

## BOOK REVIEWS

McEnergy, T., & Brezina, V. (2022). *Fundamental Principles of Corpus Linguistics*. Cambridge University Press. xii + 314 pp.

Reviewed by Niall Curry (Coventry University)

*Fundamental Principles of Corpus Linguistics* makes a valuable contribution to the fields of linguistics and applied linguistics, by offering an in-depth, philosophical perspective on corpus linguistics as a field and methodology. As indicated in the *Preface* (p.xi), the book was written with a view to building on the authors' previous work, namely, McEnergy and Hardie (2011) and Brezina (2018), by focusing on the foundational concepts that underpin corpus linguistics as a methodology. To do so, the book draws on the work of Karl Popper (e.g. Popper, 1976, 2002) and interrogates the position of corpus linguistics as a (social) science, offering a line of reasoning that can be adopted by readers who wish to build and analyse corpora, based on notions of logic and common sense approaches. McEnergy and Brezina develop their ways of thinking and their approach to corpus linguistic analysis by reflecting on the epistemological perspectives embodied by researchers who make use of this now well-espoused, transdisciplinary methodology. As such, the book is designed as a complimentary resource, not replacing or reiterating the views shared in their earlier works, but, instead, offering a further perspective to equip readers with conceptual approaches to undertaking corpus building and corpus linguistic analyses e.g. falsification, common sense approaches, repeatability, replication, etc.

The iterative development of the 48 fundamental principles of corpus linguistics throughout the book's eight chapters is one of its stronger contributions. McEnergy and Brezina's approach to developing these principles takes readers on a journey from broad statements on corpus linguistics to more nuanced perspectives, meaning that even those less familiar with corpus linguistic analytical approaches can be scaffolded into more complex thinking on the topic. Moreover, as 15 of these principles are redeveloped within the text, the iterative approach to developing the principles is itself a meta-demonstration of the processes of falsification, reflection, and revision that the authors suggest as key to effective corpus building and analysis. To demonstrate the contributions of this book, in what follows, I offer a brief critique of its eight chapters. Subsequently, to exemplify the iterative development of McEnergy and Brezina's principles, I track the transformation of Principle 18 throughout the text, explaining the thinking that inspired its (re)development. Following that, I consider some

limitations of the book, signalling potential future directions for expanding the thinking in *Fundamental Principles of Corpus Linguistics*. This is then followed by a brief conclusion.

Following the introduction, *Chapter One – The First Sketch*, introduces the reader to a number of key concepts that they will need to understand in order to engage with the fundamental principles and epistemology presented by McEnery and Brezina. These include notions of linguistics and science, and interrogations of realism and common sense approaches. As the chapter continues, concepts of logic and empiricism are discussed where the *scientia realis*, a form of empiricism, is evoked as a key foundation in corpus linguistics research. *Chapter Two – What Is Science?* and *Chapter Three – How to Do Science?* return to key issues presented in Chapter One, elaborating on concepts such as falsification, probability, and corroboration. For example, McEnery and Brezina (p.42) make clear the importance of falsification in their view of corpus linguistics, stating that “[f]alsifiability is a key to building a filter through which our ideas may pass and be judged to be meaningfully testable or not”. These chapters also juxtapose notions of science and rhetoric, arguing that,

without due consideration of the scientific method linguistics would descend into a war of rhetorical flourishes, a war of words and opinions, devoid of a correspondence to reality in any systematic sense. (p.50)

Overall, by interrogating the potential alignment of corpus linguistics with approaches to science, these chapters pave the way for *Chapter Four – What Is Social Science and the Digital Humanities?* in which the discussion centres on positioning corpus linguistics as a social science. This chapter also introduces the notion of propensity to the fundamental principles of corpus linguistics (Principle 38), arguing that the probability of language occurring in a corpus relies on propensity, whereby unknowable forces outside of the corpus data can influence the language produced by those language users who make up a sample in a corpus. In *Chapter Five – Everyday Linguistics: Form and Function*, the authors reflect on the nature of language performance as imperfect as well varieties of language, such as learner language, to illustrate the importance of recognising the functionalist foundations of corpus linguistics. Specifically, the authors highlight the role of context, noting it to be “crucial for the interpretation of linguistic production” (p.178).

