INTRODUCTION
Converging ways of applying corpus linguistics

MARÍA LUISA CARRIÓ PASTOR*
Universitat Politècnica de València
SUSAN HUNSTON
University of Birmingham

Corpus analysis is an area of research that has broadened the scope of a number of different fields of language analysis. One aspect of this research is quantitative. For more than sixty years, linguists have demonstrated that language features can be counted and frequencies calculated, and that these data are useful for the interpretation and understanding of language. For this reason, corpus analysis has been used in several fields of knowledge to support or challenge hypotheses and theories. In this volume our intention is to show that corpus analysis not only deals with a large amount of numbers and quantities, it also comprises studies that consider both quantitative and qualitative analysis.

Although the various writers use current techniques to compile and investigate corpora, our main interest is in how researchers apply corpus analysis. To this end, we include papers that cover a range of issues. The discourse types investigated include academic discourse, literary texts and teaching materials. The papers explore topics such as modality, cognition, language learning, lexicography, terminology, and typologies and employ approaches ranging from comparative analysis to genre studies. Taken together, the papers in this special issue have been selected to provide readers with an example of how researchers are developing and exploiting corpus methods to improve linguistic research.

Corpus research leads to both quantitative and qualitative insights. Since the early days of the Brown corpus in the 1960s (Meyer 2002), the calculation of relative frequency of linguistic features has enriched models of grammar, such as those of Quirk et al (1985), for example in the comprehensive Biber et al (1999), or of Halliday (1994), for example in Matthiessen (2006). Studies such as these have been particularly valuable in demonstrating differences between registers (e.g. Biber 1988) and in calculating

language change over time (e.g. Leech et al 2009). The linguistic features measured are aligned with given models of grammar, but many studies focus on semantic or pragmatic themes (e.g. Conrad and Biber 2000 on stance or Semino and Short 2004 on the representation of speech and thought) or on discourse (e.g. Baker 2006). Extensive work has also been carried out on lexis (Sinclair 1991; 2003; 2004) and in particular on the central role of collocation and phraseology (e.g. Hunston 2002; 2011). This has led to proposals of new concepts in language description, such as units of meaning (Sinclair 2004), and in psycholinguistics, such as lexical priming (Hoey 2005).

As the study of language in use, corpus linguistics is unsurprisingly of value to Applied Linguistics. Corpus findings have been useful to lexicography, language teaching, translation, stylistics, and the study of ideology in discourse, among other fields. All the papers in this volume bring together descriptions of corpora and their applications. And although the volume covers a broad range of topics, we are conscious that it gives only a glimpse of all the linguistic analysis based on corpus studies.

In the first paper included in this special issue, Alonso Almeida considers the analysis of evidential stance adverbs with the help of corpus tools. He describes the use of the single stance adverbials showing evidential meaning in a corpus of scientific computing papers, extracted from a multidisciplinary corpus of English research papers from the fields of computing, law and medicine. The specialised corpus chosen in this study allows the author to suggest an authorial tendency to use adverbs in the field of clarity, indicating authorial commitment towards the contents. This paper demonstrates the use of a corpus to investigate a single language feature in detail, and to draw conclusions about the nature of a given academic discipline.

In the second study, Candel Mora and Carrió Pastor test information about term variation from a specialised dictionary against occurrences of the terms in a corpus of texts about climate change and environmental engineering. Analysing several terms, they draw conclusions about what motivates the selection of one term from a set of possible variants. The research informs the compilation of dictionaries but also the study of lexis in specialised discourse.

Next, Cantos Gómez analyses the lexical profile used in the tales of Edgar Allan Poe. His hypothesis is that Poe's bipolar disorder is reflected in change in the language he uses in his writing at different chronological periods. Various measures, such as type-token ratio, word and sentence length are used to identify potential periods of elation and depression in Poe's life.

Criado Sánchez and Sánchez Pérez undertake an analysis of lexis in two coursebooks of English. Their aim is to discover which book offers learners more repetition of the items to be learned, building on the view that repetition is important for language learning. This research offers quantitative information upon which an evaluation of the coursebooks can be built, and by extension offers a methodology for evaluating other coursebooks in the same way.

De la Cruz Cabanillas and Tejedor Martínez examine anglicisms used in Spain, using a corpus compiled of texts from a selection of topics and genres. The work is

quantitative, answering questions about how many anglicisms, and what range of anglicisms, are used in a variety of sources, both technical and popular. It is also qualitative in that it sets out the options that exist for using and assimilating English words in Spanish. The paper demonstrates the infiltration of English into Spanish but also shows in detail the complexity of integrating foreign vocabulary into a language.

Martín Arista explains the steps followed for the construction of a three-dimensional lexicon. The author details the *Nerthus* project which analyses the Old English lexicon based on up-to-date linguistic theory. This lexical database offers an exhaustive description of the lexicon as well as an explanation for this linguistic component based on hierarchy and inheritance. Graph theory provides the geometry of the model of the derivational map in such a way that the lexical inventory is represented as a set of nodes and lexical and morphological inheritance are represented as a set of edges. Finally, the author represents the derivational map in 3D and relevant steps have been taken to construct the dynamic lexical space.

In the seventh and last paper included in this special issue, O'Donnell explains how learner corpora can be used to design a grammar for English as a foreign language. The author details the implications of the *Treacle* project: the compilation of learner corpora for the redesign of English grammar curricula in Spanish universities. The paper describes materials for assessing learner performance from these corpora. The corpora are used to identify both the clause structures used by the learner and their errors. This information is used to profile individual students and prescribe individual learning materials. An interesting feature of this paper is that it both highlights and solves the problems of treating texts in learner corpora as an undifferentiated set.

In conclusion, this collection aims to showcase just some of the ways researchers are using corpus linguistic. All the papers in this special issue contribute to the advancement of corpus linguistics itself, and also demonstrate the value of careful corpus use in the investigation of a variety of topics.

NOTES

* Correspondence to: María Luisa Carrió Pastor. Departamento de Lingüística Aplicada. Universitat Politècnica de València. Camino de Vera, 14. 46022. Valencia. E-mail: lcarrio@idm.upv.es.

REFERENCES

Baker, P. 2006. *Using corpora in discourse analysis*. London/New York: Continuum.

Biber, D. 1988. *Variation in Speech and Writing*. Cambridge: CUP.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Conrad, S. and Biber, D. 2000. "Adverbial marking of stance in speech and writing". *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Eds. S. Hunston and G. Thompson. Oxford: Oxford University Press. 57-73.

Halliday, M. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hunston, S. 2011. *Corpus Approaches to Evaluation. Phraseology and Evaluative Language*. London: Taylor and Francis.

Hunston, S. 2002. *Corpora Applied Linguistics*. Cambridge: Cambridge University Press.

Leech, G., M. Hundt, C. Mair and N. Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

Matthiessen, C. 2006. "Frequency profiles of some basic grammatical systems: an interim report". *System and Corpus: Exploring Connections*. Eds. G. Thompson and S. Hunston. London: Equinox. 103-142.

Meyer, C. F. 2002. *English Corpus Linguistics*. Cambridge: Cambridge University Press.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Semino, E. and M. Short. 2004. *Corpus Stylistics. Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. 2003. *Reading Concordances*. London: Longman.

Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.