Verdaguer, Isabel, Natalia Judith Laso, and Danica Salazar (Eds). 2013. *Biomedical English: A Corpus-based Approach*. Amsterdam: John Benjamins. xiii, 214 pp.

Research in the field of language for specific purposes (LSP) is a vital part of contemporary applied linguistics, and investigation into the communicative practices of different discourse communities has important pedagogic implications and applications, particularly for those who need or want to become participants in those communities. *Biomedical English: A Corpus-based Approach*, edited by Isabel Verdaguer, Natalia Judith Laso, and Danica Salazar, provides a valuable corpus-based contribution to the field, complementing recent, related work on (bio)medical English, its associated genres, and its sociohistorical developments (e.g. Gotti and Salager-Meyer 2006; Taavitsainen and Pahta 2004).

Biomedical English comprises an introduction, 10 main chapters, and an index. The majority of the chapters present the work of GRELIC (*Grup de Recerca en Lexicologia i Lingüística de Corpus*), the Lexicology and Corpus Linguistics Research Group at the University of Barcelona, and its development and analysis of the Health Science Corpus (HSC) and the SciE-Lex database.

The editors' introduction contextualizes the volume and provides a rationale for GRELIC, whose initial main objective was to construct a database of general English terms used in biomedical research discourse, based on research articles from "the related life sciences of biology, medicine and biochemistry" (p. ix). This database, SciE-Lex, was designed and developed to be used as a reference and pedagogic tool for biomedical researchers who are non-native speakers of English (NNSE), particularly those with Spanish as a first language. The introductory chapter also briefly summarizes and comments upon the 10 main chapters, highlighting the diversity of theoretical and methodological approaches embraced by the project.

The first main chapter—the chapters are not numbered—is written by two of the editors, Laso and Salazar. "Collocations, lexical bundles and *SciE-Lex*: a review of corpus research on multiword units of meaning" provides a summary of some of the theoretical and conceptual perspectives underpinning the creation of the SciE-Lex database. Specifically, Laso and Salazar present and discuss various approaches to

collocation—probabilistic, phraseological, and rhetorical—arguing for the relevance of each and their potential complementarities. The authors also discuss lexical bundles, and how corpus-based and corpus-driven procedures might be usefully combined to create register-specific pedagogic tools such as SciE-Lex.

Against this theoretical backdrop, the second chapter, "*SciE-Lex*: a lexical database," by Verdaguer, Laso, Guzmán-González, Salazar, Comelles, Castaño, and Hilferty, examines some of the methodological considerations underlying the project. Here, the authors restate GRELIC's initial aim to create "a reference tool to help the Spanish scientific community to publish their papers in the English language" with a "focus on high-frequency non-specialised lexical items and phraseology" (pp. 21-22). With this in mind, Verdaguer et al. describe the selection and organization of their material, a corpus of "718 scientific research articles from prestige high-impact online journals that cover different disciplines such as medicine, biology, biochemistry and biomedicine," comprising approximately four million words "produced by native speakers of English" (p. 22). Two stages of corpus annotation and analysis are then described: the first, a morphologic, syntactic, semantic, and collocational analysis of the most frequent "non-technical" (p. 24) terms in the corpus, crosschecked against the Academic Word List (Coxhead 2000) and the Academic Keyword List (Paquot 2010); the second, an analysis of lexical bundles, and their organization into structural, functional, and distributional categories.

The third chapter, "Formal and functional variation of lexical bundles in biomedical English," by Salazar, Verdaguer, Laso, Comelles, Castaño, and Hilferty, examines in greater detail some of the morphosyntactic, lexical, functional, and distributional variation of the lexical bundles in the HSC, and their relevance for SciE-Lex. Based on variables such as number, tense, and voice, as well as the interchangeability of semantically related nouns, verbs, adjectives, and prepositions in specific slots in the strings, Salazar et al. identify "canonical units of meaning" in the form of "prototypical bundles" (pp. 41-42). These prototypical bundles are then classified functionally into three major groups, based on Hyland (2008): "describing research," "organizing text," and "establishing stance and interacting with the reader" (pp. 44-45). Within each of these overarching categories, Salazar et al. identify a series of subcategories (39 in all), which they label so as

to allow for easy use by and accessibility for "non-native user[s] of a pedagogical dictionary" (p. 45). The chapter also contains an in-depth discussion of some of the challenges in categorizing multifunctional lexical bundles, including variability with regard to discipline, sentential position, textual distribution, and co-textual elements.

