John Benjamins Publishing Company

BOOK REVIEW

Herbst, T., Faulhaber, S. & Uhrig, P. (Eds.) 2011. *The Phraseological View of Language. A Tribute to John Sinclair.* Berlin/New York: De Gruyter Mouton. (xii + 324 pp.)

The volume presents a collection of papers focussing on the role of prefabricated chunks in language. It goes back to the workshop entitled *Chunks in Corpus Linguistics and Cognitive Linguistics* held at the Friedrich-Alexander-Universität Erlangen-Nürnberg in 2007 in honour of John McHardy Sinclair. The papers included in the volume cover a variety of topics, ranging from theoretical and practical aspects of chunking, to language variation and change, to chunk parsing in corpora. In that, the collection is somewhat mixed. However, what unites the contributions to this volume is, first of all, their emphasis on "the idiom principle" (Sinclair 1991), and, second, their approach to linguistic enquiry that rejects any division between theory and practical description of language; the two being fundamental to Sinclair's work.

The book is divided into four sections. Each section contains a collection of papers on a common theme written by specialists in the area. The present review follows the thematic (and chronological) order of the chapters as they appear in the volume. Section I is an introductory one; as such, it has as its aim to pay homage to Sinclair and his work. In "A tribute to John McHardy Sinclair", Stubbs provides a brief outline of Sinclair's career. The chapter discusses the major contributions Sinclair made to the three areas that are now strongly associated with his name: language and education, discourse analysis, and corpus-assisted lexicography (notably, the volume is structured in such a way that these areas roughly correspond to the three main sections of the volume). Sinclair's work, as Stubbs notes, has been instrumental in showing how to use empirical evidence to investigate what Sinclair deemed to be the fundamental question — the nature of units of meaning (also see Sinclair 1996).

Similarly to Chapter 1, the focus of the second chapter, "Corpus, lexis, discourse: A tribute to John Sinclair" by Johansson, is on Sinclair's contribution to and his influence on the field of corpus linguistics. Akin to Stubbs, Johansson stresses that Sinclair has always been concerned with how language is used (rather than how it *can* be used) and how it functions in communication. The key part of the chapter deals with *The OSTI Report* Sinclair was involved in in the 1960s. What makes the report unique, according to Johansson, is Sinclair's focus on lexis at the time when it was largely ignored, marginalised and delegated to the linguistic periphery.

Section II focuses on language and education; specifically, on theoretical and pedagogical aspects of collocation, one of the core areas Sinclair helped to define and advance throughout his career. Broadly, this section deals with the concept of collocation in the view of Sinclair's open-choice and idiom principle. The first chapter, "Choosing *sandy beaches* — collocations, probabemes and the idiom principle" by Herbst, raises a number of issues concerning the notions of single choice, meaningful units and storage. Herbst gives examples of different types of collocations and explains why they might be interesting or unique (given their frequency, predictability, limited combinability, etc.), from the point of view of a corpus linguist, a language teacher or a learner. He further positions collocation within various approaches (e.g. frequency-based vs. semantically-based), providing examples of collocations in English and German. However, as Herbst notes, traditional distinctions between collocations of different types are not helpful "when it comes to identifying single choices or semantic units in Sinclair's sense" (p. 37). In the chapter, Herbst touches upon an interesting but rather controversial issue — collocation storage — arguing that "cases such as *heavy rain* or *heavy drinking* represent an argument in favour of [collocation] storage" (p. 32), and that "a restricted set of possible collocates […] could be an indicator of storage" (p. 45). Although this is a rather tempting and, indeed, interesting possibility, it needs to be pointed out that corpus evidence alone can hardly provide a strong argument in favour of storage or processing (for a discussion of a complementary use of corpora and experimental data, see Gilquin & Gries 2009).

Much in line with Herbst, the following chapter "Sinclair revisited: Beyond idiom and open choice" by Siepmann further covers some of the core theoretical aspects of collocation. Siepmann puts forward a new take on the principles of idiom and open choice, arguing in favour of a third principle, "the principle of creativity" (p. 59), which is discussed in light of translation teaching. Siepmann distinguishes between co-creations, collocations, analogous creations (explained by the principle of creativity) and counter-creations, and provides rather compelling examples of English-French translations of some co-occurrences. A number of potential difficulties (concerning polysemy, stylistic and textual factors, frequency differences, etc.) as well as their implications for translation teaching are discussed in the final part of the chapter. Siepmann argues for the importance of the awareness of "words' combinatorial properties" (p. 76) in translation work, and encourages a systematic comparison of items from collocation dictionaries, the latter being somewhat of wishful thinking, though, given the scarcity of collocation dictionaries available.

