

Review of Cantos-Gómez, Pascual and Moisés Almela Sánchez eds. 2018. *Lexical Collocation Analysis: Advances and Applications*. Heidelberg: Springer. ISBN: 978-3-319-92581-3. <https://doi.org/10.1007/978-3-319-92582-0>

Pedro A. Fuertes Olivera
University of Valladolid / Spain & University of Stellenbosch / South Africa

The promotional leaflet of the *Quantitative Methods in the Humanities and Social Sciences* book series indicates that this is a book series “designed to foster research-based conversation” based on applications of “computational analysis, statistical models, computer-based programs, and other quantitative methods.” One of the books in this series is *Lexical Collocation Analysis: Advances and Applications*, edited by Dr Pascual Cantos-Gómez and Dr Moisés Almela-Sánchez (University of Murcia, Spain). The editors indicate in the Introduction that the book re-examines the borderline phenomenon of ‘collocation’, a concept that is subjected to different, sometimes conflicting, interpretation in linguistics. This book aims to favour an integration of perspectives that will provide a kind of standardisation of the concept of collocation, perhaps one of the most productive and difficult areas of research over the decades following the introduction of the concept, which is usually attributed to J. Firth. The editors are right to argue that collocational studies “have played a central role in the *lexicalist turn* of the last decades and in the reformulation of the boundaries between vocabulary and grammar” (v). They mention Sinclair’s *idiom principle* and Hoey’s *lexical priming* as good “epitomes of this tendency” and indicate that the results of these and other theoretical studies have had practical applications, especially in lexicography, second language teaching/learning, and computational linguistics.

The editors are also right that collocational analysis has enormously benefited from the incorporation of the new technologies into the tools of linguistic descriptions. Hence the book “lays special emphasis on the coupling of collocational research and

computational corpus tools” (vi). In other words, the common denominator of the papers presented in the book under review “is the use of computational corpora and quantitative techniques as a means to explore aspects of language patterning that overlap the boundaries between lexis and grammar” (vi). In sum, this book offers an up-to-date analysis of the concept of collocation as this is approached in current lexicogrammar analysis carried out under the theoretical framework of corpus studies.

The book is divided into 6 chapters. In “Is language a construction? A proposal for looking at collocations, valency, argument structure and other constructions,” Thomas Herbst argues “in favour of not regarding collocation and valency as strictly discrete categories but rather seeing them as near neighbours in the lexis-grammar continuum” (1). Herbst claims that collocational and valency phenomena are better understood in terms of a rather modified concept of constructions. His conclusion is that Goldberg’s credo “It’s construction all the way down” should be modified to “It’s collexemes (or items) all the way down” (18). This ‘new credo’ will especially benefit second language learners who will be involved in overcoming traditional grammar books and dictionaries. I am especially convinced that this is something that deserves more attention from both grammarians and lexicographers who must focus on presenting collexemes as central units of language and offer them the best possible description.

Chapter 2 “Bridging collocational and syntactic analysis” by Violeta Seretan proposes the “coupling of collocational and syntactic analyses” (23) because one type of analysis will benefit the other. This conclusion is emphasised after reviewing the literature on both types of analysis and surveying “the work devoted to exploiting collocational resources for syntactic parsing” (23). It is interesting to highlight the review Seretan devotes to the works “that take into account the advances made in one area to foster the other area and vice versa” (35). These works really show that we need a better understanding of the proposed coupling which will allow researchers to improve language understanding.

Sánchez-Berriel, Santana Suárez, Gutiérrez Rodríguez and Pérez Aguiar’s “Network analysis techniques applied to dictionaries for identifying semantics in lexical Spanish collocations” (Chapter 3) offers an innovative proposal, which basically consists in using Hausmann’s *collocates* and *bases* for complementing corpus data and dictionaries in the identification of collocations and their properties. Dictionary

definitions are typically used as “a source of information to support the results obtained by the automatic extraction of collocations from a text corpus” (39) but definitions do not offer information on other important aspects of collocations, e.g. they do not distinguish if the combination is a ‘collocation’ in terms of the English or Russian tradition neither do they differentiate between functional and lexical collocations. To improve the deficiencies observed, they have constructed a graph database of word relationships with which they have built a complex system. This database offers better results than relational databases, especially because the “design of the lexical database model has facilitated the use of network analysis tools that discriminate different categories of collocations, particularly functional and lexical collocations” (53). The chapter convincingly shows the approach adopted and can be an inspiration for researches who want to replicate it, perhaps using different dictionaries and corpora to those employed in this chapter.

