



## Introduction

Corpus linguistics is an advancing discipline and a rapidly changing field. Its evolution over the last three decades has gone through different stages, and since the late 1990s it has been paced up vertiginously, largely as a result of the Information Technology boom. In the 1980s, when it was still in its infancy, corpus research was almost exclusively focused on lexical aspects of language description, and virtually all languages except English remained neglected from the corpus linguistics literature. Most practical applications were oriented towards lexicography, and theoretically, the use of corpora in language description was strongly associated with one specific research tradition, namely, with neo-Firthian linguistics and British Contextualism. The OSTI Report (re-edited in Krishnamurthy, 2004) and the COBUILD Project (Sinclair, 1987, 1991) might well be mentioned as perfect exponents of what we could describe as pioneer, first-generation corpus research.

Currently, corpus linguistics has grown to the extent of covering literally all areas of language research, including phonology, morphology, syntax and even pragmatics and discourse analysis, among others. Admittedly, English remains by far the most studied language in the corpus linguistics literature, but more and more attention is being paid to other languages as well, and in fact, there are dozens of languages for which large-scale corpora have already been compiled. Moreover, the applicability potential of corpus linguistics nowadays has reached areas as diverse as natural language processing (Souter and Atwell, 1993), foreign language teaching (Harris and Moreno Jaén, 2010), translation studies (Granger et al., 2003) and clinical linguistics (Cantos, 2010), *inter alia*.

Another consequence of the expansion of corpus linguistics has been a more complex relation between theory and methodology. The term “corpus linguistics” is no longer a synonym for “Sinclairian linguistics”, that is, for what Tognini-Bonelli (2001) called “corpus-driven linguistics”. More and more experts are developing and using corpus-based techniques as a means of testing hypotheses elaborated in other theoretical frameworks. As a result of this evolution, a discussion has been set up between those who still see a theoretical dimension in the concept of “corpus linguistics” (Teubert, 2005; Tognini-Bonelli, 2001) and

those who see it only as a methodological framework that is in principle compatible with any theoretical approach to the language, even with generative grammar (Meyer, 2002).

The rapidity with which innovations and expansions of corpus linguistic areas of inquiry take place justifies the need to regularly compile monographs and special issues providing an update on latest developments. The present volume is conceived as a contribution to a picture of current advances made in the discipline. The papers have been selected according to their relevance for offering a broadening perspective on corpus linguistics. It has also been ensured that the topics cover both a diversity of applied fields (including foreign language teaching, ESP and automatic terminology extraction) and a variety of areas of language description (including phonology, lexis, grammar and text/discourse). The spectrum of contents covered by the volume is further enriched with the inclusion of both synchronic and diachronic studies, as well as with the use of more than one language as a research object. The main interest of the volume is in the English language, as expected in this journal, but it was also considered relevant to include several articles dealing with different aspects of language contact, contrast or interference (transfer) between English and Spanish.

The opening article is co-authored by **Michael J. Harris and Stefan Th. Gries**. This study evaluates the utility of various metrics of vowel duration variability. More specifically, they focus on how well these metrics allow to compare the speech rhythms of monolingual speakers (Mexican Spanish speakers) and bilingual English/Spanish speakers (speakers born to Mexican parents in California). The authors suggest that durational variability metrics alone may be too simplistic, and that ideally they should be complemented with corpus data in the form of lemma frequencies. The second article also deals with English-Spanish language contact, albeit from a very different perspective. **Isabel Balteiro** explores the impact of English on Spanish considered in the language of sports. She examines the appearance of sports Anglicisms in two dictionaries and in a Spanish monolingual corpus. In this analysis special attention is paid to the specific forms Anglicisms adopt in the dictionaries and in the corpus. This allows the author to compare different attitudes and behaviours towards the incorporation of Anglicisms. The characteristic stance of prescriptivists and of linguistic authorities is compared with the behaviour of the language community, as reflected in attested language use, and with the treatment of Anglicisms in an innovative dictionary.

