

Review of Fuster-Márquez, Miguel, Carmen Gregori-Signes and José Santaemilia Ruiz eds. 2020. *Multiperspectives in Analysis and Corpus Design*. Granada: Comares. ISBN: 978-84-1369-009-4

Moisés Almela-Sánchez
University of Murcia / Spain

The growth of a discipline is usually welcomed by the specialised academic community, but success is quickly followed by new challenges, and accomplishment gives way to the difficult task of defining new goals. This task is often a source of controversy, because setting new goals may involve redefining boundaries. As the research scope of the discipline is expanded, its limits with neighbouring disciplines are blurred, and old debates about the genuine aims and foundational principles of the discipline may be reignited.

The evolution of corpus linguistics provides a good illustration of this process. The debate about the nature of corpus linguistics and the different ways of approaching its definition dates back to earlier stages of the discipline (see Leech 1992), but the question took on a new dimension at the turn of the century as corpus methods came to be incorporated in studies from an ever wider diversity of theoretical backgrounds (including cognitive and structural linguistics, among others), and disciplines which had remained remote from a corpus linguistic approach to language, such as psycholinguistics, turned more and more frequently to corpus research in search of triangulated evidence. This proliferation of roles attributed to corpus evidence has not been free of controversy, as different influential voices in the field hold diverging views on whether certain ways of using corpora are more genuine than others. The debate between conflicting versions of corpus linguistics was particularly intensive —and proportionally fertile— in the first decade of the new century, and the relationship between theory and methodology was soon established as a central issue in the discussion (see, among others, Tognini-Bonelli 2001; Meyer 2002; Teubert 2005; Parodi 2008; Gries 2010).

It is plausible to affirm that, over the last decade, the more expansive definitions of corpus linguistics have taken the lead. The idea of a privileged bond between corpus linguistics and particular linguistic traditions or theoretical approaches has waned in recent years, and the field has accelerated the pace of its advances in a multiplicity of directions. Today, corpus linguistics is predominantly regarded as a framework of methodological resources compatible with, and valuable to, diverse paradigms of linguistic research and areas of inter-disciplinary exchange.

The edited collection under review is an eloquent testimony to this rich diversity. The selection of papers in the volume gives concise expression to the multiplicity of perspectives and approaches that have fed the growth of corpus research and stimulated its spread across disciplinary boundaries. The volume has relatively compact dimensions. It consists of nine contributions occupying a space of less than 130 pages, a size which is not larger than average among edited volumes. Remarkably, within these compact dimensions, the editors have managed to fit a collection of papers which represent diverse areas of research, both theoretical and applied, and which serve to illustrate some of the key trends observed in contemporary corpus linguistics. Thus, the volume strikes a difficult balance between comprehensiveness and focus. The collection is both succinct and informative. In a condensed manner, it conveys a sense of the polyvalent character of corpus methods, and it shows how they can be adapted to meet the needs of varying and highly specific research demands.

The volume covers topics in various areas of linguistic research (historical linguistics, sociolinguistics, pragmatics, discourse analysis, specialised languages, translation), but there is a common thread running through the diverse parts. All the contributions contained in the collection exploit the flexibility of corpus tools and show how they can be adapted to suit the particular needs of highly specific research goals. There are three main ways in which this strategy is implemented in the contributions contained in the volume. In most of them, the authors have compiled a corpus which is specifically designed for a particular research purpose or project. This is the case of the chapters authored by Arinas Pellón and Anesa (pp. 1–13), Pérez Ruiz and Ortego Antón (pp. 15–31), Verdaguer, Castaño and Laso (pp. 62–72), Serrat Roozen (pp. 73–88), Moreno-Sandoval, Gisbert and Montoro (pp. 89–102), and Vázquez García and Fernández-Montraveta (pp. 115–127). In other studies, the authors take full advantage of the internal structure of existing corpora. The contributions by Rodríguez-Abruñeiras (pp.

33–45) and Tamaredo (pp. 47–60) are paradigmatic examples of how to exploit the potential of subcorpora divisions for conducting comparisons of multiple descriptive variables. Finally, the contribution by Romero-Barranco (pp. 103–114) represents a third way of exploiting the versatility of corpus tools, since it highlights the possibility to adapt the use of particular tools to heterogeneous types of corpora. In particular, he shows that corpus tools which were originally designed to process Present-day English can also be employed in historical linguistics, provided the appropriate techniques are applied.

As befits a volume on corpus linguistics, all the contributions devote substantial attention to the description of methodological aspects. In some chapters, this special emphasis includes a detailed account of the criteria applied in the design of a specially created corpus. In other chapters, the emphasis on methodological aspects takes a different form, with a focus on the process of corpus annotation, on the adaptation of part-of-speech tagging tools, or on the selection of subcorpora. Overall, the collection highlights the potential of the corpus linguistic methodological framework for providing tailor-made solutions to highly specific research objectives.

