

The Cambridge Handbook of Corpus Linguistics. Edited by Douglas Biber and Randi Reppen. Cambridge, UK: Cambridge University Press, 2015, xviii+623. ISBN 978-1-107-03738-0

Reviewed by: Rachelle Vessey, Birkbeck, University of London, U.K.

The Cambridge Handbook of Corpus Linguistics is a collection of chapters authored by leading researchers in the field who draw on recent corpus linguistic research and case studies to report insights about the characteristics of the English language and its varieties. Written specifically for practicing scholars and advanced students in the field, the goal of the handbook is to survey the breadth of research questions relating to the study of the English language that can be approached using corpus linguistics. The aim is also to offer a critical discussion of the “state of the art” with respect to four areas that form the sections of the book: Part I (Methodological considerations); Part II (Corpus analysis of linguistic characteristics); Part III (Corpus analysis of varieties); and Part IV (Other applications of corpus analysis). While the handbook succeeds in meeting these goals, it might have benefited from greater consideration of the rapidly evolving nature of the English language and how corpus linguistics must adapt in order to remain a relevant research approach.

In the Introduction, Douglas Biber and Randi Reppen establish some essential groundwork for readers, notably proposing a contentious argument that corpus linguistics consists primarily of a “methodological perspective,” rather than being a theoretical subdiscipline of linguistics. Despite this, they explain that the goal of the handbook is not to include methodological discussions of corpus construction and analysis in each chapter, which means that methodological detail is generally absent from chapters. This leads to issues in some chapters, as noted below. Nevertheless, the book does tackle core

methodological considerations in the three chapters that comprise Part I (Methodological considerations).

In chapter 1, Mark Davies introduces the notion of corpora and explores the advantages and disadvantages of large, small and genre-based corpora. Speculating on the (near) future of corpora as datasets, he suggests that researchers will soon be able to access quantities of data and metadata online at a rate not yet imaginable by today's corpus linguists; this access will dramatically alter the nature of corpora and, as a result, the research that is possible to undertake using them. In chapter 2, Paul Rayson explores the tools and methods available for corpus construction and analysis and suggests that the tools driving the field might be limiting the kinds of research questions being asked. In chapter 3, Stefan Th. Gries notes that despite the acceptance that language consists of distributional/frequency-based probabilistic data, "most of our curricula do not contain even a single course on statistical methods" (71). Accordingly, he calls for greater understanding and consideration of statistics by corpus researchers.

The chapters in Part II present corpus analyses of linguistic characteristics, beginning with chapter 4 by Winnie Cheng. Perhaps due to its exclusion of methodology, Cheng's chapter has limited corpus linguistic detail and does not explain what goes into the compilation and analysis of prosodically annotated corpora. In chapter 5, Jonathan Culpeper and Jane Demmen provide an overview of the notion of "keywords," with a particular focus on the procedure used in the programme WordSmith Tools (Scott 2016). Due to this focus, there is unfortunately a lack of engagement with other corpus programs and their keyword functions. Also, the authors do not review alternative statistics used to derive "keyness," for example those proposed by Kilgarriff (2009) and Gabrielatos and Marchi (2012). In chapter 6, Xiao reviews the notion of collocation and the relevant statistics used for calculations, usefully including a critique of the statistics and a well-justified argument for the adoption of

log likelihood over other measures. Chapter 7 is a thorough review by Bethany Gray and Biber, who discuss the range of methods relevant to phraseology.

Chapters 8, 9, 10 and 11 all discuss grammar. Leech provides a historical survey of corpus contributions to descriptive grammar in chapter 8, which serves as a useful overview of the areas covered in later chapters, such as grammatical variation (Daniela Kolbe-Hanna and Benedikt Szmrecsanyi), grammatical change (Martin Hilpert and Christian Mair), and lexical grammar (Susan Hunston). The final three chapters of Part II cover discourse and pragmatics, starting with Alan Partington and Anna Marchi's review of using corpora in discourse analysis (chapter 12). Although it is an area ripe for discussion (see Brookes 2014), there is unfortunately only a very brief mention of critical discourse analysis. In chapter 13, Brian Clancy and Anne O'Keeffe discuss corpus pragmatics research, but the lack of methodological detail obscures some points and there is overlap between this and the previous chapter. Finally, in chapter 14, Irma Taavitsainen introduces pragmatics and historical pragmatics, including some of the overlap between sociolinguistics and pragmatics.

In Part III, the handbook turns to the corpus analysis of English varieties, starting with spoken discourse (Shelley Staples) and academic English (Ken Hyland), before turning to Susan Conrad's overview of register variation in chapter 17. Conrad begins with a distinction between the notions of register and genre and then turns to the range of different approaches available. Due to the specific focus on the multidimensional analysis technique, the accessibility of this chapter might have been improved by the inclusion of greater methodological detail. In chapter 18, Merja Kytö and Erik Smitterberg discuss the challenges of undertaking corpus linguistic analysis of diachronic registers, but insist that corpora have allowed for huge advances in register analysis. In her discussion of corpus stylistics in chapter 19, Michaela Mahlberg stresses that although even very basic corpus tools can enhance stylistics research, interdisciplinarity and qualitative and literary insights are

essential for explaining corpus findings. Next, Jack Grieve (chapter 20) explains that empirical research on naturally-occurring data and the patterns identified there through frequency have challenged some established notions of variationist research. For example, he notes that corpus research has identified frequency variables, which are measured as the frequency of linguistic forms relative to the length of a sample of data (i.e., the total number of words). This type of variable had been explicitly excluded from variationist research based on the assumption that it cannot exhibit social patterns, but corpus research has shown that this assumption is false (371).

