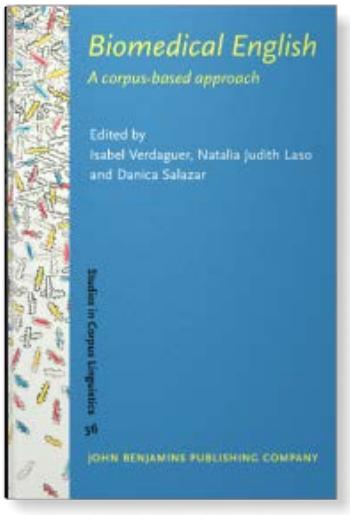

Book review:

Biomedical English: A corpus-based approach

Isabel Verdaguer, Natalia Judith Laso and Danica Salazar (eds.).

John Benjamins. Amsterdam 2013. 214 p.
https://benjamins.com/
ISBN: 9027271925, 9789027271921.

Reviewed by:
Jose Antonio Jodar Sanchez
Department of Linguistics and Language Development, San José State University, California, United States of America.
jjodarsa7@alumnes.ub.edu

English for specific purposes is a sometimes neglected discipline which, nonetheless, has important practical implications for many sectors of society. From the first studies in the 1960s and 1970s on what were then considered almost distinct language varieties of English, this field has witnessed a shift of focus towards considering the students, and not the varieties taught, as specific and distinct from one another (Robinson 1991). Nowadays, medical English, business English, legal English, aviation English and English in science and technology, *inter alia*, are considered registers which use almost the same grammatical features but make an uneven use of them in varied patterns and constructions. Even though claims on the status of these Englishes as separate registers is controversial, specialized studies, dictionaries and online resources, *SciE-Lex* being among them, facilitate the tasks of the members of those communities whose job crucially involves English.

Dictionaries in the field of biomedicine including the online *Diccionario Médico-Biológico, Histórico y Etimológico* and the Stedman's and Dorland's dictionaries are numerous and well-known to researchers, as pointed out by Navarro (*IntraMed* 2010). A few and more scarce published volumes touch upon the linguistic skills needed by biomedical scientists writing in English including Ribes, Iannarelli and Duarte (2009), Ross (2008) and Roubík, O'Neill and Smith (2005). However, their approach is usually less focused and exclusively monolingual (in English), without specific applications to the Spanish-speaking scientific community. That makes *SciE-Lex*, on which *Biomedical English* reports, one of the first attempts to ease and facilitate the task of Spanish researchers in that field. This project serves as a guiding point in their English academic skills through a simple and accessible, yet fully equipped, interface. One of the corollaries of this volume is to show how dependent meaning is on the context in which speakers use it (Barnbrook 2007: 291) and on the function for which they use it. Features such as the discourse function, the collocational environment and real examples of some of the lexical units in the database boost the usefulness of *SciE-Lex* and the informativeness of the volume with regards to the project. An added asset is its dynamic and

concrete use in the writing process, less like that of the lexicographic volumes cited above.

The volume is comprised of 10 chapters (numbered here for convenience). A bibliographic review of phraseology, collocations and their application in the field of biomedicine serves as an introduction to the following three chapters, which present the *Health Science Corpus* (*HSC* hereafter) and the database *SciE-Lex*, the authors' main project within the context of the *Lexicology and Corpus Linguistics Research Group*. The bulk of the book comprises six chapters dealing with practical applications of those two resources. Some of them target different constructional features, such as collocations of abstract nouns and adjectives in chapter 4, negation in chapter 6, passive and active sentences in chapter 7 and gender in chapter 8. Some others target specific lexical units including the verb *describe* in chapter 5 and the adjectives (*un*)*likely*, (*un*)*clear* and (*un*)*able* in chapter 6. Chapters 9 describes some of the argumentation in the *HSC* texts in metaphorical terms. The final chapter offers a follow-up on the current stage of development of the *Spanish Framenet* (*SFN* hereafter) and stresses new efforts placed on developing the syntactic interface and constructional patterning of Spanish in the project.

