Book Review

The Routledge Handbook of Corpus Linguistics
Anne O’Keeffe & Michael McCarthy (Eds.)
Routledge, New York, 2012

Giuliano Vivanel†
giuliano.vivanet@unica.it

The Routledge Handbook of Corpus Linguistics is edited by Anne O’Keeffe (University of Limerick, Ireland) and Michael McCarthy (University of Nottingham, UK and Pennsylvania State University, USA). The main authors aim at providing a comprehensive and timely overview of studies on Corpus Linguistics and Applied Linguistics (cf. previous collected works by Facchinetti, 2007 and McCarthy & Sampson, 2004). In recent years, especially thanks to the development of technologies for the analysis of large bodies of texts, this field has proved to be very dynamic and with a broad range of application fields, such as computational linguistics, discourse analysis, forensic linguistics, pragmatics, language pedagogy and translation studies (Biber, Conrad, & Reppen, 1998). The Routledge Handbook of Corpus Linguistics includes forty-five contributions (each of them enriched by suggested readings and an updated bibliographies) by experts coming from the key-areas of the discipline. The chapters are organized in eight sections. In the first section, the editors provide an introduction to corpora from an historical perspective and a theoretical overview of the evolution of corpus linguistics.

In the second section, Building and designing a corpus: what are the key considerations?, Reppen introduces the basic questions underlying the construction of corpora, such as problems related to the corpus size, how to collect texts, and how much mark-up is needed. Adolphs and Knight discuss the basics of developing a spoken corpus and the strategies for analysing and transcribing spoken data. In the following chapter, Nelson introduces the problem of building a written corpus and analyses the required steps: planning,
sampling, balancing and making the corpus representative, gathering, computerising and organising texts. Then, Koester provides an introduction to the development of small specialised corpora and analyses several questions involved in this process. Clancy considers the problem of building corpus with the aim of representing a variety of a language, starting from an analysis of variety of language meaning and continuing with issues like size, diversity, representativeness and balance. Interestingly, Thompson focuses his attention on the building of specialised audio-visual corpus and on collecting data, the preparation of transcriptions and annotations.

The third section, *Analysing a corpus: what are the basics?*, begins with a contribution by Lee who offers a general survey of currently available corpora (the chapter is divided into several paragraphs related to major English language corpora, lingua franca corpora, and non-English corpora and multilingual corpora). Evison introduces problems concerning the analysis of a corpus, such as exploring word frequency, key-word lists, concordance lines. In the following chapter, Scott introduces the use of software, showing what computers do really well, what they cannot do at all, what they can do with difficulty, and finally things that computers are expected or believed to be able to do in corpus linguistics. Hunston is interested in the strategies to find patterns in concordance lines and issues related to the assessment of their frequency. The nature and the use of concordances are discussed by Tribble in his contribution, by distinguishing two main phases (before the computer age and after the introduction of computer applications in this field). Lu is instead interested in language development and the use of corpus software in this application scenario.

The fourth Section, *Using a corpus for language research: what can a corpus tell us about language?*, begins with a contribution by Moon on the intersection between corpus and lexis. The author faces problems in lexical linguistics, such as the number of words comprised in the main vocabulary of a language or what we can learn about a word from looking at its co-occurrences, and the relationships between the meaning of words and their context. Greaves and Warren focus their attention on corpora and the study of multi-word units, based on the idea that we best know the meaning of a word, not by considering it as an isolated entity, but as a more holistic unit. Conrad considers what a corpus can tell us about grammar, by means of patterns in contexts and investigating multiple features and/or conditions simultaneously. Biber covers the theme of the use of corpora for analysing registers and genres,
distinguishing between these different perspectives. In the following chapter, Handford, after a brief introduction on genres in corpora, proposes a discussion of the methodological advantages of specialised corpora in analysing genres. Moreover, he focuses his attention on the use of corpora to analyse academic, professional, and non-institutional genres. The study proposed by Thornbury introduces the use of corpora to analyse discourse, also taking into account the limits of using a corpus in such a subject. Interestingly, Rühlemann investigates the relationship between pragmatics and corpus linguistics. In particular, he focuses his attention on the following areas of pragmatic phenomena: conversational encounters, discourse marking and speech act expressions. Anh Vo and Carter develop this perspective by studying creativity through the use of corpora and considering both spoken and written aspects of creativity.