The next two contributions belong to the field of specialized (academic) discourse analysis. **Carmen Soler-Monreal and Luz Gil-Salom** report the results obtained from a cross-language analysis of citation practice in PhD theses written in English and Spanish. The study reveals interesting differences between tendencies of English writers and Spanish writers. The authors of this investigation also suggest that the differences observed can be explained by cultural factors which shape the social position of PhD writers in relation to their examiners and to the leading scholars of the discipline. Like the previous paper, the article written by **María José Luzón Marco** has important implications for ESP teaching,

particularly for the teaching of academic writing. The author investigates the use of non-native-like combinations by Spanish learners of technical English. The study concludes that students are unaware of collocations that are typical of technical English. On this basis, the author argues in favour of using phraseological repertoires in the teaching of academic vocabulary.

The next set of articles is devoted to historical linguistics. **Ana Elina Martínez Insua** examines the distribution of *there*-constructions in the timespan from late Middle English to Present-Day English. The findings indicate that this type of construction has been productive in all the periods covered by the study, and that its frequency has increased progressively. The author also analyses in detail the distribution of *there*-constructions among text types and its discourse functions. The contribution by **Nila Vázquez, Laura Esteban-Segura and Teresa Marqués-Aguado** fills a different gap: instead of using (historical) corpora for the diachronic study of English, what they provide is a survey of the most important (computerized) English historical corpora. The authors conduct a thorough analysis of more than 20 corpora compiled in several countries. The description refers both to the texts selected (language variety, genre) and to the corpus query tools made available to researchers. The article concludes with an assessment of potentialities and limitations of the existing resources.

Next in the volume is the research carried out by **María Belén Díez-Bedmar** on errors made by Spanish students of English. This research is based on a well-established error tagging system (the *Error Tagging Manual, version 1.1.*) and a learner corpus consisting of English exams submitted as part of the University Entrance Examination. The findings lead to interesting conclusions regarding which error categories and types are the ones which pose more difficulties to Spanish students of English. The author also underlines the comparability of her findings with those of previous computer-aided analyses of errors in English exams.

The closing paper, by **Rogelio Nazar**, makes the case for an approach to term extraction based on general (language-independent) statistical principles and a method for obtaining the corresponding language-specific parameters. The method is evaluated in an empirical study of English medical terms. Despite the theoretical simplicity and the low computational cost of the algorithm, which makes no use of linguistic or ontological knowledge, the results reported indicate that the method outperforms other well-known term extraction systems.

To conclude, I would like to express my gratitude to the General Editor of IJES, Aquilino Sánchez, and to the Deputy Editor, Raquel Criado, for their helpful assistance. I would also like to extend my grateful thanks to our reviewers for their expert advice and to the authors for their cooperation. With the valuable effort made by all of them, I hope that this volume can offer a picture of some of the most recent advances made in the field of corpus linguistics.

MOISÉS ALMELA  
*Issue Editor*

## References

- Cantos, P. (2010). Analysing linguistic decline in early stage Alzheimer's disease: a corpus-based approach. In A. Sánchez & M. Almela (Eds.), *A Mosaic of Corpus Linguistics. Selected Approaches* (pp. 165-192). Frankfurt a. M.: Peter Lang.
- Granger, S., Lerot, J. & Petch-Tyson, E. (Eds.). (2003). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
- Harris, T. & Moreno Jaén, M. (Eds.). (2010). *Corpus Linguistics in Language Teaching*. Bern/New York: Peter Lang.
- Krishnamurthy, R. (Ed.). (2004). *English Collocation Studies. The OSTI Report*. London/New York: Continuum.
- Meyer, C. F. (2002). *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.
- Sinclair, J. (Ed.). (1987). *Looking Up. An Account of the COBUILD Project in Lexical Computing*. London/Glasgow: Collins.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Souter, C. & Atwell, E. (Eds.). (1993). *Corpus-based Computational Linguistics*. Amsterdam: Rodopi.
- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics* 10(1), 1-13.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.