

- Chomsky, Noam. 2019. Puzzles about phases. In *Linguistic variation: Structure and interpretation*, ed. Ludovico Franco and Paolo Lorusso, 163–168. Berlin, Boston: De Gruyter Mouton.
- den Dikken, Marcel. 2007. Phase extension. *Contours of a theory of the role of head movement in phrasal extraction*. *Theoretical Linguistics* 33(1): 1–41.
- Gallego, Angel. 2010. *Phase theory*. Amsterdam: John Benjamins.
- van Urk, Coppe, and Norvin Richards. 2015. Two components of long-distance extraction: successive cyclicity in Dinka. *Linguistic Inquiry* 46(1): 113–155.

**Garcia, Guilherme D.** 2021. *Data visualization and analysis in second language research*. New York, NY: Routledge. Pp. 286. USD \$44.95 (softcover). Also available electronically.

Reviewed by Ariane Sénécal and Laura Sabourin, *University of Ottawa*

*Data visualization and analysis in second language research* grew out of workshops developed and delivered by Guilherme D. Garcia at McGill, Concordia, and Ball State universities. These origins are evidenced by the conversational tone and exercise-based approach used throughout the book. Garcia aims to train readers to use the R programming language (through the R Studio interface) to construct robust statistical analyses while producing clear and informative plots and figures. Instead of traditional *t*-tests and ANOVAs, this book focuses on regression analyses, hierarchical models (a.k.a. mixed-effects models), and Bayesian statistics.

Part I, “Getting Ready”, includes a straightforward introduction detailing some of the author’s rationale and offers a brief review of some fundamental notions: Garcia assumes his readers have some basic knowledge of statistics (e.g., sample vs. population means, *p*-values, effect sizes, confidence intervals, standard errors, *t*-tests, ANOVA) but no prior experience with R or the more advanced statistics detailed in later sections. The second chapter, “R Basics”, explains why this specific program is preferred over others and walks readers through the R/R Studio setup, including the installation of important packages and some initial calculations and script blocks, as well as more general concepts relating to data organization (e.g., tidy data, as in Wickham 2014).

Part II, “Visualizing the Data”, includes chapters 3–5: “Continuous Data”, “Categorical Data” and “Aesthetics: Optimizing Your Figures”. It focuses on the presentation of data through plots and figures, which Garcia rightly insists need to communicate clearly the focus of the research and should help inform decisions relevant to later analyses. In this section the author aspires to convince readers of “the numerous advantages of visualizing patterns *before* statistically analyzing them” (p. 239). Some of the technical programming aspects found within these chapters include transforming binary/categorical data into continuous (hence plottable) variables, and the use of facets (i.e., layers) that allow the combination of plot types and the improvement of figures’ explanatory potential.

Part III, “Analyzing the Data”, constitutes the most dense and demanding portion of the book. It begins with generalized linear models in chapters 6–8: “Linear

Regression”, “Logistic Regression”, and “Ordinal Regression”. Here Part III runs parallel to Part II, first handling continuous data before adapting the relevant techniques to binary and ordinal data. Chapters 9 and 10 may seem to come late, considering that they represent the author’s stated end goals for his readers: “Hierarchical Models” and “Going Bayesian”. However, by the time these are introduced, they flow quite naturally and convincingly from earlier chapters. As Garcia reassures readers, “Running hierarchical versions of our models in R is actually quite easy, i.e., the code is nearly identical to what we have already seen. The tricky part is not the code, but the statistical concept behind it” (p. 194) – which the author works to explain through comparisons and his usual walkthrough-style. He does the same when bringing Bayesian statistics into the fold, insisting on the shortcomings of the Frequentist statistics that constitute the norm, and the intuitiveness of the alternative. Finally, the essence of the author’s intent for this book is summarized in chapter 11, which contains brief “Final Remarks”.

The book packs quite a lot of information in its less than 300 pages. Garcia has a good sense of how much material is digestible at one time and leaves more in-depth explanations to others through several references to additional readings (e.g., Wickham and Golemund 2016, for R; Sonderegger et al. 2018, or Winter 2019, ch. 14–15, for mixed-effects models; Kruschke, 2015, or McElreath 2020, for Bayesian stats, etc.). Indeed, there exist other, more thorough (and voluminous) guides to R or statistics (or statistics using R), but this book’s strength lies in giving readers just enough to enable them to quickly apply their newly acquired knowledge and skills to their own data in order to produce complex, journal-worthy analyses. The book is timely, with increasing expectations for more refined accounts of the diverse populations and intricate results stemming from studies of second language acquisition and bi/plurilingualism, as well as other fields of linguistic research. That said, the author’s general disregard for standard reporting of *p*-values (his preference being towards confidence intervals and effect sizes) seems overly dismissive: though *p*-values have been misreported or misrepresented on occasion, they have not lost all usefulness.

Thorough and well-structured, the book encourages readers to adopt good organizational habits and includes extensive details on file management (summarized in Appendix D). Each chapter ends with a recap of its content and some extra exercises to try on one’s own. Each section builds on the previous following a logical progression, assuming readers have processed what came before. This cumulative construction could, however, limit the book’s relevance for those who might have wished to learn about specific subtopics; it can also lead to significant time spent leafing back and forth. For example, the link to access the book’s associated data files (required to follow along with the explanations and exercises, and available through the Open Science Framework) is given in the preface, with the mention that said files will not be of use until chapter 2; the link is never repeated, nor is there a later reminder where it can be found. Fortunately, references to previous material are generally clearly indicated.