The volume starts with a bibliographic note on notions including collocations and "lexical bundle" (Biber *et al* 1999) as well as on the distinction between corpus-based and corpus-driven studies. Laso and Salazar acknowledge the need of varied approaches to multiword units of meaning so as to create effective tools adapted to specific purposes. An example of that is the use of Sinclair's idiom principle, of *FrameNet* and of other tools from cognitive linguistics in the analysis of biomedical discourse in *SciE-Lex* as presented in the volume. Sinclair's theoretical backbone is his "idiom principle", which states that:

> '[A] language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation'.
>
> (1987: 320)

The authors overview three approaches to the study of collocations. The London School of Linguistics, headed by Firth, Halliday and Sinclair, take statistical frequency as the crucial factor whereas the other dominant approach, that of Russian lexicography, focuses more on the grammatical patterns and taxonomy of collocations. A third approach, Gledhill's pragmatic-rhetorical one (2000), is also brought forth in the chapter to point out that *SciE-Lex* made use of all three at different stages of the construction of the database. One of this project's limitation is, as acknowledged by the authors, its corpus-based nature, that is, the analysis (although inductive) of tokens from a pre-established set of lemmas in the corpus and via a set of pre-established grammatical patterns known in English. Finally, Laso and Salazar make reference to "lexical bundles", recurrent units whose meaning is not always idiomatic but which instantiate particular structural correlates (e.g. pronoun + verb + complement). Lexical bundles, however, do not come without drawbacks since it is not clear how they are to be taught or how they are to be included in lexicographic works.

Chapter 2 focuses on the actual development of *SciE-Lex*. As Verdaguer, Laso, Guzmán, Salazar, Comelles, Castaño and Hilferty acknowledge, although the technical scientific lexis in English has been compiled in several specialized dictionaries, no tools facilitate the translation of non-technical scientific terms from and to English. First of all, the authors

undertook the morphological, syntactic and semantic annotation of the data in the *ad hoc HSC*, composed of 718 health science research articles in English. They use the coding letters *C* for 'word class', *M* for 'morphological variant', *E* for 'Spanish equivalent', *S* for 'clarification of senses', *Ver* for 'cross-references to related entries', *C* for 'patterns of occurrence', *L* for 'list of collocates', *Ex* for 'examples of real use' and *N* for 'notes to clarify usage', as illustrated with the headword *approach*. Since 'having a good command of the characteristic word combinations of a specific genre is crucial in order to establish the author's membership within the scientific community' (2013: 29), the second step targeted phraseological units. The headword *see*, in this case, illustrates this second tagging process, with the categories *Bundle*, *Discourse Function*, *Text Distribution*, *Example* and *Note* structuring the database at the syntagmatic level. The practical uses and applications of the *SciE-Lex* database described here are presented in the rest of chapters along the volume.

"Formal and Functional Variation of Lexical Bundles in Biomedical English" addresses a stage of functional organization in the process of the construction of *SciE-Lex*. In the first part, the authors explain how they changed from grouping lexical bundles according to frequency, in which case very frequent bundles had very infrequent variants, to grouping them according to their keyword, that is, the central item in the multiword units. Morphological (in number, tense, polarity, voice and person) and lexical variation (in different nouns, verbs, adjectives and prepositions being used) was taken into account to group these bundles. For example, *results show that*, *our results show that*, *these results show that* and *the results show that* were all taken to be variants of the prototypical bundle *results show that* (Salazar 2011). *It is clear* and *it is not clear*, however, are not taken as variants of a prototypical bundle because their positive and negative polarity establishes a functional difference. In the second part, a series of features and techniques used to search functional variation in the bundles including an innovative drop-down list with options for each function are detailed. Field of study, sentential position, textual position and surrounding words (Salazar & Verdaguer 2010) are used to discriminate those functions. An enhancement therein cited with regards to prior functional taxonomies of bundles is the bottom-up approach of 'assigning lexical bundles to functions rather than linking pre-established functions to individual bundles' (2013: 51). The authors endorse a view of language, in general, and English collocations, in particular, which contrasts with some logical- and generative-oriented frameworks based on *a priori* philosophies. Their project is effectively enhanced by the creation of the taxonomy from the data, that is, descriptively, and not to fit the data, that is, prescriptively.

Stemming from, a prior study of the noun *conclusion* of one of the authors (Verdaguer & Laso 2006, Laso 2009), Laso and John present in chapter 4 a study of that and other abstract nouns (*agreement*, *comparison* and *decision*) together with their concordance lines and collocations. The authors mostly find them with either descriptors indicating size, opinion, quantity and extent, or with classifiers indicating topic or relation. *Agreement* and *conclusion* frequently collocate with the former ones as in *They are in broad agreement* or *A similar conclusion was reached when the morphology of double-mutant plants was examined*. These nouns are then at odds with the general tendency for scientific language to be objective and unbiased. *Comparison* and *decision*, however, appear more frequently with relational and topical adjectives as in *direct comparison* and *behavioral food-sharing decision*. This fact is in agreement with Biber *et al*'s (1999) belief that the use of neutral adjectives like relational and topical ones reflects the attitude of describing and discussing evidence in scientific writing more than the use of descriptive adjectives.

