

Isabel Verdaguer, Natalia Judith Laso and Danica Salazar. 2013. *Biomedical English. A Corpus-based Approach*. Amsterdam: John Benjamins. xiv + 214 pp. ISBN: 9-789027-203625.

ISABEL DE LA CRUZ CABANILLAS

Universidad de Alcalá

isabel.cruz@uah.es

This volume by Verdaguer, Laso and Salazar is one of the first attempts in contemporary Spain at combining the study of a highly specialised language and the corpus-based approach. The authors aim to provide a comprehensive overview of the work carried out by the research team GRELIC (Grup de Recerca en Lexicologia i Lingüística de Corpus / Lexicology and Corpus Linguistics Research Group). In fact, this is a compilation of different pieces of research conducted by the editors and other members of the team, with the collaboration of Carlos Subirats. The collection is framed within a corpus linguistics approach exploring theoretical, methodological, lexicographic and pedagogical issues of biomedical English lexicology.

The book opens with an introduction where the editors offer the rationale for the publication. They clarify the use of the term biomedicine, which encompasses “the related life sciences of biology, medicine and biochemistry” (ix), and state that the original aim of the research group was to provide non-native speakers of English with a lexicographic tool that includes information on biomedical English. In order to do so, the *Health Science Corpus* was compiled and a lexical database designed. Their main focus of study is discourse and phraseological conventions. There have been other attempts at characterising biomedical English, some of which include the use of a small sample corpus of biomedical English (Vázquez del Árbol 2006), but, in general, the characterisation of this language for specific purposes has had a different aim, for example: Gutiérrez Rodilla (1997) and Alcaraz Ariza (2000) both concentrated on the influence of the English language on medical Spanish, whereas Montalt and González (2007) were mainly concerned with the translation of medical English terms and constructions into Spanish.

In addition to the introduction, there are ten essays. The three editors contribute generously, co-authoring seven chapters, with the other three being written by other specialists in the field of corpus linguistics.

In the first chapter, “Collocations, Lexical Bundles and *SciE-Lex*: A Review of Corpus Research on Multiword Units of Meaning,” Laso and Salazar begin by reviewing the notions of collocation since Sinclair first defined the term up to the present (1991, 170), and establish the differences between collocations and other related concepts, such as lexical bundles

and idioms, as well as the various approaches in trying to classify multiword expressions. The latter are identified through automatic extraction from a given corpus; they must be dispersed in multiple texts within a register; there is correlation between the length of the bundle and its frequency, that is, the longer the bundle, the lower its frequency. Other features are also highlighted, such as their fixedness, the fact that lexical bundles are not necessarily idiomatic and their strong structural component. Finally, the authors present several functional classifications of lexical bundles. All these aspects were taken into consideration to create the database *SciE-Lex*. The chapter is well documented and helps to understand the studies carried out by the group and reported in the sections that follow.

The next chapter, “*SciE-Lex: A Lexical Database*,” by Verdaguer, Laso, Guzmán, Salazar, Comelles, Castaño and Hilferty, must also be read as a kind of introduction; it is devoted to the methodological issues underlying the design and implementation of the database, *SciE-Lex*. This is a lexicographic database developed at the Universities of Barcelona, Illes Balears and León that intends to cater for the needs of Spanish-speaking researchers and scientists in the field of biomedicine when writing academic articles in English. The innovation lies in the fact that the research team was not interested in specialised terminology that can be found in specialised lexicographic works, but in “high-frequency non-specialised lexical items and phraseology” (22).

The authors continue to explain details about the compilation of the corpus. We learn that it consists of a total of 718 “scientific research articles from prestige high-impact online journals that cover different disciplines such as medicine, biology, biochemistry and biomedicine” (22), and that the total amount of words is four million approximately. However, even though some details have been described thoroughly, such as the procedure followed to process the texts, other relevant information is missing. When researchers make use of a new corpus, they need to know its characteristics before initiating any linguistic analysis. Thus, if this database is intended for use of other researchers, further specific data about the sources is required: the titles of the publication, the number of articles or words belonging to each publication and the variety of English the texts may represent. It is true that the compilers presented the database elsewhere (Verdaguer et al. 2005 and 2008), but such important information should be present here as well.