In his chapter “Collocation graphs and networks: Selected applications,” Vaclav Brezina “explains the potential of collocational graphs and networks both as a visualization tool and as an analytical technique” (vi). He indicates that the notion of collocation graphs and networks goes beyond the traditional representation of collocational relationships in tabular forms. Instead, the collocational graphs and networks is a technique that can be used in, say, (i) discourse analysis; (ii) language learning research; and (iii) lexicography. He provides three case studies of how this technique really works. For instance, in lexicography he refers to the *Sketch Engine* which implements word sketches, “i.e. collocations of a word of interest categorised according to their syntactic position” (74). He indicates that dictionaries, including dictionaries that use word sketches as a methodology, do not typically include words semantically related to the entry, for instance, they do not include conceptual relations. He claims that the inclusion of conceptual relations will be very positive as they can greatly help lexicographers in lexicographic descriptions of words “beyond the usual parameters observed in electronic lexicography” (74). He supports this claim by building collocation networks related with metaphors such as TIME IS MONEY, LOVE IS A JOURNEY and ARGUMENT IS WAR. The results of his analysis are displayed in several figures and allow him to conclude that corpora provide evidence of conceptual metaphors in everyday language use and that collocation networks “automatically identify the overlaps between collocates in multiple nodes (shared collocations)” (80).

In other words, the identification of these overlaps will allow lexicographers to identify word relationships which demonstrate that “relationships between words makes [sic] collocation networks an ideal lexicographic tool” (81).

Alexander Wahl and Stefan Th. Gries propose in “Multi-word expressions: A novel computational approach to their bottom-up statistical extraction” (Chapter 5) a data-driven bottom-up approach “to the identification/extraction of multi-word expressions in corpora” (85). They present a recursive algorithm to identify multi-word expressions (MWE) called MERGE (Multi-word Expressions from the Recursive Grouping of Elements), which is based on “the successive combination of bigrams to form word sequences of various lengths” (85). In the chapter they explain the use of their created algorithm on two corpora and test its performance for extracting MWE. The chapter is a good example of how to perform high quality research in this field. Firstly, they offer their own definition of MWE. Secondly, they explain how they extracted MWE from corpora. Thirdly, they offer an empirical evaluation of the algorithm. Finally, they discuss their results and conclude that MERGE exhibits strong similarities to humanlike knowledge of formulaic language. In other words, this chapter offers an interesting example of how to combine linguistic theory, corpus technology, and statistical knowhow to identify MWEs and work with them.

Peter Uhrig, Stefan Evert, and Thomas Proisl’s “Collocation candidate extraction from dependency-annotated corpora: Exploring differences across parsers and dependency annotation schemes” (Chapter 6) evaluates several parsers on two corpora with twenty different association measures plus several frequency thresholds. For carrying out such an investigation they analyse six different types of collocations against the second edition of the *Oxford Collocation Dictionary for Students of English*. Their analysis shows that although the extraction of collocation candidates is subjected to different possibilities, they recommend the use of *spaCy*, “a robust parser with good results on all relations” (135). Regarding the association measures, they also conclude that log-likelihood works well and therefore they recommend it for collocation research. This recommendation, however, must be handled with care, especially in lexicography as lexicographers would benefit if they select “different association measures for the different relations” (135).

As a conclusion of this review, I can say that this collection of book chapters is well-selected, offers up-to-date research on collocational analysis, presents very good

examples of high-quality research, and is well-edited and proofread. I should highlight that the book is a must for those interested in collocational analysis, especially researchers interested in understanding the role played by “automatic linguistic annotation (part-of-speech tagging, syntactic parsing, etc.) and using semantic criteria to facilitate the identification of collocations” (Promotional leaflet). In addition, the book offers definitions of MWEs, focuses on them for “capturing the intricacies of the phenomenon of syntagmatic attraction” and considers that collocation and valency are “near neighbours in the lexis-grammar continuum.” Finally, the book illustrates the use of quantitative methods for linguistic research as this is currently done by leading scholars in the field.

Reviewed by

Pedro A. Fuertes-Olivera
Facultad de Comercio
Plaza del Campus, 1
47011 Valladolid
e-mail: pedro@emp.uva.es