

## Book review: Paul Baker (ed.), 2012, *Contemporary Corpus Linguistics*

MARIA-CRINA HERTEG

---

### Para citar esta reseña:

Herteg, M.-C. (2023). Book review: Paul Baker (ed.), 2012, *Contemporary Corpus Linguistics*, *ELUA*, 39, 207-215. <https://doi.org/10.14198/ELUA.21511>

---

### Autoría

#### MARIA-CRINA HERTEG

1 Decembrie 1918 University of din Alba Iulia, Romania  
crina.herteg@gmail.com  
<https://orcid.org/0000-0002-3929-0055>

#### Ficha bibliográfica:

*Contemporary Corpus Linguistics*,  
Paul Baker (Editor), Continuum, 2012, Print  
ISBN 9781441181336, 368 pp.

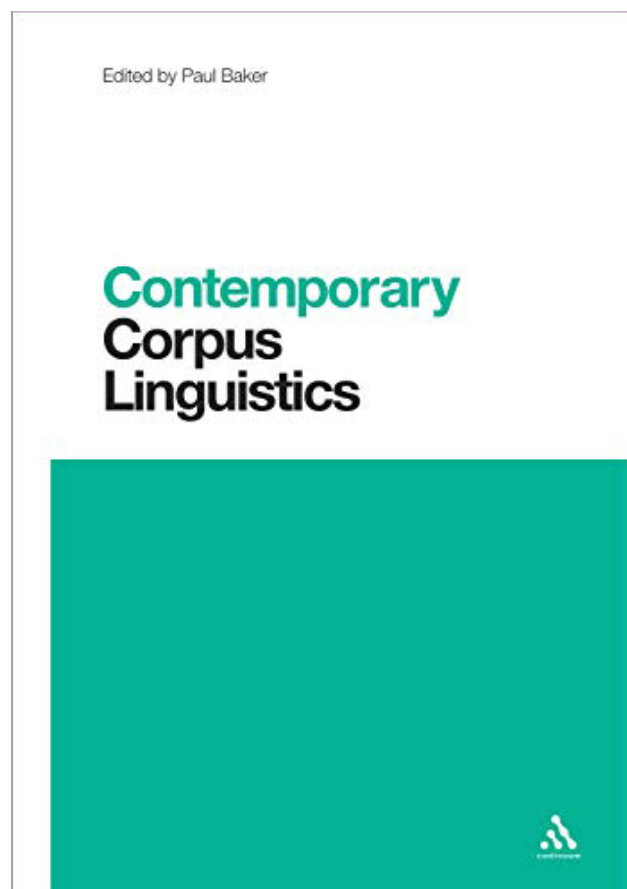
Recibido: 12/12/2021

Aceptado: 26/01/2022

© 2023 Maria-Crina Herteg



Licencia: Este trabajo está sujeto a una licencia de Reconocimiento 4.0 Internacional de Creative Commons (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



**Palabras clave:** corpus linguistics, annotation, corpus-building procedures.

This work is an instrumental resource in the field of Corpus linguistics, tackling different facets of corpus linguistics. The underlying idea is that almost every aspect pertaining to linguistics can be approached from a corpus-based perspective.

This book is well structured and balanced and it makes a smooth transition from one

chapter to another in terms of using corpora in different types of linguistic, literary and SLA analyses, and analyses of collocational patterns.

In the Introduction, the editor announces the aims of the book and analyses the articles which are included in the volume. The volume comprises contributions

tackling linguistic issues with the help of corpora such is the chapter authored by Alice Deignan. It also contains theoretical contributions in the field of corpus linguistics, annotation, corpus-building procedures and contributions pertaining to computational linguistics. Jonathan Culpeper shows how pragmatics can be analysed with the help of corpora, while Laurence Anthony focuses on the tools corpus linguistics works with, concordancers. Paul Baker identifies three types of contributions pertaining to the broad area of corpus linguistics: computational linguistics, corpus linguistics theory, linguistics using corpora.