While the first three main chapters describe and discuss various theoretical and methodological considerations in constructing and annotating the HSC and SciE-Lex, many of the subsequent chapters in *Biomedical English* deal with corpus-analytic studies of the HSC. The first of these is "A corpus-based analysis of the collocational patterning of adjectives with abstract nouns in medical English," by Laso and John. Here, the authors examine the collocational patterns of the abstract nouns/nominalizations *conclusion*, *agreement*, *comparison*, and *decision*. Laso and John study how these specific nouns are modified by certain types of adjectives: descriptors and classifiers, broadly speaking (cf. Biber et al. 1999: 508-509). The nouns *comparison* and *decision* tend to collocate with classifiers rather than descriptors, e.g. *direct/statistical/valid comparison*. But for *agreement* and *conclusion*, this tendency is reversed. In the case of *conclusion*, which Laso and John describe as particularly noteworthy (p. 67), the noun is commonly modified by evaluative descriptors such as *misleading* and *controversial*. Indeed, on the whole, the authors note that the "number of occurrences of evaluative descriptors is surprisingly high in [their] data" (p. 68), especially in comparison with Biber et al.'s (1999) observations for academic prose, in which there is "extreme reliance on classifiers" (Biber et al. 1999: 511). Awareness of these and similar collocational patterns, and their inclusion in tools such as SciE-Lex, the authors conclude, is "extremely valuable to the community of scientists whose first language is not English" (p. 69).

The fifth chapter, "*As described below*: a corpus-based approach to the verb *describe* in scientific English," by Ventura, explores the complementation patterns associated with this commonly used verb. The author also compares the use of *describe* with its nominalized form *description*. Based on data from the HSC, Ventura proposes four main patterns for *describe*, each of which comprises two subtypes: group 1, 'active' or 'passive'; group 2, 'V-*ed*' as a pre- or postmodifier; group 3, 'N V-*ed as* N' (simple categorization pattern) or 'N (be) V-ed as N' (complex categorization pattern); and group 4, '*as* [Adv] V-*ed* [Adv]'

(temporal guiding pattern) or '*as* V-*ed* [Adv]/[PP]' (spatial guiding pattern, cf. example in chapter title, *As described below*). Ventura goes on to associate these different patterns with different meanings of the verb *describe*, following the approach of Hunston and Francis (2000). One meaning, which Ventura terms prototypical, "give[s] an account or representation of in words" (p. 92). This is associated with the patterns of groups 1, 2, and 4, e.g. *this report describes a novel pathway…* (group 1, active). A second meaning is "to identify, characterise, and label" (p. 92), which Ventura associates with group 3 patterns such as *DNase IV has also been described as an essential replication factor…* (group 3, complex categorization) (p. 92). These two senses of *describe*, Ventura notes, are comparable to the semantic frames for 'describe.v' in FrameNet: 'Statement' and 'Communicate_categorization.' A shorter section in the chapter compares *describe* and *description* ('V *description*' and '*description* PP'), with the verb *describe* being the most preferred or most highly selected form in the HSC (95.3% versus 4.7%). Ventura concludes by emphasizing the interrelation of lexis, pattern, and meaning, and by arguing that awareness of variations in patterns of a particular lexical unit (in this case *describe*) helps to determine the meaning of a word, and that knowing the semantics of a lexical unit can in turn help distinguish different patterns of use (p. 99).

Laso, Comelles, and Verdaguer present a chapter on "Negation in biomedical English." They begin with a short, exploratory study, in which they find that *not* is the most frequent marker of negation in the HSC (76.4% of all clausal and affixal negative markers), followed by *no* (12.7%) and *un-* (10.8%). They then provide a more detailed investigation of three adjective pairs—*likely*/*unlikely*, *clear*/*unclear*, and *able*/*unable*—and the phraseological patterns associated with them. Laso et al. describe and compare differences in frequency of occurrence as well as differences in the patterns associated with each pair. These patterns are then assigned discourse functions based on Salazar et al.'s taxonomy (same volume; see above). The authors find, for example, that negative bundles, i.e. phraseological units that contain a negative element such as *un-* or *not*, tend to be used in clauses of cause, consequence, and contrast (subcategories of Salazar et al.'s "organizing text"). They are also commonly used in connection with hedging, e.g. *they were presumably unable to bind at the active site* (p. 113), and of course as a direct hedge in the case of *likely*/*unlikely*. Moreover, these negative

bundles are more commonly found in the Results and/or Discussion sections of the research articles in the HSC.

