

Book Review

Stephen Carradini and Jason Swarts

Text at Scale: Corpus Analysis in Technical Communication

—Reviewed by

LUKE THOMINET  AND GISELL DEL CASTILLO 

Index Terms—*Qualitative research, quantitative research, research methodology.*

In *Text at Scale: Corpus Analysis in Technical Communication*, Stephen Carradini and Jason Swarts provide a well-structured introduction to an increasingly important research methodology. The book could function as a foundational resource for academic researchers new to corpus analysis methods, a reference volume for experienced researchers, and a textbook in graduate courses. It also points to ways that industry practitioners, such as content managers, technical editors, and user-experience designers, could use corpus analysis to improve efficiency, quantify user concerns, and assist with large-scale data editing.

Corpus analysis is a research methodology that identifies patterns of discourse in textual datasets that are too large for a human to analyze alone. It has been used extensively in fields such as computational linguistics. Technical communication studies have also employed corpus analysis over the past decade, although it is less common than other approaches in the field. Carradini and Swarts's book complements and expands our existing literature while uniquely filling a gap in technical communication's increasing attention to big data methods.

Chapter 1 discusses the growing importance of corpus analysis in technical communication. First, the authors argue that methods of scale are necessary for technical communication to have a "critical reflection on what the field has accomplished without oversimplifying the scope of

the field" (p. 4). The authors also say that corpus analysis can help the field respond to the increasing scale of technical documentation and discourse online. The remainder of the chapter describes potential applications for corpus analysis in research, pedagogy, program administration, and professional practice. These applications function as an excellent primer for readers to imagine how they might use the methodology.

Chapter 2 offers a more detailed introduction to corpus analysis. The authors begin by describing their methodological assumptions, including the lexical significance of individual words, the analytical value of quantification, and the relationship between dataset size and the generalizability of research results. Then, the authors briefly introduce a range of techniques to demonstrate what corpus analysis looks like in practice. Overall, this chapter could be used as a stand-alone text in classes that dedicate only a single week to the topic due to its clear, useful overview of the methodology.

Chapter 3 describes eight categories of questions that corpus analysis can answer, including questions about the "kinds of content [that] make up a corpus" (p. 49) or the "meaningful patterns of discursive action that arise around particular work practices" (p. 56). The authors also summarize in-field studies related to each type of research question, offering an invaluable resource for researchers seeking to build connections to past work. The chapter concludes by discussing the limitations of corpus analysis and the questions it cannot answer. This section may be valuable to researchers who need to frame their work for readers resistant to distant reading approaches.

Chapter 4 describes the process of building a corpus. First, the authors establish the qualities of a good corpus, including sufficient size, diversity of sources, and balance across desired phenomena.

Received 17 September 2025; accepted 17 September 2025.
Date of current version 26 November 2025. (*Corresponding author: Luke Thominet.*)

Luke Thominet is with Florida International University, Miami, FL 33199 USA (email: lthomine@fiu.edu).

Gissell Del Castillo is with the University of Pittsburgh, Pittsburgh, PA 15260 USA (email: gid63@pitt.edu).

IEEE 10.1109/TPC.2025.3612130

Book publisher: Fort Collins, CO, USA: The WAC Clearinghouse and the University Press of Colorado, 2023, 144 pp., including index.

They then address ethical considerations of corpus creation, including when and how researchers can gather online texts, while directing readers to additional ethical guidelines from organizations such as the Association of Internet Researchers. They also chronicle various methods for gathering texts from manual approaches to building tools and web scrapers. Finally, they discuss how data cleaning and textual annotation techniques can contribute to deeper insights.

Chapter 5 details how corpora can be analyzed. The authors include subsections on each of the following methods:

- Word and keyword analyses, which examine the frequency and importance of terms within a corpus
- Concordance analyses, which look at the context or location of terms
- Cluster analyses, which calculate the frequency of multiword phrases
- Visual collocation analyses, which graph the relationships between terms that frequently occur together
- Dispersion analyses, which identify the degree to which a term is spread throughout texts in a corpus

In each subsection, the authors define the method and give examples of applications in technical communication projects, including a fictional study of transcripts of collaborative meetings. The chapter includes useful screenshots of two corpus analysis tools, Lancsbox and AntConc, as practice demonstrations.

Chapter 6 presents an extended example of a corpus analysis of topic-based writing. First, the authors review relevant literature to identify potential research questions. Then they describe methods for collecting high-quality corpora of topic-based and book-based documentation for comparative analysis. Next, they discuss potential approaches for analysis. This pseudoresearch section explains the authors' research decisions and processes, giving it a more narrative structure than a typical article. The first approach analyzes referential language in topic-based documentation and demonstrates how frequency and dispersion analyses can help researchers identify differences in corpora. Before turning to their second approach, the authors list various corpus analysis attempts that failed, saying "exploration of the data produced a number of dead ends" (p. 115). This transparency is wonderful. It reinforces corpus analysis as an exploratory method while reassuring researchers whose initial attempts might not lead to usable results. The authors then describe a

second set of results around infinitive phrases in topic-based documentation. After an extended analysis of the pattern, they close the chapter by discussing how it could lead to further studies and pedagogical applications.

The concluding chapter considers the future of corpus analysis in technical communication. The authors argue for additional resources, including classes and textbooks, to prepare future researchers, as well as financial and computing assistance for those engaging with this methodology. They note the need for researchers to develop a background in linguistics to get the most value out of corpus analysis techniques and call for the creation of additional corpora of technical communication texts.

Although the book is an invaluable resource for technical communicators, it also has a few limitations. First, the book's short length might limit its ability to be a comprehensive resource for conducting a corpus analysis. As Carradini and Swarts note, researchers will have to supplement it with additional sources on statistical tests and linguistic patterns. Based on our experiences with the methodology, we would argue that researchers will also need additional support with storing and structuring corpora. Specifically, readers might benefit from further guidance on how to categorize original data, clean data, and analyze outputs. In addition, the book could offer further direction on how to maintain records of methodological and technical decisions to support distributed and collaborative corpus analysis projects. Finally, as teachers of methods courses, we wish that a digital companion accompanied the book to provide learners with working corpora and sample exercises, though this might be an opportunity for future publication.

Regardless of these quibbles, *Text at Scale* is brilliant. It can be used as a foundational text for students and researchers looking for a new method, while also acting as a strong reference source for researchers already engaging in this work. The book offers one of the best methodological introductions available in our field. Key concepts are described throughout, with many references to the field's literature. It also offers a transparent assessment of the assumptions, goals, and limitations of corpus analysis that could function as a model for the active interrogation of other research methodologies. We will be keeping the book on our shelves and recommending it for graduate-level courses. It is the kind of book that invites return visits, both for reference and inspiration.