

**Joanna Kopaczyk** and **Jukka Tyrkkö** (eds.). *Applications of pattern-driven methods in corpus linguistics* (Studies in Corpus Linguistics 82). Amsterdam/Philadelphia: John Benjamins. 2018. 313 pp. ISBN 978 90 272 00136. Reviewed by **Christer Geisler**, Uppsala University.

This edited volume consists of eleven chapters and is divided into three parts: I. Methodological explorations, II. Patterns in utilitarian texts, and III. Patterns in online texts. Apart from the introductory chapter by the editors, **Jukka Tyrkkö** and **Joanna Kopaczyk**, the book contains an additional ten articles. The book is a welcome and important contribution to the study of pattern-driven linguistic research, which here refers primarily to investigations of n-grams and lexical bundles. The editors place pattern-driven methods between corpus-based and corpus-driven approaches, with the term ‘pattern-driven’ implying a study of language patterns that are not linguistically pre-determined, but without the theory-neutral aspects of corpus-driven research. That said, most of the authors state that they use corpus-driven methods, since the theoretical framework is determined by the findings.

In the first part of the volume, **Gerold Schneider** and **Gintare Grigonyte** analyze lexical bundles in learner and native language corpora. The study investigates formulaicity versus expressiveness, and introduces the concept of surprisal, an information-theoretic measure. The results of their study indicate that L2 speakers overuse most frequent lexical bundles, but underuse infrequent ones.

**Lucasz Grabowski** addresses methodological issues in the analysis of lexical bundles in a corpus of pharmaceutical drug-drug interactions, i.e. texts that describe cases where one drug affects the action of another. The author investigates a small set of lexical bundles and finds that the sampling method used can have important effects on the representativeness of the results.

A diachronic approach is taken by **Ondřej Tichý**, who investigates lexical obsolescence between 1700 and 2000 in data from Google Books. The study treats methodological problems in diachronic corpus-driven research and shows that n-grams can be used to track lexical loss over time.

The second part of the volume focuses on investigations of utilitarian texts. In a paper by **Antonio Pinna** and **David Brett**, part-of-speech grams (POS-grams) are studied in newspaper language. The authors use POS-grams of six members, exemplified by *with the palm of your hand*. The paper argues that new insights into general schema of specialized registers can be uncovered by analyzing POS-grams. These results would probably not have been found by using lexical n-grams.

**Stanisław Goźdz-Roszkowski** studies semantic sequences in judicial opinions of the US Supreme Court. In particular, the paper focuses on noun complement clauses (such as *assumption that* and *argument that*) and their functions in legal discourse.

**Anu Lehto** analyzes three-word lexical bundles in Early Modern and Present-day English Acts of Parliament. The investigation shows that, in the historical acts, anaphoric bundles are the most common type, while the most common bundles in the present-day data are bundles that limit the extent of the law in question. In both time periods, the predominant bundles are made up of noun phrase and prepositional fragments. The lexical bundles reflect diachronic change originating from a sociohistorical and discourse-related context.

In the final part of the volume, the language of online texts is analyzed. Lexical bundles in Wikipedia articles are investigated in **Turo Hiltunen's** article, and the author compares the results with other types of expert writing. The study shows that Wikipedia articles are markedly different from other forms of academic text. Overall, Hiltunen's article provides evidence that Wikipedia articles have a different communicative purpose than academic research articles.

**Joe McVeigh** investigates lexical bundles in email marketing texts to legal professionals in the US. One finding is that many lexical bundles are highly specific to that particular genre. In fact, the lexical bundles in these marketing texts are unique to that register. As in several articles in this volume, McVeigh discusses methodological questions related to lexical bundles, such as frequency and range.

**Federica Barbieri** focuses on lexical bundles in American blogs and finds that blogs contain mainly verb-based bundles, as well as referential and narrative expressions. The most common functional category of bundles includes stance expressions with first person pronoun *I*.

In the final article of the volume, **Joanna Kopaczyk** and **Jukka Tyrkkö** study 3-grams (n-grams of three words) in blogs in the GloWbE corpus. Based on the analysis of 3-grams, four geographical clusters emerge: an Inner Circle cluster, an African cluster, a South-East Asian cluster, and an Indian Peninsula

cluster. The authors also discuss their findings in relation to various models of English worldwide.

To sum up, *Applications of pattern-driven methods in corpus linguistics* provides important contributions to the study of n-grams and lexical bundles, applied to various types of corpora, ranging from learner data to Wikipedia articles. Several of the contributors address methodological concerns and the benefits of the corpus-driven approach. The articles of the volume are of high quality and they can serve as useful models for corpus linguists in the future.