The seventh chapter in *Biomedical English* is "A cross-disciplinary analysis of personal and impersonal features in English and Spanish scientific writing," by Salazar, Ventura, and Verdaguer. As the title indicates, the authors look beyond biomedical English, to compare research writing across disciplines (medicine and mathematics) and across languages (Spanish and English). In order to do so, the study uses four specifically designed comparable corpora (E-MED, S-MED, E-MATH, S-MATH), rather than the HSC, and limits its examination of personal and impersonal features to first-person pronouns and passive constructions, respectively. The overall frequencies of these two features, across the four corpora, indicate clear disciplinary differences, with a marked preference for the passive in E-MED and S-MED, and a general preference for first-person pronouns in the mathematics corpora, particularly in E-MATH. The authors then investigate the usage patterns and rhetorical functions of these features. Their overall findings suggest, in general, a relatively impersonal medical literature with a focus on "replicability and universality" (p. 137), and a more explicitly interpersonal mathematics literature that guides the reader through the arguments and emphasizes a sense of solidarity and shared intellectual goals. Awareness of these disciplinary differences and, importantly, the language that construes them needs to be reflected in pedagogic materials, an argument the authors use as part of a rationale for the development of discipline-specific reference works such as SciE-Lex.

The eighth chapter, by Guzmán-González, examines the role of assigned gender in zoology research discourse. Entitled "Gender assignment in present-day scientific English: a case study in the field of zoology journals," the chapter investigates variations in the choice of the pronouns *he*, *she*, and *it* in referring anaphorically to nonhuman animals in the zoology subsection of the HSC (HSC-Z). *It* is the most common anaphoric reference to nonhuman animals in the HSC-Z (65.73% of all pronominal references to nonhuman animals), but the relatively high frequency of *he* and *she* (19.38% and 14.88%, respectively) belies, it would seem, the notion of a "detached, impersonal, neutral style" (p. 159) in the present-day English-language research discourse of zoology. Guzmán-González's findings seem to attest to a literature in which the choice of anaphoric reference to nonhuman animals is not only one of a

default neuter gender, which Guzmán-González refers to as the standard according to some scholars. Rather, it appears to depend on a variety of factors, including animacy hierarchies, sexual differentiation, and markers of acquaintance, interest, and detachment.

In "The metaphorical basis of discourse structure," Castaño, Hilferty, and Verdaguer investigate how the notion of discourse, in its text-analytic rather than, say, Foucauldian sense, can be metaphorically conceptualized as a form of motion through space and time, a journey with a starting point, a trajectory, and a destination. Based on Cognitive Metaphor Theory (e.g. Lakoff & Johnson 1980), Castaño et al. explore the hypothesis that DISCOURSE IS A FORM OF MOTION ALONG A PATH INFLUENCED BY FORCE DYNAMICS. They do so by examining six research-article abstracts from the *Journal of Cell Biology*, and show how organizational structure and certain linguistic resources reflect source-path-goal and force-dynamics image schemas. According to the authors, previous research is the main starting point of the journey, but progress is usually hampered by an obstacle, a gap in the research. Collision with such obstacles causes the discourse to deviate from its initial path, toward a new destination, the intended goal (or goals) of the research. This new path is the method or means of reaching the intended goal. The results and evidence that emerge from these methods are conceived of as forces, effecting in various ways the direction the path takes, pushing the discourse in one way and not another, until the final destination, which, in Castaño et al.'s cases, coincides with the intended goals described at the start of the abstract. In the conclusion to this chapter's own 'discourse as journey,' the authors note that "[j]ust as a well-marked path avoids detours or secondary roads and leads us straight to our destination, the way the empirical data are presented help to prevent unnecessary diversions in the process of reasoning" (p. 182).

The final chapter, Subirats' "Frames, constructions, and metaphors in Spanish FrameNet" differs somewhat from preceding chapters. Subirats' contribution does not deal with biomedical English or the compilation, annotation, or analysis of the HSC. Instead, it discusses the Spanish FrameNet (SFN) project. Subirats describes the processes of manual and automated semantic annotation, as well as recent efforts to incorporate grammatical constructions into SFN. An extensive section of the chapter deals with frame semantics and metaphor, and how metaphors are treated as mappings between different source and target semantic frames in

SFN; an interesting example for the Spanish verb *adentrarse* is provided, in which a particular metaphorical meaning is dealt with as a mapping between a 'Penetrating_into' frame and a 'Treating_difficult_topic' frame. The chapter concludes with a brief summary of current and future developments, which include cross-linguistic analysis with other versions of FrameNet, the incorporation of large-scale public-domain corpora, and the further development and integration of grammatical constructions.

Overall, *Biomedical English* is a fascinating volume that provides a variety of corpus-based perspectives on a specific mode and text-type associated with this rather broad register. The first three chapters are thorough in their documentation of the theoretical and methodological foundations of the project. Subsequent chapters, including the somewhat 'off-topic' final chapter, which the editors include for "the new horizons" it offers (p. x), explore a wide range of lexicogrammatical and semantic features. Yet, in a sense, *Biomedical English* really just scratches the surface of what a corpus-based and/or corpus-driven approach to register and discourse studies might offer. The chapters provide an intriguing sample or glimpse into the possibilities afforded by the development and analysis of the HSC and SciE-Lex database. Some of these possibilities will no doubt be explored in future work, with further valuable and revealing studies sure to emerge from this rich material.