The chapter that follows, "Accessing second-order collocation through lexical co-occurrence networks", would arguably be a better fit for the section that deals with the computational aspects of collocation (Section IV), rather than one on

theoretical and pedagogical aspects, given the computational approach employed by Mollet, Wray and Fitzpatrick. The chapter makes use of network modelling in order to shed further light on patterns of lexical co-occurrence. To this end, Mollet, Wray and Fitzpatrick propose, define, and explore "second-order collocations" (p. 87) — word combinations that depend on the presence or absence of other items, modelled in the form of networks. They put forward the principles of the network algorithm. Admittedly, a less mathematically apt reader might struggle with their interpretation (not least because the formula provided lacks explanation). It is argued that the network model, while a rather insufficient way of exploring first-order collocations, offers a valuable resource for exploring second-order collocations. The latter are important because they can, for example, help us to separate out different senses of the same word, or provide an additional tool to study linguistic intuition and L1 versus L2 language usage. Overall, the method, still scarcely used, promises to offer new insights into the very nature of linguistic patterning.

Granger's chapter entitled "From phraseology to pedagogy: Challenges and prospects" nicely follows those by Herbst and Siepmann, in that it introduces the reader to pedagogical aspects of collocation. Granger discusses a rather influential approach associated with teaching and learning collocation, the lexical approach, highlighting its major strengths (e.g. fluency, ease of learning) and weaknesses (e.g. lack of generative power). In doing so, Granger focuses on the complex relationship between lexis and grammar in the context of foreign language teaching and learning. A key part of the chapter is dedicated to the implementation of the lexical approach, with respect to, for example, the criteria to be taken into account when selecting phrases to be taught. Unsurprisingly, the major deciding criterion is frequency of occurrence. However, as Granger argues, it is crucial to counterbalance corpus-based frequency with other factors, such as learner variables (e.g. aptitude, learning style), learnability (e.g. processing and storage of phrases), and teachability (e.g. how to choose items that do not overload teachers and learners). With regard to the notion of learnability, research on phrasal processing and storage is still at an early stage and more psycholinguistic work is needed for us to be able to draw any valid conclusions.

The pedagogical aspects of collocation are further explored in "Chunks and the effective learner — a few remarks concerning foreign language teaching and lexicography" by Götz. Specifically, Götz discusses strategies learners should adopt when learning idiomatic language. The chapter does indeed offer a very general picture of the difficulties learners encounter acquiring collocations. However, the chapter seems to lack a clear focus and, critically, substance. Götz's proposition that it is essential for language learners to be aware of chunks and a few, somewhat

mundane, recommendations as to how to tackle the problem fail to capture the essence of the problem or engage the reader.

The final chapter of the section, "Exploring the phraseology of ESL and EFL varieties" by Nesselhauf, complements Granger's and Götz's focus on collocation and pedagogy, in that it compares the use of phraseological units by ESL and EFL learners. Both ESL and EFL, viewed as legitimate varieties of English, are evaluated against each other, as well as the baseline — native speaker data. Nesselhauf reports that the behaviour of ESL learners resembles that of EFL learners much more closely than it does native speaker data (e.g. both ESL and EFL learners overuse *play a role* and underuse *play a part*, while native speakers apparently use both phrases equally frequently). What accounts for such differences, according to Nesselhauf is not frequency, but the degree of language-internal regularities of a given phrase (i.e. *play a role* is more regular than *play a part*). Nesselhauf's arguments offer an interesting insight into ESL and EFL varieties versus native speaker usage. On the downside, the small sample size does not always warrant some of the conclusions drawn. It is further worth noting that parts of the chapter are based on previously published material (Nesselhauf 2009).

Section III covers the topic of variation and change in the context of chunking. More broadly, this falls within the domain of discourse analysis — another area that Sinclair worked in and contributed to in such an outstanding manner. The section is remarkably homogeneous, in that the contributions focus on different, yet interrelated topics, allowing the reader to grasp and appreciate the very potential and worthwhileness of the approach. The chapter by Mair, "Writing the history of spoken standard English in the twentieth century", focuses on spoken language versus written language, and discusses how the two can develop and change independently of one another over a certain period of time. This is demonstrated by means of a comparison of various linguistic phenomena (e.g. specificational cleft and modal verbs) in spoken and written corpora of British and American varieties. It is argued that changes in written and spoken discourse may: (1) occur in parallel (admittedly, quite rarely), (2) happen exclusively in one domain, (3) first happen in one domain, and, with a delay, in the other. As such, the chapter offers a refreshingly novel perspective on the real-time diachronic changes in written and spoken language.