In addition, it is a plea for using corpora in a plethora of analyses, touting their merits and warning against the perils of intuition in CMT (Deignan, 2012) and dictionary compilation (Hanks, 2012). Consequently, another idea analysed throughout the book is the difference between invented vs. authentic data. Three major areas of linguistics praise the merits of authentic data, CMT, as evident from Deignan's contribution, dictionary compilation, as evident from the chapter authored by Peter Hanks and last, but not least teaching, i.e., designing teaching material on the basis of corpora as evident from Randi Reppen's contribution. The book also pinpoints the idea that the advent of corpus linguistics and the use of corpora is the result of the increasing electronic data. These data can shed a new light on the way linguistic phenomena are perceived and can open new vistas of analysis. From the book we could observe that corpora facilitate all types of analysis. The comparison between the traditional approach and the corpus-based one to develop linguistic and literary studies is present in the book, for example, in Chapter 11, authored by Yukio Tono, which highlights the fresh perspective corpora bring in the analysis of SLA.

The advantages of using corpora as evident from the book are: they enable the statistical measurements and deepen the analysis, offering, for this reason, a fresh perspective. The book extols the virtues of corpora in undertaking linguistic, pragmatic (Jonathan Culpeper), stylistic analyses (Michaela Mahlberg) and contribution to teaching.

The methods the contributors resort to and introduce are: keyness method (Jonathan Culpeper), statistical analysis (Yukio Tono) and the types of corpora introduce in the book range from self-made corpora to ready-made ones (Oxford English corpus), also referred to as off-the-shelf corpora (Oakey, 2012). In addition, corpora are classified according to their use in monitor, parallel, specialised, and comparative corpora.

On the other hand, the book also addresses the status of corpus linguistics. In particular, Laurence Anthony discusses the standpoints adopted along the time by corpus linguists who see Corpus linguistics as a theory or as a branch of applied linguistics. Another idea tackled throughout the book is that corpora can be applied in the analysis conducted in different genres.

From the outset, Paul Baker notes the increasing number of contributions in the field of corpus linguistics mentioning the fact that the boundaries between them become more and more flexible. Paul Baker also notices that the number of linguists using corpora in their analysis is increasing. The author points out the idea that almost every field of linguistics can be approached from a corpus perspective. This idea pervades throughout the whole: cognitive linguistics, metaphor analysis, language learning, corpus stylistics, dictionary creation, translation.

The second chapter is authored by Alice Deignan and is entitled *Searching for metaphorical patterns in corpora*. This contribution approaches the theory of Conceptual metaphors from a corpus perspective. In the Introduction, Alice Deignan stresses the fact that the study of metaphors has been a topic of considerable interest across a range of academic disciplines. She then inventories academic disciplines which tackled metaphor analysis: corpus linguistics, applied linguistics, philosophy. Alice Deignan discusses how metaphor theory can contribute to corpus-based language description, the challenges that investigating metaphors pose for corpus linguistic techniques can contribute to metaphor theory. The author also explains that CMT works by establishing mental connections at the level of thought between

semantic area or domains known as Source Domains and Target Domains. The author also touches upon metonymy and the contribution of corpus linguistics to the study of metaphors. She underlines the fact that corpus linguistics contributes to and enhances the study of metaphors.

Deignan also makes reference to metaphor identification procedures and she indicates the method developed by the Pragglejaz group and the method propounded by Deignan (2005a). She then points out the difference between these two methods which basically consists in the role of the surrounding context in metaphor identification, namely the Pragglejaz method takes into account the contextual meaning of each word in the text, while Deignan (2005a) does not consider the surrounding context. The author examines other identification methods, the traditional approach to analysing word meaning in a corpus, the sampling based approach used by Charteris-Black (2004). Referring to the automatic identification of metaphors in a corpus the author discusses the difficulty of the undertaking and states that a small number of researchers are making progress in this area. Finally, Alice Deignan resumes the statement that cognitive linguistic theorists rely on intuition, while corpus linguistic theorists consider intuition a poor guide to language in use.