Chapter 5, the most extensive in the volume, presents over 31 pages a very thorough analysis of the verb *describe* in scientific texts from the *HSC*. A strong preference for the verbal instantiations of *describe* (95.3%) rather than its nominalized forms (4.7%) is found in these texts. From a lexeme-specific perspective, *described* outranks any other form by being overwhelmingly used in 90.93% of cases. One of Ventura's main points, following Pattern Grammar (Hunston & Francis 2003), is that the different senses of a word will appear in different patterns, and different words instantiated in the same pattern will exhibit some core shared meaning. In one of the sections the four main patterns displayed by *describe* are discussed. On the one hand, he shows that the more prototypical meaning of *describe* as 'to depict in words' appears very frequently in patterns with active and passive verbs, with adverbs and prepositions pre- and post-modifying it, and as discourse markers or, as he labels them, "spatial and temporal guiding patterns". An instance of the former active pattern is *In conclusion, this report describes a novel pathway for the regulation of ion transport in PDEC*. On the other hand, *describe* as 'to identify and label' appears only with the pattern [(N) V-*ed* as N] as in the sentence *The animals described as wild type were C. elegans, variety Bristol, strain N2*. Ventura's descriptive contribution to the volume undoubtedly helps exemplify the varied range of syntactic and semantic patterns captured by the *HSC*.

"Negation in Biomedical English" offers some observations on the behaviour of negation, especially in relation to the adjectives (*un*)*likely*, (*un*)*clear* and (*un*)*able*. Interestingly, *unlikely* is more common than *not likely*, probably because its appearance with the pronoun *it* makes it more tentative and evasive with regards to scientific commitment, as Laso, Comelles and Verdaguer point out. It is also noticeable that both are used mostly when indicating the consequences of a phenomenon and illustrate the function of textual hedging. With the function of showing (un)certainty, *clear* and *unclear*/*not clear* show very similar results in terms of frequency in the *HSC*. The latter is usually followed by a wh-clause and shows up in the discussion, but not the results section, evincing the more assertive attitude of scientists in their results than in the discussion stemming from those results. As with *unlikely*, *unable* is favoured against *not able* with regards to frequency, and it is usually accompanied by modals in the function of hedging devices. A notable use, as illustrated by *Although the rate was low, we were able to recover three clone from cell lines*, is that of *able* with human subjects in the superordinate clause to indicate success of the authors. Overall, negative bundles are frequently used in constructions indicating cause and consequence, being affixal negation more common than clausal one. The section where these adjectives in their negative manifestation feature more prominently are the results and discussion sections.

Chapter 7, the only one in the volume that analyzes data coming from a language other than English, addresses the frequency, type of verbs, functions and language- and field-specific preferences of personal and impersonal (active/passive) constructions. Journals in medicine and mathematics in English and Spanish are used to extract data that show 'the prevalence of passive features that serve to reduce the authorial presence in the description of scientific procedures' (2013: 121) and the 'marked preference for personal rather than impersonal forms in the more abstract, logic-based discipline of Mathematics' (*ibid*). The former may be, according to Salazar, Ventura and Verdaguer, due to the medical focus on research procedures instead of in the researchers themselves, a captured by Tang and John's (1999) role of "recounter". The latter may be due to the more needed role of the researcher as guide and "originator" (in Tang and John's terminology) in the reasoning of mathematical argumentation, as attested by the frequent use of mental, existence and communication verbs

in the articles on mathematics. The authors evince as well the propriety of each type of construction, personal and impersonal ones, for different types of field. They note the enhancing feature that *SciE-Lex*, their self-designed database for Spanish scientists, presents by providing information on, among others, the contexts in which passive constructions may be adequate in scientific biomedical discourse.

Guzmán deals with what she labels "assigned gender" in chapter 8 as expressed deviantly in nouns denoting animals in English Zoology journals. She manually inspects 47 nouns with anaphoric reference, illustrating what the author calls "Units of Anaphoric Reference". Among those nouns, the most common ones are *gorilla*, *bugs* and *rhinoceros*. Although overall a 65.73% of them have *it* as their anaphora, the remaining percentage of *he*/*she* prevents us from considering the neutral anaphora as default. Following Quirk *et al*'s (1985: 314-318) taxonomic scale, the author notes that the category comprising names of bugs, amphibians and fish use *he*/*she* predominantly. Similarly, nouns of mammals and birds use *he*/*she* as well when there are proper nouns, gender-marked nouns or sex-specific activities being referred to such as being pregnant. The evidence from these nouns extracted from the *HSC* shows that, at least in the academic register in zoology journals, neuter anaphora is not the default one.