While the data files, once located, are freely available for download, such is not the case for the R scripts and code blocks – nor can they be copy-pasted from the

book's electronic version. The author acknowledges this as a conscious decision: "You should manually type the code in R every time: it can be a little time-consuming, but it's certainly the best way to learn R – make sure you double-check your code, as it's very easy to miss a symbol or two." (p. xix). While arguably pedagogical, this choice may again reduce the book's audience to those who have the time and energy to fully commit to Garcia's approach.

The code transcriptions work as they should for those using the versions of R and R Studio indicated in the book, that is, the latest available at time of publication. Users with previously downloaded versions will probably want to update theirs; otherwise, some commands may not operate exactly as stated. Future users of the book are also likely to encounter such issues: R evolves rather quickly, being open-sourced and widely used. However, it should not lead to anything but minor adjustments.

If some of these observations seem petty, it is that such irritants may aggravate what could already be an exercise in frustration for some users: R remains a programming language, the syntax of which is rarely straightforward for non-programmers. If we truly wish to strive for across-the-board better statistical analyses in linguistic research, more user-friendly tools will be required. While Garcia lauds the merits of R over, say, IBM's Statistical Package for the Social Sciences (SPSS), he does not mention options such as JASP (JASP Team 2021): a free, open-source program with a point-and-click interface that allows for Bayesian calculations and hierarchical modeling. While staunch R advocates will argue that the inability to personalize code in such programs means sacrificing too much control, they may suffice for a great number of users.



Still, though requiring the patience to tackle what may be an intimidating learning curve, an investment towards mastering R will pay off for the career academic. Garcia's book can assist with this process: the author succeeds in making accessible to a broader audience the coding of scripts yielding nicely designed plots and well-conceived statistical analyses. If the hand-holding approach and informal style of writing may put off some more established researchers, the explicit instructions as well as the casual and supportive language (such as the frequent "don't worry") is meant to comfort those daunted by the book's subject matter.

While directed at second language researchers and graduate students, with examples and sample data pertaining to their field, the material covered could easily be adapted to other specializations who also handle quantitative data. *Data visualization and analysis in second language research* would serve well as a main course book for a data analysis class or a specialized seminar, or for a lab or research team's reading group. The lone reader will also benefit, though may need to consult with the large, "active online community" (p. 15) of R users – one of the selling points for using this program. Overall, those who are relatively new to hierarchical models and/or Bayesian statistics, and are willing to take on R as long as they are guided through it, should seriously consider Garcia's workshops-cum-handbook.

## REFERENCES

- JASP Team. 2021. JASP (Version 0.16). Available for download at <<https://jasp-stats.org/download/>>.
- Kruschke, John K. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. London: Academic Press, 2nd edition. <<https://doi.org/10.1016/B978-0-12-405888-0.09999-2>>
- McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: Chapman & Hall/CRC, 2nd edition. <<https://doi.org/10.1201/9780429029608>>
- Sonderegger, Morgan, Michael Wagner, and Francisco Torreira. 2018. *Quantitative methods for linguistic data*. <<http://people.linguistics.mcgill.ca/~morgan/book/index.html>>.
- Wickham, Hadley. 2014. Tidy Data. *Journal of Statistical Software* 59(10), 1–23. <<https://doi.org/10.18637/jss.v059.i10>>
- Wickham, Hadley, and Garrett Grolemund. 2016. *R for data science: Import, tidy, transform, visualize, and model data*. Sebastopol, CA: O'Reilly Media, Inc.
- Winter, Bodo. 2019. *Statistics for linguists: An introduction using R*. New York: Routledge, <<https://doi.org/10.4324/9781315165547>>

**Igor Mel'čuk**. 2021. *Ten studies in Dependency Syntax*. Berlin: De Gruyter Mouton. Pp. 444. US \$160.99 (hardcover).

Reviewed by Jianwei Yan , *Department of Linguistics, Zhejiang University, Hangzhou, China*  
and Tsy Yih , *School of Foreign Studies, Tongji University, Shanghai, China*

The volume under review is authored by Igor Mel'čuk, professor emeritus of linguistics at the University of Montreal. Serving as Volume 347 of “Trends in Linguistics Studies and Monographs”, it investigates the structural description of human languages from a typological perspective based on the Meaning-Text Theory (MTT) (Mel'čuk 1988).

The monograph consists of 11 chapters, among which the first is a brief introduction to the framework of MTT; the other ten are typological case studies based on dependency syntax. The volume is divided into four parts: “A Brief Overview of the Meaning-Text Model”, “Surface-Syntactic Relations”, “Some Hard Nuts in Syntax Cracked by Dependency Description” and “Word Order – Linearizing Dependency Structures”.

The first part, composed of Chapter 1, “Meaning-Text linguistic model”, introduces the Meaning-Text model (MTM), providing an overview of the theoretical framework for what follows. The author first proposes the postulates that underlie

---

Tsy Yih is the corresponding reviewer; ‘Tsy Yih’ is the transliteration of the name of the reviewer in his mother tongue, Shanghai Wu Chinese. He is also known as ZI YE in Mandarin pinyin. This work was supported by Zhejiang Provincial Philosophy and Social Sciences Planning Project (Grant No. 24NDQN15Z) and the MOE Project of the Key Research Institute of Humanities and Social Sciences at Universities in China (Grant No. 22JJD740018).