Chapter 3 is entitled *Corpora and Critical Discourse Analysis* and is authored by Gerlinde Mautner. It analyses how CDA can benefit from corpus linguistic techniques. The author adopts a critical stance and stresses the limitation of this approach. The author mentions the fact that both Corpus linguistics and CDA are relatively new approaches to linguistic enquiry and shows that they developed independently. However, she believes that despite their independent development these two branches can cooperate fruitfully as they share the social aspect of language. The author emphasises the fact that CDA's interests mainly pertain to in-depth analysis and the importance of extralinguistic context. At the other point of the continuum Corpus linguistics works with large sets of data and for this particular reason it cannot meet CDA's requirements. Gerlinde Mautner observes that

CDA researchers were reluctant to resort to the use of Corpus linguistic potential in their investigations at least at the inception of using corpora. She then enumerates and inventories corpus-based contributions to CDA. In the Concluding Caveats, the author touts the merits of corpus linguistic techniques in CDA as opposed to manual analysis. The author concludes by stating that corpus-driven approaches give researchers the perspective to make the results more reliable.

Chapter 4, *Corpus Stylistics and the Pickwickian Watering Pot*, is a fine example of how corpus linguistic techniques can contribute to stylistic analysis. By resorting to corpus linguistic techniques, Michaela Mahlberg extends the object of analysis which otherwise would have remained restricted to only one text. In other words, the author analyses the words watering pot in Pickwick Papers in terms of frequency and collocations. Michaela Mahlberg offers an interesting analysis of literary pieces by resorting to corpora. The extended analysis consists in the fact that with the help of concordances the author analyses the word watering pot in a corpus consisting of 23 novels written by Dickens. The concordancing software enables the author to analyse the distinction of watering pot across the text in the Dickens corpus in terms of occurrences, collocations. The author complements her analysis by investigating clusters, repeated sequences of words. The author believes that the clusters in each groups share formal and functional similarities. Such clusters can be used as pointers to local textual functions. They can be interpreted as building blocks of textual worlds in the sense that the clusters point to aspects of the world of Dickens that is built through the labelling of characters and themes, characters' speeches, references to time and place. In the Conclusion, Michaela Mahlberg mentions the main aim of the article and reinforces the basic idea presented throughout the article, that the potential of corpus stylistics lies in the combination of different approaches.

Chapter 5, *The Metalanguage of Impoliteness: Using Sketch Engine to Explore the Oxford English Corpus*, pertains to the area of pragmatics analysed with the help of



corpora. Jonathan Culpeper opines that large corpora are needed in order to conduct studies on impoliteness devices and impoliteness metalinguistic expressions. The author conducts the study on OEC Oxford English Corpus and works with Sketch Engine. From the beginning Culpeper underlines that the chapter contributes and expands on corpus analysis of metalanguage. The author states the idea that scholars have not done much to investigate lay people's metalanguage for politeness and even less for impoliteness. Also, he generates a comprehensive range of terms using *thesauri*. The next step the author takes is to use the corpus-based approach on a subset of the term list, this subset consists of the words rude(ness) and impolite(ness) and he examines their frequencies and co-textual characteristics such as collocations and genre distribution. Based on previous research conducted by Culpeper *et al.* (2003), the author provides a list of labels of impoliteness. The author does not pay heed to the frequency of the labels being more interested to discover the range of labels that were used. To widen the scope of his research, Jonathan Culpeper used Social Sciences Citation Index combined with the Arts and Humanities Citation Index.

Chapter 6, *Issues in the design and development of software tools for corpus studies: The case for collaboration*, is written by Laurence Anthony and describes AntConc, a freeware, standalone, multiplatform corpus toolkit developed by the author in collaboration with some of the leading corpus linguists in the world.

The author is particularly interested in the technical aspects of corpus linguistics and describes the AntConc tool from its inception to the present day. In the Introduction, the author refers to the role corpora played in the study of language over the past 40 years. The emergence of Corpus linguistics in 1960s entailed a revolution in the study of language and it was the result of two factors: corpora and the way in which corpora are stored and analysed. Laurence Anthony underlines the fact that nowadays most of the corpora are electronic and shows that computer science and engineering contributed a lot to the way in which corpora are stored and analysed.