In *Chapter Six – Repetition and Replication: Laying the Groundwork for an Empirical Study* and *Chapter Seven – Replication: Carrying out an Empirical Study*, the authors unpack two key concepts that have become fundamental in contemporary corpus linguistics research: repetition and replication. Moving between the ideals of research and its pragmatic realities, they frame repetition

as the processes of redoing the same analysis with the same data and note that among the challenges of repetition are issues such as unavailable data. Distinguishing replication from repetition, the authors identify the former as the notion of reapplying the same approach to different data in order to attempt to falsify the hypothesis that emerged from a previous study. Challenges in replication appear manifold, in their view, and key among these challenges, for example, is the alignment of the new research design with that of the original study. Finally, *Chapter Eight – Conclusion* reflects on the contributions of the book, discusses the affordances of the 48 principles, and calls on the field to advance our collective understanding of epistemologies for corpus linguistics. Overall, as new ideas are discussed and critiqued in the book, they are used to posit new principles and develop, where relevant, existing principles, adding greater degrees of complexity, e.g. the case of Principle 18.

Principle 18 is introduced three times in the book, once in Chapter Two and twice in Chapter Three. Initially, Principle 18 states that,

[r]eporting falsifications is arguably more important than reporting corroboration. At the very least, both should be reported. (p.59)

Principle 18 argues that, as opposed to focusing on evidence corroborating expectations, it is arguably more important to show when researchers' expectations are falsified by the evidence they produce i.e. when hypotheses are not supported by the data. In its first redevelopment, Principle 18' adds more nuance to the original Principle 18, stating that,

[r]eporting falsifications is arguably more important than reporting corroboration. At the very least, both should be reported *unless the report is on a previously falsified hypothesis; then neither corroboration or falsification is of value.* (p.105; emphasis added)

The modification to Principle 18, emphasised in italics, is made amid a discussion of the value of intersubjective testing for falsification in corpus linguistics, whereby, following Popper, an "intersubjectively testable falsification is final [and] [c]orroboration of the system made past that falsification are of no value" (p.104). This means that if a hypothesis is falsified and the process of falsification is replicable and repeatable across individuals, no further falsification is needed. The final iteration, Principle 18", states that,

[r]eporting *conditional* [emphasis added] falsifications is arguably more important than reporting corroboration. At the very least, both should be reported *unless the report is on a previously rejected hypothesis – then a falsification of a falsification that led to rejection is of particular interest.* (p.109; emphasis added)

This final modification to the principle, again emphasised in italics, adds nuance to how one falsifies hypotheses in corpus research. The notion of conditional falsifications is introduced which, in the authors' view, is necessary given that all falsifications depend on whether the data used to falsify a hypothesis is of sufficient quality and relevance to do so. Overall, this whistle-stop tour through the positing, development, and redevelopment of Principle 18 serves to illustrate how this book offers readers a scaffolded pathway to developing nuanced and critical principles to inform corpus linguistics research.

While the book's contributions to shaping corpus linguistic analytical approaches are manifold, there remain some limitations to its scope and reach. Firstly, one might wonder to what degree all 48 principles proposed in this book can be applied to all corpora and types of corpus analysis. For example, Principle 42' states that,

[i]n situ falsifications regarding beliefs of appropriacy and well-formedness of a linguistic feature or structure based on Principle 40 may occur where appropriacy and well-formedness is rejected by a person in the intended audience of the linguistic production in question or the producer of the language. (p.164)

The notion underpinned by this principle is that there is evidence in the documented interaction within corpora to help researchers identify in situ falsifications. However, for corpora that contain student academic written texts for example, such as the British Academic Written English corpus (BAWE; Nesi & Gardner, 2018), this principle seems less applicable. This is because even though the texts within BAWE are indeed dialogic and addressed, their addressees (typically the lecturers marking the students written assignments) do not directly engage with or respond to the texts as part of the corpus. This means that the addressees cannot help us, as researchers using BAWE, to identify in situ falsifications. Arguably, there is no requirement to apply all 48 principles at all times or to all corpus studies. Rather, as, McEnery and Brezina note, the principles may be seen as "load-bearing pillars" (p.252) that can eclectically support a range of approaches to corpus linguistic research.

A second point of note is that while this book offers a valuable resource, investigating the notion of epistemology specifically for corpus linguistics research, the authors largely draw on English language research to support their investigation. Given that epistemologies are known to vary across cultures, contexts, and languages (Bennett & Muresan, 2016), the anglocentric focus may limit the applicability of some of McEnery and Brezina's 48 principles to corpus studies of languages other than English. For example, in Principle 6", McEnery and Brezina note that,

[c]orpus linguistics, drawing on *scientia realis*, works, as a social science, in a way which is informed by concepts from science – it is the study of observable language on which experience may be tested in accordance with Principles 7 and 11. (p. 133)

The reference to *scientia realis* is noteworthy as it relates directly to empiricism. Therefore, if one were to apply such thinking to a French language academic context for example, challenges may emerge owing to the key role of poststructuralist epistemologies therein (Bennett, 2015).