In chapter three, “Formal and Functional Variation of Lexical Bundles in Biomedical English,” Salazar, Verdaguer, Laso, Comelles, Castaño and Hilferty explore the morphosyntactic, lexical and functional variation of lexical bundles and address the methodological difficulties this variability may entail. They aim to demonstrate that lexical bundles are subject to formal and functional variation. This piece of work is intended to be a qualitative analysis of the topic, hence there is no reference to quantitative data. However, it would also have been enriching to learn about the frequency of these lexical bundles, in order to establish their true presence in biomedical discourse.

Chapter four by Laso in collaboration with Suganthi, “A Corpus-based Analysis of the Collocational Patterning of Adjectives with Abstract Nouns in Medical English,” has a twofold purpose: to analyse the use of abstract nouns in combination with adjectives

on the part of native speakers and to exemplify how these specific patterns function in biomedical discourse. In doing so, Laso and Sunganthi wish to help authors whose native language is other than English to master the discourse conventions of this scientific genre. The findings reveal interesting issues regarding the position and the typology of the adjectives found. Evaluative adjectives are often associated to abstract nouns like *conclusion*, *agreement*, *comparison* and *decision*. The chapter highlights the value of a database like *SciE-Lex*, where not only can collocations be retrieved, but also where their real frequency is available, an element which is absent in specialised dictionaries.

Chapter five, “*As described below: A Corpus-based Approach to the verb describe in Scientific English*,” by Ventura, is devoted to the exploration of the complementation patterns of the verb *describe*, as well as its uses. Like the previous essay, it is one of the few analyses that provide the reader with quantitative data in order to show the overall frequency of the verb as well as to answer other research questions established at the beginning of the study; among them, whether there is any difference between the general pattern and other patterns related to the verb *describe*. The use of the database adds extra value to the research of complementation patterns of the verb, as it allows the author to claim that the use of the verb *describe* is preferred to other alternatives, such as the verbs *receive* and *give* followed by the noun *description*. Again, the data retrieved from the lexical database help to validate the intuitive predictions the author may have before carrying out the research. Apart from the references mentioned in the body of the text, in the final section an extensive list of articles taken from the *Health Science Corpus* is included.

In chapter six, “Negation in Biomedical English,” Laso, Comelles and Verdaguer are concerned with negative polarity. Several indicators are analysed: *not* is selected as a clausal negation, while the use of *no* and *un-* as negation exponents is also discussed. As expected, *not* was found to be the most frequent. Some other items expressing polarity, such as the adjectives *clear*, *likely*, and *able*, as well as their opposites, are examined according to the data retrieved from *SciE-Lex*. The corpus data demonstrate that some patterns are likely to appear with the negative polarity conveyed by the adjectives mentioned above. For instance, the authors claim that a proven correlation between negative bundles, such as *unable to detect*, *it is unclear* or *it seems unlikely*, and clauses of cause can be established. This finding can help researchers to organise their discourse in a more coherent and cohesive way.

In chapter seven, “A Cross-disciplinary Analysis of Personal and Impersonal Features in English and Spanish Scientific Writing,” Salazar, Ventura and Verdaguer deal with personal and impersonal features in English and Spanish scientific writing from a cross-linguistic and cross-disciplinary perspective in both a qualitative and quantitative manner. Spanish and English research articles in Medicine and Mathematics are searched in order to identify language and discipline preferences. Thus, issues such as the overall frequency of personal and impersonal constructions, the most frequently used verbs and the functions performed by these personal and impersonal constructions are explored. From the conclusions gathered from the data, the authors claim that personal constructions are preferred in Mathematics, contrary to what is common belief.

The next chapter, "Gender Assignment in Present-day Scientific English: A Case Study in the Field of Zoology Journals," provides one of the few cases in the book where detailed information on the amount of words of the subcorpus of zoology and the number of articles retrieved from each journal is given. Guzmán examines the frequencies of the pronouns *it*, *he* and *she*, when applied to nouns referring to animals and sometimes to species. The obvious conclusion is the use of *it* as the default gender, although the percentage of *he* and *she* is also relevant. According to the author, the presence of the masculine and feminine genders in Units of Anaphoric Reference cannot be considered exceptional. Guzmán defines Units of Anaphoric Reference as "fragments of texts containing an idea concerning, a description of, an event about or an action performed by the referent of a pronoun or pronouns constituting an individual *piece of communication*, a pragmatic unit which may reach one or various lines in a text and may have different structures depending on the text type" (149). The high percentage of *he* and *she* to refer to animal nouns is reinforced by the fact that they appear in scientific specialised texts. The referents are not always high-rank animals, but its use is also prevalent when sex-specific activities and states are involved (*become pregnant*).