He advances the idea that corpus linguists are more likely to contribute to the creation and development of software than engineers, arguing that corpus linguists have a better understanding of what they want to analyse, why they want to analyse it and who will be using the software (Anthony, 2012: p. 88-89).

In the first part of the chapter, the author refers to the place of Corpus linguistics as discussed by corpus linguists and underlines that a number of researchers consider it a methodology and he indicates here Kennedy (1998), Meyer (2002) and Scott & Tribble (2006), while Tognini-Bonelli (2001) considers Corpus linguistics as a branch of applied linguistics (Anthony, 2012: p. 90). The author foregrounds the idea that a poorly constructed corpus will lead to poor results (Anthony, 2012: p. 90).

On discussing about the functionalities of corpus tools, the author gives examples of previous research conducted in the field and cites Biber *et al.* (1998) who state that the programs can be written without a special aptitude in computer science or mathematics (Biber *et al.* 1998 as cited in Anthony 2012: p. 93).

The author presents the facilities, the functionalities and the aim of the concordancer: N-gram, keyword generations and the fact that it can work with almost all languages in the world (Anthony, 2012: p. 95). The software has been designed for classroom environment and has continuously been developed based on teachers' feedback. The features include: sorting KWIC concordance line by chapter position, the ability to adjust colours, font sizes and font styles so that learners would always be able to easily see the results of searches on a central projector, the ability to handle annotated data and produce lemma lists. The author concludes by saying that AntConc can be valuable for learners, but also includes a robust set of features for serious corpuswork. He also stresses the fact that AntConc is in a continuous process of development based on the increasing requests and suggestions the author of the software receives from users.

In Chapter 7, *Compatibility between Corpus Annotation Efforts and its Effects on Computational Linguistics*, Adam Meyer focuses

on the annotation of corpora differentiating between manual annotation and automatic statistics-based analysis. The author discusses the aspects which have effects on annotation: the way the information is recorded and which corpora are used. The author distinguishes between two annotation approaches: single-theory approach and *a la carte* approach focusing on the features of these approaches as well as on the research groups who use them. The author enumerates all the features the ideal corpus should have and highlights the idea that unfortunately there is no such corpus.

In Chapter 8, *Spoken Corpus Analysis: Multimodal Approaches to Language Description*, Irina Dahlmann and Svenja Adolphs aim to analyse another facet of corpus linguistics: multimodal corpora. The authors point out the idea that spoken and written corpora are concerned to textual representations while multimodal corpora “allow us to reflect on and evaluate some of the methods for analysing textual renderings of spoken discourse established so far” (Dahlman and Adolphs, 2012: p. 125). From the outset the authors emphasise the fact that Corpus linguistics have influenced the way in which we study and describe language by expanding the range of data and by enabling researchers to analyse both written and spoken texts. In addition to written and spoken corpora, the multimodal corpora enable researchers to analyse aspects such as gesture, prosody and kinesics.

The authors mention the difference between other current research on spoken corpora and their undertaking by referring to the fact that compared to other research that focuses on orthographic transcriptions they aim to illustrate the gains achieved by pause annotation for the description of MWE candidate *I think* as captured in spoken English. The corpus they use in their analysis is ENSIC, English Native Speaker Interview Corpus which contains transcriptions and recordings of 35 interviews.

