Reviews

Jarle Ebeling and Signe Oksefjell Ebeling. 2013. *Patterns in Contrast*. Amsterdam/Philadelphia: John Benjamins. Studies in Corpus Linguistics 58. 249 pages.

The monograph *Patterns in Contrast* is a recent publication in the series *Studies in Corpus Linguistics*. The patterns that are contrasted in this work emerge through quantitative and qualitative analysis—illustrated through five case studies—and involve the two languages English and Norwegian. Following the tradition pioneered by scholars such as Stig Johansson, Ebeling and Oksefjell Ebeling adopt a bidirectional contrastive method using the *English-Norwegian Parallel Corpus* (ENPC) and fully monolingual corpora of both languages.

The authors take a novel cross-linguistic approach in that they study multi-word units rather than single words; this is introduced as 'contrastive phraseology' in Chapter 1. The aim of the analysis can be said to be to (a) identify meaningful multi-word units in English and Norwegian and then (b) examine the extent to which they correspond cross-linguistically, not only with respect to linguistic form, but also with respect to semantic and pragmatic connotations. The authors provide an in-depth study of multi-word units by considering four Sinclairian aspects which are by now familiar corpus-linguistic concepts: collocation, colligation, semantic preference and semantic prosody. These four aspects form part of Sinclair's model of 'extended units of meaning', which follows the neo-Firthian tradition of seeing meaning as residing to a large extent in multi-word units rather than in single words. An extended unit of meaning consists of a 'core', as in Sinclair's own example of *naked eye*, which has other variable items co-occurring with it, such as *with the naked eye* (collocation); or preposition + determiner + *naked eye* (colligation); expressions such as *see* and *perceive* having to do with visibility (semantic preference); and co-text that indicates a perceived difficulty of seeing something (semantic prosody; this example is reviewed on p. 57ff). Thus, the phenomenon under study is an extended unit of meaning, or a 'pattern'; these terms appear to be used interchangeably. A pattern is defined as "a recurrent sequence of orthographic words that function as a semantic unit" (p. 50). This definition is said to be different from that of Hunston and Francis' (2000)

Pattern Grammar in that the focus is on lexical rather than grammatical patterns. At the same time, the authors emphasise a point often made in the neo-Firthian tradition: that lexis and grammar are inseparable. It is also typically the case that a pattern is synonymous with a simplex word; *big deal*, for example, is said to be equivalent to 'important' (p. 209-10).

The work is firmly grounded in contrastive linguistics, which is defined as "cross-linguistic studies involving a systematic comparison of two or more languages with a view to describing their similarities and differences" (Hasselgård 2010: 98). Chapters 2 and 3 give a useful overview of this area, especially as developed in the context of parallel corpus analysis. As the study aims to locate pattern X not only in language A but also in language B (and then compare them), it is a basic requirement to establish a basis for comparison—a 'tertium comparationis', that is, "some kind of constant serving as the background of sameness against which the differences are to be measured" (Ringbom 1994: 738). The authors use Chesterman's (1998) notion of 'perceived similarities' as the basis for their contrastive analysis and the evidence is in the cross-linguistic correspondences found in the empirical material, which is bidirectional and involves translations. Observable items that correlate in source and target texts are talked about in terms of 'correspondence'. The degree of correspondence can be measured by overall frequency (how often does pattern X in language A correspond to pattern Y in language B?) and by taking into consideration aspects of form, meaning and use. As the analysis is partly based on translation data, the authors thoroughly cover the pros and cons of drawing on translation in contrastive analysis.

The work is also firmly grounded in the area of phraseology, for which the authors discuss relevant key notions in Construction Grammar, Pattern Grammar and in John Sinclair's and Michael Stubbs' ideas concerning the idiom principle and extended units of meaning. Chapter 4 includes a very interesting discussion about criteria for identifying semantic units in texts and how to operationalise the notion of semantic unity. Through a rich intertextual weave, the authors show that there is currently no way of fully automatising this procedure: "Even if statistical methods have been used with some success to identify semantic unity in patterns, there is no guarantee that sequences identified in this way in fact constitute a semantic unity" (p. 63). The most reliable criteria to date are said to be using paraphrases ("including one-word near synonyms")

and intertext ("e.g. in the form of recurrent patterns from several different texts").

The methodology used for the case studies is given a great deal of space. Anybody dedicated to the scientific principle of replicability will be pleased about Chapter 5, in which the different steps involved in the analysis are meticulously described. One of the key steps is n-gram extraction, on the basis of which patterns are then selected for cross-linguistic analysis. The main justification for retrieving n-grams is that this procedure avoids taking a pre-defined lexical element as the starting point of the analysis.

Chapter 6 describes the corpus material. In addition to using the fiction part of the ENPC, which is both a comparable corpus and a translation corpus, the authors use an expanded version of the corpus including a larger sample of fiction texts. In addition to the parallel corpus material, monolingual corpus material is also drawn on (whenever relevant; p. 83). For English, this entails the fiction part of the *British National Corpus*, amounting to 16 million words, and for Norwegian, the fiction part of the *Leksikografisk bokmålskorpus*, amounting to just under 14 million words.

The case studies comprise five patterns, with each pattern given a separate chapter: *big deal* (Chapter 8), *out of the ordinary* (Chapter 9), *found* + reflexive pronoun (Chapter 10), Norwegian *FÅ tak i* ('get hold of'; Chapter 11) and antonymic binomials of continuous lateral movement exemplified through expressions such as *back and forth* (Chapter 12). The analysis reveals a cline with regard to similarity/contrast of both form and meaning of the patterns. Occasionally, 'phraseological gaps' are found, as in the case of *big deal*, where no congruent (with respect to part of speech) corresponding pattern is found in Norwegian. As may be expected, cases of polysemy are also found in the data: Norwegian *FÅ tak i*, for example, has at least four distinct senses, judging from the correspondences in English: "obtain, grasp physically, contact, grasp mentally" (p. 170).