In chapter nine, "The Metaphorical Basis of Discourse Structure," Castaño, Hilferty and Verdaguer present fundamental considerations for the metaphorical mappings occurring in discourse, particularly on the organisation of research abstracts from the field of Biology. After a qualitative assessment of research abstracts in Biology, where the authors argue that conceptual metaphors play an important role in structuring and sequencing the ideas in the text, they conclude that the discourse in this field is based on the metaphor "Discourse is a form of motion along a path influenced by force dynamics."

The book ends with an essay on "Frames, Constructions, and Metaphors in Spanish FrameNet." In this final chapter, Subirats, director of Spanish FrameNet, outlines the fundamental issues concerning the application of the frame semantics to the lexical analysis of FrameNet. To this end, the process of semantic annotation is described, as well as the software tools used. Furthermore, the analysis of metaphors as mappings between semantic frames is also explored. The volume concludes with these new approaches to the study of scientific discourse based on the tenets of Cognitive Linguistics. In this specific case, the application of FrameNet to the cross-linguistic study of scientific language may open new lines of research in the study of specialised languages.

This volume is a welcome contribution to the field of biomedical language in Spain since it provides empirical data extracted from a textual corpus. This is one of the strengths of the book: all chapters are based on the extraction of data from the *Health Science Corpus* (HSC). Another positive aspect is that the publication also relies on the *SciE-Lex*, "a lexical data-base of general English terms employed in biomedical discourse" (ix). By contrast, some of the criteria used for the compilation of the corpus, as well as the design and implementation of the database need further clarification. No reference is made to the usual criteria for the compilation of the corpus; that is, representativity, balance and sampling. The reader may imagine that the corpus has been marked up and annotated as

well, but little information is provided. It is true that these details were given elsewhere (Verdaguer 2005), but the volume would have benefited from its inclusion here.

To conclude, this book opens up a whole new line of research in the Spanish academic community because the findings cited are based on textual evidence retrieved from a corpus of research articles in biomedicine, a field where more studies using corpus linguistics were needed. The contributions in this collection also address a wide variety of issues related to biomedical language, such as lexicogrammatical patterns, lexicography, discourse analysis and its pedagogical applications. The volume is as readable and enlightening as it is well documented and informative, written by experts in the tool they are describing which was implemented from the purpose-built corpus. It would be highly advisable that the database were at the disposal of researchers and scientists interested in biomedicine in English.

#### WORKS CITED

- ALCARAZ ARIZA, María Ángeles. 2000. *Anglicismos en el lenguaje de las ciencias de la salud*. Alicante: U de Alicante.
- GUTIÉRREZ RODILLA, Bertha M. 1997. "La influencia del inglés sobre nuestro lenguaje médico." *Medicina Clínica* 108: 307-13.
- MONTALT RESURRECCIÓ, Vicent and María González Davies. 2007. *Medical Translation Step by Step*. Manchester: St Jerome.
- SINCLAIR, John 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford UP.
- VÁZQUEZ DE ÁRBOL, Esther. 2006. "La redacción del discurso biomédico: (inglés-español): rasgos principales." *Panacea* 7 (24): 307-17.
- VERDAGUER CLAVERA, Isabel and María Juan Garau. 2005. "Las combinaciones léxicas en el inglés científico: presentación de una base de datos." *Revista española de lingüística aplicada*, 1 (vol. extra): 67-84.
- VERDAGUER CLAVERA, Isabel, Anna Poch Higuera, Natalia Judith Laso Martín and Eva Giménez Domínguez. 2008. "SciE-Lex: A Lexical Database of Collocations in Scientific English for Spanish Scientists." Paper delivered at the *XIII EURALEX International Congress*, Barcelona, IULA, 15-19 July, 2008.

Received 3 September 2013

Accepted 24 February 2014

Isabel de la Cruz Cabanillas is Senior Lecturer in English at the University of Alcalá, where she teaches History of the English Language, as well as other courses on English Language and Linguistics at undergraduate and postgraduate level. Her research interests are chiefly in the field of English Historical Linguistics and related subjects, especially Lexicology and Semantics.

Address: Dpto. Filología Moderna. C/Trinidad, 3. 28801, Alcalá de Henares, Madrid, Spain. Tel.: +34 918855162. Fax: +34 918854445.