The volume opens with an introduction by the editors, as is customary in this type of collections, followed by the chapter “Advanced-fee scams: A corpus and genre analysis” by Ismael Arinas Pellón and Patrizia Anesa. This paper analyses the language used in scam emails. The data are extracted from the *Corpus of Advanced-Fee Scams* (CAFS), a corpus consisting of more than 500 emails. The analytical framework is multidisciplinary, as it combines insights from neo-Firthian linguistics, genre analysis, and psychology. The identification of linguistic patterns is based on the classical Sinclairian model of extended lexical analysis—expounded also by Stubbs (2002)—with its distinction of four main descriptive categories: collocation, colligation, semantic preference, and semantic prosody. The patterns detected in the corpus are then related to categories of motivational choices and persuasion strategies. One of the most interesting conclusions from the study is that scam emails can be analysed as a variant of sales promotion letters, since they contain similar rhetorical moves, offer similar types of incentives to the recipients and use similar strategies to generate credibility. As the authors point out, research of this type, which identifies patterns of language use in fraudulent emails, can contribute to the development of systems capable of detecting and neutralising these attempts. Another potential application of this type of research is to help educate and alert the public about the typical characteristics of scam emails.

The second paper is “El sabor de las manzanas: análisis contrastivo (español-inglés) de la terminología objetiva referida a la experiencia sensorial del gusto” by Leonor Pérez Ruiz and María Teresa Ortego Antón. The language patterns analysed in this study correspond to the description of gustatory perceptions. The data are obtained from two comparable corpora (in English and Spanish, respectively) consisting of fact sheets on apples gathered from websites of food companies. The results from the study highlight the richness of the terminology employed to describe gustatory sensations. The conclusions also indicate that these descriptions tend to focus on four main aspects, namely, 1) the degree of sweetness/acidity, 2) the evocation of other types of food and beverage, 3) the aroma, and 4) the touch, and that they are often accompanied by lexical intensifiers and downtoners which help to convey subtle nuances. The study points to potential applications in the marketing strategies used by food companies.

The third contribution is “Two example markers in and beyond exemplification: Dialectal, register and pragmatic considerations in the 21st century” by Paula Rodríguez-Abruñeiras. This study provides a thorough analysis of the use of two example markers (*for example* and *for instance*) in two corpora representing different geographical varieties of English: *British English 2006* (BrE06) and *American English 2006* (AmE06). The author applies a threefold typology of the uses of exemplary markers—exemplification, selection, argumentation—and analyses the distribution of these uses in different text types of the two corpora. This serves to take into account the interplay of register and dialectal variables. The analysis is further enriched with the consideration of different positions occupied by example markers (before their scope domain, after their scope domain, and in the middle of the example) and an analysis of their effects on the pragmatic functions. The results indicate that different positions tend to be associated with different pragmatic nuances, such as focus or mitigation. In sum, the study provides a valuable contribution to the analysis of discourse markers in English, since it offers a highly systematic and fine-grained description and takes various relevant aspects into account (dialect, register, position).

The study of language variation is also at the heart of the next contribution “Probabilistic grammars across registers: Pronominal subject expression in some varieties of English” by Iván Tamaredo. This paper investigates which factors, both language-internal and language-external, act as the most effective determinants of the choice between overt and omitted pronominal subjects. The data analysed are obtained from

three components (British, Indian, Singaporean) of the *International Corpus of English* (ICE), and the analytical framework combines elements of probabilistic grammar and of research into World Englishes. Following a sophisticated quantitative and qualitative analysis, the author concludes that clause position and coordination are the most important language-internal constraints on the distribution of pronoun omission across varieties, modes of production, and levels of formality, and that mode of production and level of formality are the most powerful language-external factors. This paper is remarkable for its methodological rigour and depth of analysis.

The next contribution, entitled “Semantic frames in *SciE-Lex*” (Isabel Verdaguer, Emilia Castaño and Natalia Judith Laso), presents recent advances in a specialised lexicographic resource. *SciE-Lex* is a lexical database of biomedical English developed by the *GreLic Research Group* at the University of Barcelona.¹ The empirical data for this database are obtained from the *Health Science Corpus* (HSC), compiled by the same research group. In the current stage of development of this lexicographic project, the database is being enriched with information about semantic structures above the level of the individual lexical items. This will be useful for integrating the description of words that share a semantic background. The theoretical model applied is informed by the Fillmorean notion of ‘semantic frame’. This paper is thus a good example of how a corpus linguistic methodology can be combined with a theoretical framework informed by cognitive linguistics. The proposal is illustrated with the analysis of two verbs, *to block* and *to inhibit*, which in the *Health Science Corpus* are used to evoke the frame ‘Hindering’. The results of the analysis highlight the specific properties of this frame in biomedical English, compared to its description for general English in *FrameNet*. As the authors explain, the results obtained from this type of research can be used to assist dictionary users in their scientific writing.