On a more critical note, and with regard to the volume in general, this rich material could have been described and explained in more detail. Although a description of the HSC is given in the second chapter, the "prestige high-impact online journals" (p. 22) from which the material was selected are not named, nor is there mention of how this particular set of high-impact journals and the research articles therein were selected, e.g. algorithmically and/or based on recommendations from colleagues and other scholars in relevant fields. Moreover, no publication dates for the research articles in the HSC are specified. Instead, the interested reader is left to fill in these gaps by looking at the lists of referenced examples at the end of each chapter—useful in the case of specific examples, but less so in providing a more general impression of the material basis for some of the findings.

Another potential criticism regards the compilers' choice to include in the HSC only those research articles with "at least one native English speaker among their authors" (p. 23). This choice is briefly justified in terms of "native competence" (p. 23), but the potential problems of

privileging this particular group are not acknowledged, despite the insistence in a later chapter that "English is considered the lingua franca of the scientific community" (p. 55). The authors are of course clear that the HSC, like any other corpus, cannot be truly representative of the discourse as a whole, and findings from the project are therefore specific to the HSC itself (p. 23). However, the choice to exclude high-impact research articles authored entirely by non-native speakers seems to overlook or disregard the role played by these researchers in shaping the register of biomedical English. One of the pedagogic implications of this is that users of the SciE-Lex database may miss out on some of the rhetorical and linguistic diversity and hybridity of the discourse (see, for example, Mauranen 2001; Pérez-Llantada 2012), which is not necessarily bound or defined by an Anglophone center.

As the editors note in the opening of their introduction, *Biomedical English* "explores the theoretical, methodological, lexicographic and pedagogical aspects of a specific sublanguage of English" (p. ix). These first three aspects are explored in great depth, and this is undoubtedly one of the strengths of the volume. However, pedagogic considerations receive somewhat less attention. Bearing in mind the original aim of the project, as well as general interest in the pedagogic uses of corpora (e.g. Aijmer 2009; Lee & Swales 2006; Sinclair 2004), *Biomedical English* could have benefited from a dedicated chapter or two on the specific application of the HSC and SciE-Lex database for teaching, learning, or general reference purposes.

Biomedical English: A Corpus-based Approach should be of interest to a wide readership. Most obviously, this would include researchers and practitioners involved in the study and/or teaching of biomedical English. But the volume should also be of interest to scholars and practitioners engaged in other areas of LSP, particularly English for academic purposes, and who are looking to benefit from insights from other fields as a way of informing their own work. Moreover, since the book attends to many of the challenges involved in constructing and annotating corpora, it could be relevant to corpus linguists and computational linguists in general, particularly those involved in creating or developing corpora. Certain chapters are thus likely to appeal to different groups of readers. With this in mind, the fact that each chapter has a clear abstract and introduction, as well as separate reference lists, makes the volume easy to read as a whole or as a series of standalone papers.

Daniel Lees Fryer
University of Gothenburg, Sweden, and Oslo and Akershus University College of Applied Sciences, Norway

References

Aijmer, Karin (ed.) 2009. *Corpora and language teaching* (Studies in corpus linguistics 33). Amsterdam: John Benjamins.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finnegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Coxhead, Averil. 2000. "A new academic word list." *TESOL Quarterly* 34: 213-38.

Gotti, Maurizio and Françoise Salager-Meyer (eds) 2006. *Advances in medical discourse analysis: oral and written contexts* (Linguistic insights: studies in language and communication 45). Bern: Peter Lang.

Hunston, Susan and Gill Francis. 2000. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hyland, Ken. 2008. "As can be seen: lexical bundles and disciplinary variation." *English for Specific Purposes* 27: 4-21.

Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago, Illinois: University of Chicago Press.

Lee, David and John M. Swales. 2006. "A corpus-based EAP course for NNS doctoral students: moving from available specialized corpora to self-compiled corpora." *English for Specific Purposes* 25: 56-75.

Mauranen, Anna. 2001. "Descriptions or explanations? Some methodological issues in Contrastive Rhetoric." *Academic writing in context: implications and applications - papers in honour of Tony Dudley-Evans*, ed. M. Hewings. Birmingham: University of Birmingham Press. 43-54.

Paquot, Magali. 2010. *Academic vocabulary in learner writing: from extraction to analysis.* London: Continuum.

Pérez-Llantada, Carmen. 2012. *Scientific discourse and the rhetoric of globalization: the impact of culture and language*. London: Continuum.

Sinclair, John McH. (ed.) 2004. *How to use corpora in language teaching* (Studies in corpus linguistics 12). Amsterdam: John Benjamins.

Taavitsainen, Irma and Päivi Pahta (eds) 2004. *Medical and scientific writing in late medieval English*. Cambridge: Cambridge University Press.