In a similar vein, the following chapter by Mittmann, "Prefabs in spoken English", uses British (the spoken part of the *British National Corpus*) and American (the *Longman Spoken American Corpus*) corpora with the aim of studying phraseological differences between the two varieties (e.g. synonymous word combinations: *I reckon* vs. *I guess; have a look* vs. *take a look*). Mittmann argues that if speakers of different varieties (e.g. British and American) use different word combinations to express the same concept or idea, this can be taken as an

indication that the expression in question is formulaic and retrieved from memory. Clearly, more empirical evidence is needed to support this claim (and a wider spectrum of examples should be provided). Mittmann's chapter makes two useful suggestions. First, it is important not to disregard those word combinations, which on the surface may seem uninteresting and fully analysable, but are nevertheless formulaic (e.g. *Yes, it is*). Second, future research should focus on phonological properties of chunks (with respect to intonation, pauses, hesitations, stress, etc.) (also see Lin 2012).

In line with the general theme of this section, the chapter by Römer, "Observations on the phraseology of academic writing: Local patterns — local meanings?", compares the "phraseological profile" (p. 211) specific to academic writing with a less specialised, more general language corpus. The chapter focuses on the selection of common meaningful units in a corpus of linguistic book reviews, in particular, on the study of their semantic prosody. It is noted that phrases that seem neutral in isolation, may, in fact, be more likely to introduce positive or negative evaluations (e.g. *at the same time, on the other hand* can be followed by positive or negative information). Römer's findings imply that some of the patterns, as well as the distribution of their evaluative nature (more positive or more negative), are, indeed, specific to a particular kind of restricted language. Römer argues that such patterns serve as evidence for a "local lexical grammar" (also see Mahlberg 2006). Similar to Nesselhauf (this volume), however, the small sample size is an issue.

In the final chapter of this section, "Collocational behaviour of different types of text", Uhrig and Götz-Votteler further explore the quality and quantity of collocations found in different text types: fiction, newspaper and academic articles, EFL essays, and an automatic on-line translation of a novel. Uhrig and Götz-Votteler tested a number of hypotheses. First, there is an interrelation between the collocation behaviour of a text and its difficulty. This was not found to be the case. Second, there is an interrelation between the collocational behaviour of a text and the text type. Again, this was not found to be true. Third, there is an interrelation between the collocational behaviour of a text and its idiomaticity. The results suggest that this was indeed the case, with texts produced by English language learners being collocationally weaker that those written by native speakers. Arguably, this finding was somewhat expected given a large body of existing research. Although the results are far from being conclusive (or convincing, for that matter), the chapter nevertheless nicely illustrates how computer software can be used to explore and "play" with your data. The graphs used in the chapter are pleasantly clear, practical and informative.

The final section of the volume deals with yet another area strongly associated with Sinclair's name — corpus-assisted lexicography. More precisely, Section IV

centres on computational aspects of chunk parsing and extraction from corpora. In "Corpus linguistics, generative grammar and database semantics", Hausser presents the theoretical framework of Database Semantics (DBS) — a natural language processing system. In basic terms, the ultimate goal of DBS is to enable an artificial cognitive agent (a computer) to comprehend natural language (a human). Hausser discusses a number of components that constitute the DBS architecture, and, briefly, how the system deals with collocations and other types of chunks (it is pointed out that in DBS, collocations are in the language, not in the grammar). Although the chapter makes it clear what the potential of DBS is, the reader would have benefited more had the core part of the chapter (rather than a peripheral one) been dedicated to the treatment of collocation in the context of DBS.

In "Chunk parsing in corpora", Görz and Schellenberger discuss chunk parsing in Natural Language Processing (NLP). First, the concept of chunking in the context of NLP is defined and explained, and limitations and problems of chunking in English and German are discussed. Following this, Görz and Schellenberger introduce and evaluate a method called Transformation-based Learning. Overall, a reader with little or no parsing experience will find the chapter difficult to follow. The lack of a proper conclusion (particularly in light of the content being highly technical) is rather disappointing.