In Chapter 9, *Fixed Collocational Patterns in Isolexical and Isotextual Versions of a Corpus*, David Oakey shares the belief that in order to conduct comparative studies of the discourse functions of fixed collocational

patterns a comparable amount of corpus data is necessary. The author introduces two approaches for conducting such research: 1. to subdivide existing normative corpora to allow comparisons to be made between different registers or between genres. The author draws attention to the fact that the unequal data analysed might lead to the fact that the results of the analysis might be biased in favour of the subcorpus with the greatest number of texts or tokens. The author also identifies a second approach to conducting an analysis of discourse functions of fixed collocational patterns, namely to balance the number of tokens in each subcorpus by using text samples. The aim of the chapter is to report on a comparative study of fixed collocational patterns. The author conducts his analysis on a corpus comparing research articles from eight academic disciplines. In the Introduction, the author makes reference to previous work on collocation in English and highlights the importance of corpora in conducting such research. There are two methodological approaches to corpus-informed comparisons of the role of such collocations. The author showcases two aspects that may influence the result of the analysis of collocations using data from existing “off-the-shelf” corpora. Thus, according to the author, the amount of data from each corpus is unequal and the studies may be inaccurate. Another aspect pointed by the author is that corpus compilers achieve a balanced number of tokens by using text samples with the result that discourse features from the parts of the texts are not included in the sample are not represented in the corpus. The author outlines the fact that the development of corpora has facilitated the analysis of collocations. Furthermore, the analysis of collocations has been refined focusing on their role in discourse rather than in language as a whole. Oakey (2002) conducted a corpus-based work on a selection of lexical phrases in a subset of the BNC. The findings of the research were that extended collocations with the same form as Nattinger and DeCarrico’s lexical phrases were indeed used in academic writing, but these extended collocations performed more than one identifiable discourse function. David Oakey provides some tips to be taken into account



by corpus designers in order to avoid bias, mention is made of the size criteria in the case of comparative corpora.

Chapter 10, *Corpus Linguistics and Language Variation* is written by Michael P. Oakes. The author points out the similarities between the Brown corpus and the LOB corpus, referring to the number of words they comprise, the contents the fact that they were assembled using identical sampling methods.

Chapter 11, *Integrating Learner Corpus Analysis into a Probabilistic Model of Second Language Acquisition*, authored by Yukio Tono deals with research taken in the field of learner corpora. The author makes reference to the JEFLL corpus (Japanese EFL learner corpus) and provides details with reference to its incorporation and to its structure. The author discusses the issue of multifactorial analysis in relation to SLA and Tono (2002) is indicated as one of the main authors to have dealt with the issue and to have emphasised the value of multi-factorial corpus analysis as a methodological innovation. The author enumerates the factors that influence the SLA pointing out the following ones: environmental variables, linguistic or cross-linguistic factors. The author claims that corpus linguistics can provide a very strong empirical basis for SLA. According to Yukio Tono, the data driven nature of corpus linguistics is one of its strengths, namely findings are supported by a large amount of attested language use data. In the final part of the chapter the author argues that learner corpus research should be able to make a significant contribution to a probabilistic view of SLA theory based on Bayesian reasoning.

Chapter 12, *English Language Teaching and Corpus Linguistics: Lessons from the American National Corpus* is authored by Randi Reppen. The chapter in its entirety praises the merits of corpora in the classroom as an efficient tool of teaching and showcases the benefits of corpora and corpus information in classroom teaching. The author exemplifies with activities designed on the basis of ANC and analyses the dissimilarities and similarities between ANC and BNC. The major similarity is that ANC was built following the same structure as BNC. Among the dissimilarities Randi Reppen

mentions the age of the electronic texts included in the two corpora pointing out that ANC contains newer information. Referring to the use of corpora in the learning process Randi Reppen mentions the concept DLL, Driven Language Learning introduced by Tim Johns and which prompts learners to act as “language detectives” and discover patterns of language use that are revealed by concordance lines. The author provides examples of activities that rely on the use of corpora or corpus analysis used as a source of teaching materials. The author also addresses a series of questions which cover the following aspects: the existence of a particular grammar structure that students need to practise. The author presents a set of activities aimed to outline the way that corpora can be used in order to bring register awareness into the classroom. Among the advantages of using corpora in the classroom, mention should be made of the fact that they raise students’ awareness of different registers, variation of English. In addition to helping students differentiate between different registers, corpus-based activities facilitate their comprehension of the linguistic features pertaining to different registers. The author provides insights and designs tasks based on corpora which are meant to be used as examples by teachers embarking on teaching relying on corpora.