The fifth section, *Using a corpus for language pedagogy and methodology*, starts with a work by Cheng on what a corpus can tell us about language teaching. The chapter argues that corpora, corpus-analytic tools and corpus evidence have been increasingly used in English language teaching and learning for the last two decades (cf. Sinclair, 2004). The contribution by Walsh analyses in depth the topic of the creation of language teaching materials. The following chapter, written by Chambers, is devoted to data-driven learning. This approach does not limit the use of corpora to the development of language-learning materials, but also gives learners access to substantial amounts of corpus data that can be found, for instance, in a dictionary or in a course book. The pedagogical functions of data-driven learning are discussed by Gilquin and Granger, who firstly introduce the operationalization of data-driven learning and then face the problem of assessing its effectiveness. At the end of the section, the contribution by Sripicharn is devoted to the preparation of learners for using language corpora, covering issues like assessment, learning objectives and available tools.

The sixth Section, *Designing corpus-based materials for the language classroom*, starts with the chapter written by Jones and Durrant on what a corpus can tell us about vocabulary teaching materials. Moreover, the authors face the problem of identifying the types of corpora which are considered as more suitable for academic vocabulary learning, as well as of selecting the right approach to design teaching materials (cf. Knowles, Mcenery, Fligelstone, & Wichman, 1997). Hughes offers an analysis of the role of corpora in relation to grammar teaching materials, considering their benefits, limits and looking to
their future development. McCarten takes in account the use of corpora in order to write a course book and discusses the future of corpus-informed course books. The following chapter, written by Walter, is strictly related to the latter and concerns the application of corpora in writing dictionaries. He discusses the use of corpora by lexicographers (also considering current limits and future developments) and analysing tools that can be useful for lexicographers in this context. Flowerdew deals with the corpus-based and corpus-driven approaches to design writing materials, in relation to English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP) instruction. Coxhead investigates the usage of corpora in relationship with the English for Academic Purposes (EAP), addressing several major questions related to this topic (what can corpora reveal about aspects of academic language in use; how can corpora influence EAP pedagogy; how can corpora be used in EAP materials; what can a corpus tell us about EAP learner language; and what might the future be for corpora in EAP). Vaughan, in his chapter, analyses the use of corpora by teachers for their own research. The author’s work is aimed at practitioners who are interested in conducting their own professional or pedagogical research by exploiting corpora.

The seventh section, Using corpora to study literature and translation, starts with a chapter written by Kenning on parallel and comparable corpora (defined as collections of electronic texts that are closely related to each other) and the strategies for their usage. Then, Kübler and Aston examine applications of different types of corpora in the practice and pedagogy of translation (in particular, they are interested in pragmatic translation). In the following chapter, McIntyre and Walker outline different techniques for using corpora to study poetry and drama with the aim of demonstrating their value in stylistic analysis. This section is completed by Amador-Moreno who analyses the use of corpora to explore literary speech representation (introducing similarities and differences between real/fictional speech and presenting a case study of an Irish novel).

The last section, Applying corpus linguistics to other areas of research, is introduced by Andersen’s work whose aim is to give an account of how corpora and corpus linguistic methods can be used by researcher who wants to pursue a sociolinguistic research question. O’Halloran examines how to use corpus linguistics in the study of media discourse, and focuses on Critical Discourse Analysis. In the following chapter, Cotterill explores the issues related to the use of corpus linguistics in forensic linguistic analysis, including both its
potential advantages and also some of the methodological challenges associated with its use. Àdel’s chapter is related to the application of corpus linguistics in the study of political discourse (the focus is on linguistic analysis rather than on political comments). In the chapter written by Atkins and Harvey, the authors are interested in the usage of corpus linguistics in the study of health communication and illustrate how some of the corpus linguistic methodologies presented in this book can be usefully adopted in the field of healthcare communication studies. Farr analyses instead the use of corpora in teacher education (with major emphasis on Language Teacher Education) and introduces three types of relevant corpora: corpora of classroom language; learner corpora; and pedagogic corpora. Moreover, the author examines the use of specialized corpora, aiming at developing language awareness skills. The last contribution, written by Barker, is related to the use of corpora in language testing (a field concerned with the assessment of language proficiency and knowledge, performance or application of individuals in a variety of contexts and for a range of purposes).

The overall book is a complete and well-organised guide in corpus linguistics and offer to the reader a good overview of the field, in its relationships with the mains aspects of language, such as lexis, grammar and discourse. A well-argued discussion of the merits and the limits of computational tools in language analysis is provided, as well as a presentation of the future developments of the discipline. Unfortunately, most of the examples and materials come from English: a variety of examples coming from corpora in other languages would have been useful to better understand the power corpus linguistics can have in the analysis of (more complicated) language structures. In particular, the chapters dedicated to pragmatics are just a few when compared to other linguistic phenomena: for instance, it would have been interesting a more detailed discussion on the detection and analysis of (lexicalised/dead) metaphors, irony, and others figure of speech.

REFERENCES