The metaphoric pattern DISCOURSE IS A FORM OF MOTION ALONG A PATH INFLUENCED BY FORCED DYNAMICS, originally presented in Castaño's (2012) doctoral dissertation, is introduced in chapter 9 to explain discourse organization in scientific texts. Castaño, Hilferty and Verdaguer's analysis focuses on six abstracts from articles in the *Journal of Cell Biology*. After an abridged account of the interacting image schemas "Force Dynamics" and "Source-Path-Goal", the authors provide the ontological correspondences of the metaphor, that is, the links between one entity and its corresponding entity in the target domain, and the epistemic correspondences, that is, the inferences drawn from those ontological correspondences. According to these correspondences, the source is the previous knowledge, obstacles are knowledge gaps, the path is the scientific method and forces are evidence pushing towards the destination or conclusion. For instance, in [...] *much effort has been directed towards defining their molecular organization. Unfortunately, major uncertainties remain regarding their true structure in living cells*, the knowledge gap entity in the target domain is illustrated because *major uncertainties* are obstacles in the path where efforts have been moving. Similarly, in *We have applied this technique to define the structure of TIs operating from* [...], the method, that is, the way of carrying out the research, in the metaphoric target domain is instantiated by the *technique*, which corresponds to the means, that is, the way in which the trajector moves, in the MOTION ALONG A PATH. After their revealing analysis, the authors conclude by asserting that 'the rhetorical structure of scientific abstracts in the field of biology is shaped by the topology of source-path-goal and force-dynamics image schemas' (2013: 182), thus conforming to the metaphorical conceptualization of discourse as a path on which entities move.

The last chapter of the volume, "Frames, Constructions and Metaphors in Spanish FrameNet", offers an informative presentation of the most updated features of the *SFN* by its director, Carlos Subirats. First he presents an automatic semantic-role program for labelling sentences, the so-called "Shalmaneser" (Erk & Padó 2006), that requires prior syntactic tagging of the data. That program is useful in the current step of SFN, that of storing grammatical patterns associated with frames. This chapter shows how the aim of the Framenet project to account for the syntax-semantics interface, in Spanish in this particular case, can be attained through a

detailed analysis of grammatical constructions and their constructs or fillers stemming from the consideration of the semantic frames those constructions instantiate. New notions such as "construction evoking element" (CEE), "construction daughter" (CD) and "construct phrase type" (CstrPT) are used in the syntactic-constructional analysis of sentences. The author illustrates the *Comparison_equality* construction through the Spanish example *Este niño es tan alto como su padre* 'This kid is as tall as his father', which contains the "construction daughter marker" *tan* and the "construction daughter base expression" *alto*. Finally, Subirats illustrates how metaphors can be explained through frames. For instance, the still unregistered frame Penetrating_into captures the source frame of a metaphor where by penetrating into a place is trying to understand a difficult topic as in *En esta novela el autor se adentra en la particular problemática del exiliado político* 'In this novel the author delves into the special problems of political exile'. The future horizon of the SFN points towards 'working on developing more ways to integrate semantic and constructional annotation' (2014: 207).

Even though some of the chapters fit into the traditional categories of morphology (chapter 8) and syntax (chapters 4 and 5) or in the interface of both (chapters 6 and 7), others do not. The pervasive cognitive operation of metaphor that was brought about with the publication of Lakoff and Johnson's seminal work *Metaphors we Live By* (1980) explains much of chapters 9 and 10. The former accounts for scientific discourse being metaphorically understood as movement along a path. Another tool from cognitive linguistics, the *FrameNet* database originally developed by Fillmore and his collaborators, serves as the basis for Subirats proposal, which applies frames to the analysis of metaphors. As the variegated topics covered by Verdaguer, Laso and Salazar's volume show, language entails more than form. Those forms, the most visible part, are mapped onto systematic meanings in entrenched patterns, giving rise to collocations, lexical bundles and multiword units in, among other types of texts, English articles in biomedical journals.

Overall, this volume shows that *SciE-Lex* can be a useful tool for Spanish scientists in the field of medicine, biology and related disciplines. Its taxonomic criteria for English collocations and expressions could be enhanced by tools such as the *BioLexicon* (Thomson *et al* 2011), which is used in text mining tasks such as processing of biomedical texts, linguistic tagging and text extraction of events and facts. Notwithstanding such minor gains, the human component in the process of selection and classification of collocated units is amply justified and proven successful. The reviewed volume and the project involved, *SciE-Lex*, represent, without a doubt, a concise, well-structured and encompassing attempt to facilitate the task of Spanish scientists reporting their research in English. Not only are the benefits of this enterprise directed towards individual researchers and teams whose credibility and ease of communication will be enhanced but also towards Spanish research in the field of biomedicine in general.

