall, the second edition of Sebastian M. Rasinger's *Quantitative Research in Linguistics: An Introduction* is an attractive introductory book to quantitative analytical techniques that can be used in linguistics, and, furthermore, it gently introduces the reader basic principles in statistics. It would be a good choice for a coursebook in general quantitative linguistics and can be used at both undergraduate and postgraduate levels. It is so accessible that advanced students and professional linguists should be able to tackle it on their own (in particular if they make use of the companion website as well), while undergraduates might need some guidance from their teacher or a tutor. Easy to read and a very gentle introduction to the use of spreadsheets in quantitative linguistics, Rasinger's book should be particularly attractive to instructors, students, and other readers who are not necessarily interested in using superior platforms such as R, Python or SPSS software.