The title of the sixth contribution in the volume is “Accesibilidad, traducción audiovisual y normas en la subtitulación online: EMPAC (*EuroparlTV Multimedia Parallel Corpus*)” by Iris Serrat Roozen. The goal of this paper is to find out whether the subtitling of the online television channel EuroparlTV conforms to the norms of audiovisual translation commonly accepted in more traditional media (TV, DVD, cinema, etc.). The corpus compiled for this purpose is the *EuroparlTV Multimedia Parallel Corpus* (EMPAC), consisting of audiovisual documents hosted in the aforementioned

¹ http://www.ub.edu/grelic/eng/?page_id=13

television channel. In particular, the study focuses on the analysis of four features related to reading speed —characters per second, characters per line, pauses between subtitles, and segmentation— and it sets out to determine whether they comply with standard recommendations. The conclusion is that, in general, they do not follow such norms, although the extent to which they deviate from them shows variations depending on the year and on the particular feature under scrutiny. The author discusses implications for the accessibility of online content.

The compilation of a specialised financial corpus is the focus of the next contribution: “*FinT-esp*: A corpus of financial reports in Spanish” by Antonio Moreno-Sandoval, Ana Gisbert and Helena Montoro. The paper provides a detailed description of the steps taken in the process of creating a corpus of Spanish financial narratives. The corpus (*FinT-esp*) consists of annual reports and financial statements published on corporate websites of companies listed in the *Madrid Stock Exchange* for the 2014–2017 period. Additionally, the authors explain the reasons for creating a more specific corpus consisting of letters to shareholders, which constitute a particularly relevant section in annual reports. A further distinction is made between two subcorpora consisting of letters to shareholders written by Presidents and by CEOs, respectively (these are expected to articulate different types of narrative). The paper offers a meticulous justification for the decisions made in the design of the corpus, and it illustrates how this resource can facilitate the application of corpus linguistic and computational techniques to analyse financial texts in Spanish.

The contribution by Jesús Romero-Barranco addresses a problem which specifically affects the creation and analysis of historical corpora. The title of this chapter is “Spelling normalisation and POS-tagging of historical corpora: The case of GUL, *MS Hunter 135* (ff. 34r-121v).” The paper highlights the benefits that the normalisation of spelling can offer for POS-tagging. This is illustrated with the processing of a specific manuscript: *MS Hunter 135*, a medical volume written in the first half of the sixteenth century. The tool for normalising spelling which is applied in this study is VARD, developed at the University of Lancaster, and the POS-tagging system is CLAWS. The results indicate that the accuracy of this POS-tagger for specific parts of the *MS Hunter 135* text can be increased by approximately 15 per cent if spelling is normalised. Based on these results, the author argues that tools which were originally designed to process

Present-day English can be adapted to historical corpora if they are complemented by appropriate strategies.

The collection ends with the chapter “Annotating factuality in the TAGFACT corpus” by Glòria Vázquez García and Ana Fernández-Montraveta. This contribution provides a detailed account of the annotation scheme devised in the TAGFACT project. The aim of this project is to create an automatic tool for the annotation of factuality, i.e. the degree of certainty with which situations are presented in texts. In principle, the tool has been created for the annotation of a Spanish corpus, but the authors argue that it can also be applied to other languages. The paper explains the criteria used for selecting the predicates to be annotated and the type of linguistic clues employed to establish the factual status. Another important aspect which receives special attention from the authors is the classification of situations into dynamic and non-dynamic ones. The authors underline the innovative character of their contribution by remarking that there is no other resource with equivalent characteristics for Spanish.

Through this diversity of topics, lines of research and applications, the selection of papers covered in the volume will give the reader an accurate portrayal of one of the key aspects that is marking the evolution of contemporary corpus linguistics, namely its tendency to cross the traditional boundaries of the discipline and to be diversified with the incorporation of a broad range of linguistic paradigms and inter-disciplinary exchanges. This does not mean that the idea of corpus linguistics as a theoretically specific and relatively homogeneous field, defined by a close connection with a particular linguistic tradition, has been completely abandoned. In fact, a substantial amount of the corpus linguistic literature produced today has a clear neo-Firthian background. However, the broader approach to the concept of corpus linguistics has been gaining ground in recent years. The number of scholars undertaking corpus research from diverse perspectives has been increasing in the last decade, and this has contributed to highlighting the potential of corpora as a pool of methodological resources compatible with multiple theories and paradigms. The volume reviewed here is a reflection of this trend and, therefore, it will be useful for readers who want to keep up-to-date with developments in the field.

REFERENCES

- Gries, Stefan Th. 2010. Corpus linguistics and theoretical linguistics. A love–hate relationship? Not necessarily... *International Journal of Corpus Linguistics* 15/3: 327–343.
- Leech, Geoffrey. 1992. Corpora and theories of linguistic performance. In Jan Svartvik ed. *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, 105–122.
- Meyer, Charles F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Parodi, Giovanni. 2008. Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada* 46/1: 93–119.
- Stubbs, Michael. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Teubert, Wolfgang. 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics* 10/1: 1–13.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Reviewed by

Moisés Almela-Sánchez

Departamento de Filología Inglesa

Facultad de Letras

C/ Santo Cristo s/n

30011 Murcia

e-mail: moisesal@um.es