References

Barnbrook, G. (2007): Sinclair on collocation. *International Journal of Corpus Linguistics*, 12(2): 183-199.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999): *Longman grammar of spoken and written English*. Harlow: Pearson.

Castaño, E. (2012): *The embodied basis of discourse coherence*. Doctoral Dissertation, Universitat de Barcelona. http://diposit.ub.edu/dspace/bitstream/2445/35028/3/ECC_PhD_THESIS.pdf.

Erk, K. & Padó, S. (2006): Shalmaneser: A toolchain for shallow semantic parsing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy. http://www.lrec-conf.org/proceedings/lrec2006/.

Gledhill, C. (2000): *Collocations in science writing*. Tubingen: Gunter Narr.

Hunston, S. & Francis, G. (2003): *Pattern grammar: A corpus-driven approach to the lexical grammar of English* [Studies in Corpus Linguistics 4]. Amsterdam: John Benjamins.

International Computer Science Institute (n.d): *FrameNet*. University of California, Berkeley. https://framenet.icsi.berkeley.edu/fndrupal/framenet_data.

IntraMed (2010): *Entrevista a Fernando Navarro. Lenguaje médico. ¿Qué será del español en el siglo XXI?*. http://www.intramed.net/contenidover.asp?contenidoID=67287.

Lakoff, G. & Johnson, M. (1980): *Metaphors we live by*. Chicago: University of Chicago Press.

Laso, N. J. (2009): *A corpus-based study of the phraseological behaviour of abstract nouns in medical English: A needs analysis of a Spanish medical community*. Doctoral Dissertation, Universitat de Barcelona. http://diposit.ub.edu/dspace/bitstream/2445/35016/1/NJLM_THESIS.pdf.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985): *A comprehensive grammar of the English language*. London: Longman.

Ribes, R. Iannarelli, P. & Duarte, R. F. (2009): *English for biomedical scientists*. Springer.

Robinson, P. C. (1991): *ESP today: A practitioner's guide*. New York: Prentice Hall.

Ross, B. (2008): *A computational linguistic analysis of biomedical abstracts: Differences between native and korean speakers of English*. Memphis State University.

Roubík, K., O'Neill, C. & Smith, S. (2005): *English for biomedical professionals*. České vysoké učení technické v Praze.

Salazar, D. (2011): *Lexical bundles in scientific writing: A corpus-based study of native and non-native writing*. Doctoral Dissertation, Universitat de Barcelona. http://www.tdx.cat/bitstream/handle/10803/52083/DJLS_DISSERTATION.pdf;jsessionid=993BE7DA6849D1325BF365F638AA3C4B.tdx2?sequence=1.

Salazar, D. & Verdaguer, I. (2010): Multifunctionality of lexical bundles in health-science research writing. Talk at the Annual Conference of the American Association for Applied Linguistics, Atlanta, GA.

Sinclair, J. (1987): Collocation: A progress report. In R. Steele & T. Threagold (eds.): *Language topics: Essays in honour of Michael Halliday* (pp. 319-332). Amsterdam: John Benjamins.

Tang, R. & John, S. (1999): The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18: S23-S39.

Thompson, P., McNaught, J., Montemagni, S., Calzolari, N., Del Gratta, R., Lee, V., Marchi, S., Monachini, M., Pezik, P., Quochi, V., Rupp, C. J., Sasaki, Y., Venturi, G. Rebholz-Schuhmann, D. & Ananiadou, S. (2011): The BioLexicon: A large-scale terminological resource for biomedical text mining. *BMC Bioinformatics*, 12(1): 397-426. http://www.biomedcentral.com/1471-2105/12/397.

Universidad de Salamanca (2007-2011): *Diccionario médico-biológico, histórico y etimológico*. http://dicciomed.eusal.es/.

Universitat Autònoma de Barcelona (n.d): *Spanish FrameNet*. http://sfn.uab.es:8080/SFN.

Verdaguer, I. & Laso, N. J. (2006): Delexicalisation in a corpus of scientific English. In A. Honero, M. J. Luzón & S. Murillo (eds.): *Corpus Linguistics* (pp. 417-428). Bern: Peter Lang.

Verdaguer, I., Laso, N. J. & Salazar, D. (2013): *Biomedical English: A corpus-based approach* [Studies in Corpus Linguistics 56]. Amsterdam: John Benjamins Publishing.