

Daniela Landert, Daria Dayter, Thomas C. Messerli, and Miriam A. Locher
(eds.), *Corpus Pragmatics*. Cambridge: Cambridge University Press, 2023. 84 pages,
ISBN: 978-1-009-09508-2

Thanks to the ever-growing processing power of modern computers, the increasing accessibility of large amounts of data and the advancement of corpus tools, the use of corpus methods is becoming increasingly common in linguistic and related disciplines (Tognini Bonelli, 2010). Corpus pragmatics combines two complementary methodologies: pragmatics, which has traditionally relied on qualitative analysis, and corpus linguistics, which is essentially quantitative. *Corpus Pragmatics*, edited by Daniela Landert, Daria Dayter, Thomas C. Messerli, and Miriam A. Locher, is one of the open access books in the *Cambridge Elements in Pragmatics* series. It addresses key topics in corpus pragmatics, with an emphasis on the challenges faced by pragmatic research and the possible ways of overcoming these challenges.

The volume focuses on “the challenges and opportunities that different types of corpora offer for the study of pragmatic phenomena” (Abstract). It provides a practical overview of the many different ways in which corpora can be used for pragmatic research, covering both established methods and innovative approaches. In addition, it identifies areas in which new methodological developments are needed in order to make new types of data accessible for pragmatic research. Case studies are included to illustrate how empirical pragmatic studies are conducted. Thus, this book is highly recommended for those carrying out research in the new linguistic discipline that stands at the interface between corpus linguistics and pragmatics.

The volume comprises eight well-crafted chapters, each consisting of an introductory section, main discussion along with case studies and a short conclusion. Chapter 1 (Introduction) first stresses the role of corpora in pragmatic research. It then discusses the challenges faced by corpus pragmatics, focusing on the accessibility to contextual information, which presents a great challenge for corpus pragmatics. Next, the different ways in which the tension between qualitative and quantitative approaches can be resolved are discussed. The authors note that, overall, corpus pragmatics is form-to-function oriented, while function-to-form approaches face many challenges. This chapter also raises the issue of scalability, which is at the core of every challenge due to “the poor scalability of context-dependent interpretations and the manual work involved in identifying pragmatic functions” (p. 7). Despite all these issues, this chapter concludes that corpus pragmatics is a vibrant field of research that can yield meaningful insights. It ends with a brief summary of the advantages that corpus approaches offer and the outline of the Element.

Chapter 2 (Corpora and Their Characteristics: Challenges and Opportunities for Pragmatic Research) provides an overview of the different types of corpora and how their characteristics influence pragmatic research. Following the introductory section, it centres on the discussion of the nine types of corpora that are of interest to pragmatics, which include balanced reference corpora, topic- and domain-specific corpora, spoken corpora, multimodal corpora, pragmatically annotated corpora, learner corpora, parallel multilingual corpora, corpora including unsystematic large text collections and purpose-built self-compiled corpora. For each type, its advantages and limitations for pragmatic research are discussed. Examples of relevant previous studies are provided. This chapter concludes that since corpora come in all sizes, forms and structures, and with a wide variety of different content, the methods applied to the analysis and the skills needed by researchers differ vastly. “Any discussion of corpus pragmatic methods needs to first establish what kind of corpus data is involved in a given study” (p. 20).

Chapter 3 (Corpus-Driven Approaches with Self-Compiled Corpora) and Chapter 4 (Corpus-Assisted Approaches with Self-Compiled Corpora) elaborate on two different ways in which purpose-built corpora can be used. Chapter 3 focuses on “corpus-driven research on project-specific data” (p. 8). Based on a case study of persuasive discourse in the comments posted on the popular online forum Reddit /changemyview (Dayter and Messerli, 2022), this chapter explores the detailed process of corpus pragmatic research: a) how to design a corpus, b) ethics and copyright issues, c) the technical challenges of building a corpus and d) the analysis of formality-informality and overt expression of persuasion. The manifold benefits of corpus-driven approaches to pragmatics as well as the challenges of working with data compiled from online sources are illustrated, which include issues concerning ethics, copyright and data documentation.

Chapter 4 turns to “much smaller purpose-built corpora and the ways in which they can be used for corpus-assisted discourse analysis” (p. 9). In this chapter, two prominent approaches to pragmatics, corpus-based discourse analysis (CBDA) and corpus-assisted discourse analysis (CADS), are discussed. Drawing on a research project entitled *Language and Health Online: Typing Yourself Healthy* (Locher and Thurnherr, 2017), this chapter describes how a scaffolded and quantified analysis of online health practices is conducted step by step. The case studies demonstrate one of the possible ways to “work with a custom-made corpus to gauge pragmatic research questions” by “combining corpus linguistic methodologies and the possibilities of layered annotation” (p. 38). It concludes that the key feature of the corpus-assisted

approach to pragmatics is combining qualitative with quantitative steps of analysis.