Chapter 13, *The Impact of Corpora on Dictionaries*, by Patrick Hanks tackles the use of corpora in designing dictionaries and the benefits corpora bring to dictionary compilation. Patrick Hanks warns against the perils of using invented data in the compilation of dictionaries. The author stresses the idea that nowadays with such a variation of corpora and benefitting from large corpora, it is very unwise to resort to intuition rather than benefit from a large body of evidence. The author provides insights on lexicographical work on the basis of his experience as a lexicographer.

Chapter 14, *Using Corpora in Translation Studies: The State of the Art* is written by Richard Xiao and Ming Yue and deals with the issue of corpora in translation studies. The chapter intermingles theoretical aspects pertaining to corpora, such as types of corpora as well as practical ones. The

theoretical aspects pertaining to translation studies comprise translation strategies and translation universals. The chapter starts with an incursion in translation studies, the different branches of translation studies. In the first part of the contribution the authors intersperse theoretical aspects of translation studies with the major trends in translation studies, i.e., the descriptive and prescriptive trends. The authors categorise the main types of corpora used in translation studies and clarify the confusion surrounding the classification parallel vs. comparable corpora and comparable vs. comparative corpora, illustrating these classifications and definitions with examples. The authors identify the factors contributing to the convergence between corpus research and translation studies, showing that corpora applied in translation studies favour contrastive analysis. Among other advantages that corpora bring to translation studies mention should be made of the fact that they contribute to the creation of terminological databases and banks, pinpointing the fact that terminology management tools are the product of corpus research. In addition, they can also be used as a basis for translation evaluation and review. The authors also make an inventory of the types of corpora used in translation studies, as well as of research and researchers that resorted to corpora in translation studies. The authors particularise the analysis of the corpus data with evidence from the English-Chinese translation of fiction.

Chapter 15 is written by Andrew Hardie and is titled *Corpus Linguistics and the Languages of South Asia: Some Current Research Directions*. This contribution addresses the application of corpus linguistics methodologies to the languages of South Asia and as well as challenges and difficulties in corpus analysis in South Asian countries. Andrew Hardie mentions the stumbling blocks in compiling corpora in South Asian countries: the writing system of the South Asian countries and the fact that those countries use multiple languages. The author undertakes a short corpus-based investigation in order to show the kinds of avenues of analysis that are open to the South Asian researcher using corpus data. Andrew Hardie makes reference to aspects related

to corpus construction and analysis, such as corpus annotation, more specifically POS tagging and parsing.

The contribution authored by Robert Lew, *The Web as Corpus vs Traditional Corpora: Their Relative Utility for Linguists and Language Learners*, (Chapter 16) compares the world wide web as a corpus to traditional corpora, looking at the advantages and disadvantages they bring forth. The author observes the expansion of the world wide web as a corpus witness and considers it is beginning to become a real competitor to traditional corpora. The chapter also examines the advantages and disadvantages of the world wide web as a corpus focusing on the target users, whom the author divides into experienced users, linguists and less experienced ones, language learners. The author draws attention to the fact that corpora are becoming more and more accessible to language-related professions: translators, teachers. Technical aspects pertaining to corpus compilation are analysed such as linguistic representativeness and the balancing of corpora.

In Chapter 17, *Building and Analysing Corpora of Computer-Mediated Communication*, Brian King outlines the incompatibilities that emerge between CMC data and corpus study. The author makes an inventory of scholars who have provided general and specialised advice on the topic of corpus construction. Among them, mention should be made of Biber *et al.* 1998, Kennedy 1998, McEnery and Wilson 2001, Baker 2006. The author resumes an idea often dealt with throughout the volume, that corpus analysis minimises intuition (King 2012, p.303). The author opines that the sociolinguistic analysis he was carrying out could have been done without resorting to corpora, but he chose to resort to corpus analysis for the specific reason that corpus analysis eliminates intuition. Moreover, he opts for a combination of qualitative and quantitative analyses. The use of intuition and corpus analysis is shrouded in metaphor “the lavender-coloured glasses” metaphor which encompasses the following entailments: the eyes = the intuition, the spectacles/ the lenses are those which hide the unexpected