As the title suggests, the final chapter of the volume "German noun+verb collocations in the sentence context: morphosyntactic properties contributing to idiomaticity" by Heid, discusses the extraction of German collocations with the aim of identifying linguistic properties of multi-word expressions that contribute to their idiomaticity (e.g. modification, negation, word order). The second half of the paper deals with the procedures used to extract collocations. Heid suggests using a number of extraction criteria, arguing that none of them performs satisfactorily when used in isolation (e.g. one of the challenges is the ability to ignore phraseologically trivial, but highly frequent, word combinations). To this aim, Heid presents a parsing-based architecture for the extraction of unambiguous noun+verb collocations from a German corpus. Of the three chapters on computational aspects of chunk parsing and extraction, the chapter by Heid seems to be more focused and better organised than those by Hausser and Görz & Schellenberger; critically, it is also more accessible to a non-specialist reader. On the downside, the chapter clearly has an issue with terminology — too many unconventional terms are used, which is likely to lead to some confusion (e.g. the term 'multiwords' is not really used in literature, instead 'multiword expressions' should be used; it is not clear how 'idiomatized multi-word expressions', 'collocational multiwords', and 'idiomatized collocations' differ from each other and from 'multiwords', and why one would want to use so many different, but seemingly synonymous, terms in one paper).

One of the general criticisms of the volume is that one cannot help noticing that certain information is repeated throughout (e.g. the definition of the idiom principle and open choice, the notion of creativity). However, this is perhaps unavoidable given the nature of the volume and the authors' common focus on the phraseological view of language in light of Sinclair's work. A bigger criticism, however, is that no study reported in the volume makes use of an additional method, adopting instead a purely corpus-based approach (researchers have recently argued for a complementary use of corpora and psycholinguistic techniques, e.g. Gilquin & Gries 2009). A more psycholinguistic approach to the problem, motivated by a wealth of corpus evidence, is bound to enrich and deepen our understanding of the very concept of the idiom principle, and, crucially, shed important light on the many issues raised in the collection, such as collocation processing, storage and retrieval. As is clear from the volume, many applied and corpus linguists entertain the idea of collocations and other types of multi-word expressions being processed as *unanalysed* wholes. While it has been demonstrated in a few recent studies that multi-word expressions are indeed processed quantitatively faster than novel phrases (due to their frequency and predictability; e.g. Arnon & Snider 2010; Janssen & Barber 2012, Siyanova-Chanturia et al. 2011), it is important to bear in mind that these studies do not support (and neither were designed to test) the idea of such chunks being processed as *unanalysed* units.

Overall, however, the collection offers a valuable insight into the legacy of John McHardy Sinclair and his enormous contribution to the fields of lexicography, language pedagogy and discourse analysis. Given a wide range of topics covered, the volume will prove to be a useful reference book for anyone interested in the phenomenon of prefabricated language in the context of corpus linguistics.

References

Arnon, I. & Snider, N. 2010. "More than words: Frequency effects for multi-word phrases". *Journal of Memory and Language*, 62, 67–82.

Gilquin, G. & Gries, S. Th. 2009. "Corpora and experimental methods: A state-of-the-art review". *Corpus Linguistics and Linguistic Theory*, 5 (1), 1–26.

Janssen, N. & Barber, H. 2012. "Phrase frequency effects in language production". *PLoS ONE,* 7 (3), e33202.

Lin, P. M. S. 2012. "Sound evidence: The missing piece of the jigsaw in formulaic language research". *Applied Linguistics,* 33 (3), 342–347.

Mahlberg, M. 2006. "But it will take time… points of view on a lexical grammar of English". In A. Renouf & A. Kehoe (Eds.), *The Changing Face of Corpus Linguistics.* Amsterdam/Atlanta, GA: Rodopi, 377–390.

Nesselhauf, N. 2009. "Co-selection phenomena in New Englishes: Parallels (and differences) to foreign learner varieties". *English World Wide*, 30 (1), 1–26.

Sinclair, J. McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. McH. 1996. "The search for units of meaning". *Textus: English Studies in Italy*, 9, 75–106.

Siyanova-Chanturia, A., Conklin, K. & van Heuven, W. 2011. "Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multi-word sequences". *Journal of Experimental Psychology: Language, Memory, and Cognition*, 37 (3), 776–784.

Reviewed by Anna Siyanova-Chanturia,
Victoria University of Wellington, New Zealand