The concluding chapter sums up the five case studies. It also includes a section discussing challenges related to semantic prosody, especially as seen from a cross-linguistic perspective. Furthermore, the chapter presents new material in the form of a pilot study which applies the method to other languages, comparing English with German and Portuguese. The only pattern sufficiently frequent for examination is

found + reflexive pronoun, the analysis of which points to both "similarities and differences between the languages under study" (p. 221).

The degree of complexity of the work as a whole is relatively high and the reader needs to navigate between different areas of study— primarily contrastive linguistics, phraseology and corpus linguistics— and keep track of a large number of specialist terms. The authors themselves refer to the "terminological apparatus" (p. 209) applied in their work. Contrary to what may be expected, this rarely poses a problem, much thanks to the clarity of presentation and the care taken to provide not only definitions but also background information. The literature reviews on both contrastive linguistics and phraseology provide a great deal of food for thought.

The reading is made especially pleasurable by an engaging argumentative style and by discussions where both pros and cons are regularly brought up. A case in point is Section 3.3.1 on the use of parallel corpora, which records multiple voices from the literature in a balanced way, while still arriving at a clearly stated position in favour of using such data. The authors' ambition to not brush aside potential problems is evidenced throughout. One problem that is repeatedly referred to is sparse data: most of the patterns selected for analysis are relatively infrequent. *Out of the ordinary*, for example, occurs "17 times in the English original texts and four times in the translated texts" (p. 113). The corpus material, despite the added fiction texts, is not sufficiently large to yield a very large number of examples of the pattern in question. This presents a problem for this type of analysis, as rather sizeable corpora are needed in order for patterns to emerge in a reliable way.

As mentioned above, the method of extracting potential pattern candidates is done by bootstrapping based on n-grams which involves no pre-selected lexical material. While the identification of potential patterns follows strict criteria, the ensuing step of selecting patterns for further analysis is more subjective in character, which is also clearly stated: "intuition is part of the process of selecting which patterns to explore cross-linguistically" (p. 66). The authors still offer a list of six criteria for an n-gram to qualify as a pattern, thus making the process less arbitrary. However, the procedure for selecting the five patterns illustrated in the case studies from the rather unwieldy lists of over

12,000 n-grams is not very clear, although the authors state that "since we are doing cross-linguistic analysis of patterns, we concentrated on combinations that exhibited potentially interesting differences between the source and target texts or between the two languages" (p. 95).

Given the solid work that has gone into devising the method of the study, it is not easy to find ways in which it could be improved. There is, however, one step which could have been added to the retrieval stage: a dispersion check. Currently, the method considers frequency as a criterion, with a cut-off point at eight occurrences in order for a given item to be considered for analysis. However, it does not take into consideration the spread of a given item in the corpus, as is done for example in Biberian extraction of multi-word units (so-called "lexical bundles", see e.g. Biber 2010). With an added dispersion criterion, stating that a given item needs to occur in at least X number of texts in the corpus, items such as *the men on the hills* (Table 5.1, p. 68) would likely automatically be excluded from the lists.

My understanding of the role of the five case studies is that they are meant to illustrate the method ("demonstrate the potential of exploring patterns in contrast" as stated on the back cover). However, the case studies still leave me with more general questions: What do we do with all the detailed and intricate facts about the five patterns? Where does all this information fit into linguistic theory? Is it feasible to carry out this type of in-depth analysis for a large number of patterns? It is easy to see how the level of detail may be useful to translators specialising in the selected language pair or to advanced language learners whose first and second/foreign languages involve English and Norwegian, but it is more difficult to see beyond these applied target audiences. (At this point, it needs to be said that knowledge of Norwegian is not necessary in order for the reader to appreciate the book, although, needless to say, a reader who is familiar with Norwegian will be able to appreciate the intricacies of the case studies.) A related question concerns what predictions linguistic theory might make with respect to how the idiom principle, as opposed to the open-choice principle, works cross-linguistically. Perhaps it is unfair to be requesting discussion about this, as the notion that the word is the basic linguistic unit very much predominates in traditional linguistic theory building, but such a discussion would nevertheless have been a welcome addition.

The main contribution of the work is in the presentation of a methodology for identifying and contrasting patterns cross-linguistically. This should be seen as a starting point; the work does not end here. The idea is that, by applying the methodology systematically, we will be in a better position to answer more general questions about how patterns work cross-linguistically. In the case of the pattern *found* + reflexive pronoun, for example, the authors suggest that systematic cross-linguistic findings will help us understand how reflexive patterns in general work across languages. What is really exciting about this is that the workings of meaningful units above the word level are explored not only from the perspective of linguistic form, but also from the perspective of semantic and pragmatic functions.

Annelie Ädel
Dalarna University, Sweden

References
Biber, Douglas. 2010. "Corpus-based and corpus-driven analyses of language variation and use." *The Oxford Handbook of Linguistic Analysis*. Eds. B. Heine and H. Narrog. Oxford: Oxford University Press. 159-91.

Chesterman, Andrew. 1998. *Contrastive Functional Analysis*. Amsterdam: John Benjamins.

Hasselgård, Hilde. 2010. "Contrastive analysis/contrastive linguistics." *The Routledge Linguistics Encyclopedia*. Ed. K. Malmkjaer. London: Routledge. 98-101.

Hunston, Susan and Francis, Gill. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Ringbom, Håkan. 1994. "Contrastive analysis". *Encyclopedia of Language and Linguistics*. Eds. R.E. Asher and J.M.Y. Simpson. Oxford: Pergamon. 737-42.