linguistic patterns, the corpus-driven analysis favours the unbiased overview. The author shares some of the problems he encountered while he was compiling the corpus. The corpus comprises data collected from the “public” spaces of a number of queer chat-rooms. The author also explains his choice of tagger which is not quite appropriate for the content of the corpus he used as a basis for his analysis. As taggers require data that contain grammatically and orthographically “correct” constructions, while chat-room data is often “messy”. The author opted for carrying out an examination of an untagged word list and grouped words that had similar semantic functions together by hand. Brian King also raises the issue of ethical aspects in comprising a corpus, such as gaining permission to use sources. He also explains the solutions which made the compilation of his corpus both ethically and practically possible. The author explains which are the main challenges for corpus analysis highlighting the fact that turn-taking in chat-rooms makes corpus analysis difficult. He adds cyberortography and addressivity and explains what the latter refers to. The author resumes the aims of the contribution: to examine who was using various word categories, how frequently and in which situations. The author introduces the system he used for conducting his analysis.

In a nutshell, the book provides a plethora of perspectives on corpus linguistics and it pleads for the use of corpora in a large array of research, ranging from linguistic to pedagogic research.

## REFERENCES

- Anthony, L. (2012). Issues in the design and development of software tools for corpus studies: The case for Collaboration. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 87-104). Continuum.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum. <https://doi.org/10.5040/9781350933996>
- Baker, P. (2012). *Contemporary Corpus Linguistics*. Continuum.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>
- Charteris-Black, J. (2004). *Corpus Approaches to Critical Metaphor Analysis*. Palgrave-MacMillan. <https://doi.org/10.1057/9780230000612>
- Culpeper, J., Bousfield, D., & Wichmann, A. (2003). Impoliteness revisited: with special reference to dynamic and prosodic aspects. *Journal of Pragmatics*, 35(10-11), 1545-1579. [https://doi.org/10.1016/S0378-2166\(02\)00118-2](https://doi.org/10.1016/S0378-2166(02)00118-2)
- Culpeper, J. (2012). The Metalanguage of Impoliteness: Using Sketch Engine to Explore the Oxford English Corpus. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 64-86). Continuum.
- Dahlmann, I., & Adolphs, S. (2012). Spoken Corpus Analysis: Multimodal Approaches to Language Description. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 125-139). Continuum. <https://doi.org/10.4324/9780203526149>
- Deignan, A. (2005a). *Metaphor and Corpus Linguistics*. John Benjamins. <https://doi.org/10.1075/celcr.6>
- Hanks, P. (2012). The Impact of Corpora on Dictionaries. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 214-236). Continuum.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Longman.
- King, B. (2012). Building and analysing corpora of computer-mediated communication. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 301-320). Continuum.
- McEnery, T., & Wilson, A. (2001). *Corpus Linguistics* (2nd ed.). Edinburgh University Press.
- Meyer, C.F. (2002). *English Corpus Linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511606311>
- Oakes, M.P. (2012). Corpus Linguistics and Language Variation. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 159-183). Continuum.



- Oakey, D. (2012). Fixed Collocational Patterns in Isolexical and Isotextual Versions of a Corpus. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 140-158). Continuum.
- Scott, M., & Tribble, C. (2006). *Textual Patterns. Key Words and Corpus Analysis in Language Education*. John Benjamins. <https://doi.org/10.1075/sc1.22>
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins. <https://doi.org/10.1075/sc1.6>
- Tono, Y. (2002). *The Role of Learner Corpora in Second Language Acquisition and Foreign Language Learning: The Multiple Comparison Approach*. Unpublished PhD thesis. Lancaster University.
- Tono, Y. (2012). Integrating Learner Corpus Analysis into a Probabilistic model of Second Language Acquisition. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 184-203). Continuum.