Working with several corpora in combination can greatly enhance the available data and open up research perspectives that could not be pursued with any single corpus on its own. However, it also creates a range of challenges. Chapter 5 (Compatibility and Comparability: Combining Existing Corpora) discusses the practical challenges involved in this case under two headings: compatibility and comparability, i.e., differences in data formats, annotation, metainformation and access to data. In addition to discussing the challenges, the authors also suggest ways and corpus tools to deal with these challenges. A case study of a research project on epistemic stance is presented to illustrate the issues that arise when working with four Early Modern English corpora. As the authors argue, “while incompatibility of corpus formats can make it more difficult to work with different corpora side-by-side, the lack of comparability of data and results within or across corpora will negatively affect the reliability of results and interpretations” (p. 9).

Chapter 6 (Scalability: Meaningful Pragmatics with Large Data) concentrates on scalability and its consequences for quantitative and qualitative corpus methods. It first discusses sampling, a popular and well-established approach for dealing with large sets of retrieved hits. The benefits and drawbacks of random sampling and stratified sampling are compared. The authors argue that the choice of sampling unit and sampling process have a considerable effect on optimising the annotation efficiency, and they then include two case studies for demonstration purposes. One addresses the speech act of self-praise using the manual identification of high-density corpus files. The other identifies and automatically extracts high-density passages for the study of epistemic stance in a much larger corpus of Early Modern English. These recent approaches, which have the aim of retrieving texts or passages of texts from a corpus that makes use of the increasing size of corpus data, improve the scalability of qualitative methods and, thus, provide a rich insight into a given pragmatic phenomenon.

Chapter 7 (Multimodality: Integrating Non-verbal Information) centres on the possibilities and challenges of multimodal corpora for pragmatic studies. Given the complexities of such corpora, the discussion is devoted to three questions: How are modes and media represented in corpora? What examples in terms of corpora and studies are there at the time of writing? What challenges and opportunities lie ahead for multimodal corpus pragmatics? The first question is discussed along with the example of the *Swiss Memes Corpus* (SMC) to illustrate how a multimodal corpus is applied to address the pragmatic issues

in question. In discussing the second question, the concepts of multimodality and multimediality are distinguished and an overview of currently available multimodal corpora is provided. The third question is addressed with the example of a research project dedicated to the representation of the multimodality, termed humour support indicators (e.g., emojis), in textual comments to Korean TV drama. This chapter ends with an outline of the main areas of recent development in multimodal corpus pragmatic research.

The final chapter (Open Issues and Outlook) presents a summary of open issues and an outlook on possible future developments. It first restates the advantages of conducting corpus pragmatics, and then moves on to a brief summary of the challenges and opportunities of corpus pragmatics that have been discussed throughout this Element, including suggestions as to how to engage with corpus pragmatics. It reviews the ways in which corpora can be used for function-to-form approaches and the various ways of combining quantitative and qualitative methods. The opportunities and challenges of new corpus resources and the ethics of data collection are further emphasised. The authors suggest that consulting the references mentioned in this Element is probably a good starting point for a research project on corpus pragmatics. This chapter ends with the creativity and dynamics of the field and a brief outlook on the future.

The appendices are highly beneficial as they list all corpora (Appendix A) and all corpus tools (Appendix B) mentioned throughout the volume. Appendix A includes a large range of different corpora used for pragmatic research. Practical information on where to access them and/or find additional information about them has been provided. Appendix B presents an overview of the corpus linguistics tools and platforms that can be applied in corpus pragmatic studies.

To sum up, *Corpus Pragmatics* provides a comprehensive and practical guide to researchers and students who are interested in corpus pragmatics. This Element is written in accessible English, with certain depth and breadth, profound theoretical significance and broad application prospects. It provides readers with the resources for evaluating the opportunities and challenges faced by present-day corpus pragmatics. One of the prominent features of this Element is the detailed presentation of case studies that can help readers further understand the topic of each chapter. I note, however, that a clearer distinction between corpus-driven approaches and corpus-assisted/corpus-based approaches could be drawn. A more detailed explanation of the fundamental differences between the two approaches would be advisable for readers with little previous knowledge. In conclusion, this Element is a valuable contribution to the burgeoning field of corpus pragmatics.

Acknowledgement

This research is funded by Shanghai Municipal Foundation for Philosophy and Social Science, Project No. 2023BYY002.

Lin Chen | ORCID: 0000-0001-5135-3998

College of Foreign Languages, Shanghai Maritime University,
Shanghai, China

College of Foreign Languages and Literature, Fudan University,
Shanghai, China

chenlin@shmtu.edu.cn

References

- Dayter, Daria and Thomas C. Messerli. 2022. Persuasive language and features of formality on the r/ChangeMyView subreddit. *Internet Pragmatics* 5(1): 165–195. <https://doi.org/10.1075/ip.00072.day>.
- Locher, Miriam A. and Franziska Thurnherr. 2017. Typing yourself healthy: Introduction to the special issue on language and health online. *Linguistics Online* 87(8/17): 3–24. <http://dx.doi.org/10.13092/lo.87.4170>.
- Tognini Bonelli, Elena. 2010. Theoretical overview of the evolution of corpus linguistics. In: Anne O'Keeffe and Michael McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London and New York: Routledge, 14–27.