In chapters 21, 22 and 23, the volume turns to international and non-standard varieties of English, such as World Englishes, English as a Lingua Franca, and Learner English. Both the chapter by Anna Mauranen, Ray Carey, and Elina Ranta (chapter 22) and that by Gaëtanelle Gilquin and Sylviane Granger (chapter 23) make note of the challenges posed by these varieties for standard corpus linguistic procedures (e.g., the tagging of non-standard data).

The final section of the book, Part IV, focuses on applications of corpus analysis. Ron Martinez and Norbert Schmitt (chapter 24) explore the relevance of units in vocabulary studies, suggesting a move from words to bundles and chunks. Magali Paquot (chapter 25) situates lexicography and phraseology in Firthian linguistics and describes the utility of the Sketch Engine corpus program (<https://www.sketchengine.co.uk>). Thomas Cobb and Alex Boulton (chapter 26) present findings from an empirical meta-analysis, showing corpus applications in the classroom. Fanny Meunier and Reppen (chapter 27) explore the value of using native and learner corpora in second and foreign language teaching, comparing corpus-informed and non-corpus-informed ELT grammar textbooks and finding that corpus-informed materials are more comprehensive and superior in their lexical information and

exercises. Finally, Silvia Bernardini (chapter 28) provides a history of using corpora in translation studies and sets out core trends in corpus-based translation research.

As this summary shows, the expertise of the authors, the breadth of coverage, and the currency of the work mean the handbook is an excellent and mainly up-to-date resource. However, one drawback is arguably the organization. Chapters are not necessarily in an intuitive order and the volume does not entirely cohere due to sections that are not always unified and chapters that sometimes do not logically flow from one to another. For instance, Shelley Staples (chapter 15) provides a useful overview of spoken discourse that could have preceded Clancy and O’Keeffe’s chapter on pragmatics (chapter 13), which focused predominantly on spoken language. Similarly, Conrad’s work on register variation could have been instructive to readers if it had been placed at the beginning of Part III, rather than two chapters later. In all cases, the lack of volume-internal referencing means that the reader is often left to deduce links between the chapters. If authors had referred to each other’s work, this might have established connections and contributed to the cohesion of the handbook. Instead, in some places there appears to be overlap between chapters: for example, Clancy and O’Keeffe’s chapter on pragmatics and Partington and Marchi’s chapter on discourse analysis clearly cover work that is related, if not the same.

Other shortcomings arguably emerge as a result of the aims of the book. Although no volume could conceivably tackle all areas relevant to corpus linguistics, the exclusion of methodological detail means that some chapters are rather difficult to follow and the book is not necessarily useful as a reference or for teaching and learning. Also, it is unfortunate that the focus on English prevents discussions of the evolution of corpus research on other languages (although Rayson, Gray and Biber, Grieve, Paquot, and Bernardini all discuss other languages to a limited extent). Furthermore, despite the focus on English, there is little discussion of the rapidly changing boundaries of the English language. Since an entire section

of the book (IV) is dedicated to varieties, it might have been useful to have some critical reflection on how the notion “*English corpus linguistics*” has, itself, evolved alongside the findings produced by research in the field. Although most chapters usefully highlight areas for future research, there is little discussion of how the English language is changing in dramatic ways in the online world. Indeed, although there are discussions of World Englishes (chapter 21) and English as a Lingua Franca (chapter 22), neither delves into the thorny issues of English as used online. Davies speculates on the future of corpora drawn from online language, but does not discuss the increasingly non-standard English used there. Language mixing and the use of multimodal features such as emoticons, emojis and gifs allow users to draw on quickly-evolving and expanding repertoires, and the resultant array of variation poses numerous challenges to accepted corpus linguistics techniques and procedures that have been developed to deal with normative language use. Although Rayson, Mauranen, Carey, Ranta, and Gilquin and Granger all discuss the challenges of non-standard data for established corpus linguistic procedures, it would have been in the reader’s interest for the authors to tackle online language specifically as part of the future of English corpus linguistics. Such a discussion would have ensured the completeness of this otherwise invaluable handbook.

References

- Brookes, Gavin. 2014. Review of Alan Partington, Alison Duguid & Charlotte Taylor. 2013. *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. *International Journal of Corpus Linguistics* 19(2). 292-300.
- Gabrielatos, Costas & Anna Marchi. 2012. Keyness: Appropriate metrics and practical issues. *CADS International Conference*.

<http://repository.edgehill.ac.uk/4196/1/Gabrielatos%26Marchi-Keyness-CADS2012.pdf> (25 May 2016).

Kilgarriff, Adam. 2009. Simple maths for keywords. *Proc. Corpus Linguistics, Liverpool*.

URL: <https://www.kilgarriff.co.uk/Publications/2009-K-CLLiverpool-SimpleMaths.doc>

Scott, Mike. 2016. WordSmith Tools. [Computer program]. Version 7, retrieved May 2016

from <http://www.lexically.net/>