It should be noted that this critique of anglocentricity would be better applied to the field of corpus linguistics more generally, as most research therein focuses on and is written and published in the English language. Evidently, it is beyond the scope of any one book or piece of scholarly work to cover all potentially relevant epistemological, ontological, cultural, and contextual perspectives that shape research practices in corpus linguistics. In fact, recognising this, McEnery and Brezina call on readers in the final paragraph of Chapter Eight to elaborate on the position they offer, critique and falsify their epistemology, and consider ways to extend and advance thinking in this area. Therefore, seeing *Fundamental Principles of Corpus Linguistics* as a valuable starting point in bringing together these key issues in research methodologies and applying them to approaches in corpus building and analysis, there is scope for researchers to offer further perspectives on how and to what degree the epistemology delineated by McEnery and Brezina works across cultures, contexts, and languages.

Critiques aside, the contributions of this book outweigh any potential, and arguably, inevitable, limitations. Written at a critical point in the development of corpus linguistics as a field and methodology, this book offers an antidote for the potential impact of a number of worrying practices that are influencing the field. For example, as corpus linguistics has become more widespread, taken-for-granted views of representativeness in corpus research have become more evident, where the iterative interrogation of representativeness as a guiding concept appears to be increasingly assumed and unreported, as opposed to argued or evaluated in corpus research (Egbert, 2019). Likewise, the emerging, and often misguided, value attributed to big data in corpus linguistics brings with it prospective issues. For instance, there is potential for a lack of consideration for the role of interpretation when conducting corpus analyses (Egbert et al., 2020) and a growing preoccupation with quantity over quality in corpus description (Egbert, 2019). Moreover, with a focus on big data, there is a risk of relegating small, specialised corpora to a lesser status, despite small corpora remaining of critical importance when applying corpus linguistics to fields such as languages for specific purposes, contrastive linguistics, and pragmatics (Cotos, 2017; Curry, 2021; Rühlemann,

2018). Offering a counter-perspective to such worrying practices, *Fundamental Principles of Corpus Linguistics* reminds us of the need to focus not just on the quantity of corpus data, but also on their quality. In so doing, researchers can identify what a corpus represents, that for which the corpus data can account, and how to design and use corpora effectively to generate, refine, and/or answer research questions.

## References

- [doi](#) Bennett, K. (2015). Towards an epistemological monoculture: Mechanisms of epistemicide in European research publication. In R. P. Alastrué & Pérez-Llantada, C. (Eds.), *English as a Scientific and Research Language: Debates and Discourses* (pp. 9–36). Walter de Gruyter.
- Bennett, K., & Muresan, L. M. (2016). Rhetorical incompatibilities in academic writing. *Synergy*, 12(1), 95–119.
- [doi](#) Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
- [doi](#) Cotos, E. (2017). Language for specific purposes and corpus-based pedagogy. In S. Shannon & Chapelle, C. (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (pp. 248–264). John Wiley & Sons.
- [doi](#) Curry, N. (2021). *Academic Writing and Reader Engagement: Contrasting Questions in English, French, and Spanish Corpora*. Routledge.
- [doi](#) Egbert, J. (2019). Corpus design and representativeness. In T. Berber Sardinha & Veirano Pinto, M. (Eds.), *Multi-dimensional Analysis: Research Methods and Current Issues* (pp. 27–42). Bloomsbury.
- [doi](#) Egbert, J., Larsson, T., & Biber, D. (2020). *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User*. Cambridge University Press.
- [doi](#) McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- [doi](#) Nesi, H., & Gardner, S. (2018). The BAWE corpus and genre families classification of assessed student writing. *Assessing Writing*, 38, 51–55.
- Popper, K. (1976). The logic of the social sciences. In Adorno, T. (Ed.), *The Positivist Dispute in German Sociology* (pp. 87–104). Heinemann.
- [doi](#) Popper, K. (2002). *The Logic of Scientific Discovery*. Routledge.
- [doi](#) Rühlemann, C. (2018). *Corpus Linguistics for Pragmatics: A Guide for Research*. Routledge.

## Publication history

Date received: 16 December 2022

Date accepted: 20 December 2022

Published online: 2 March 2023