Sebastian Hoffmann, **Andrea Sand**, **Sabine Arndt-Lappe** and **Lisa Marie Dillmann** (eds.). *Corpora and lexis* (Language and Computers). Leiden/Boston: Brill Rodopi. 2018. 306 pp. ISBN 978-90-04-36113-3. Reviewed by **Sven Leuckert**, Technische Universität Dresden.

The edited collection *Corpora and lexis* was published in the 'Language and Computers' series by Brill Rodopi and represents the result of a number of papers presented at ICAME36 in Trier, which had the theme "Words, Words, Words – Corpora and Lexis". A topical introduction and nine chapters are rounded off by an index of concepts and an index of corpora and databases.

In their introduction, the editors **Hoffmann**, **Sand**, **Arndt-Lappe**, and **Dillmann** explain that lexis and lexicon can be studied from a paradigmatic perspective and a syntagmatic perspective: Words "form a highly connective network" in the lexicon (p. 1), but we also ask "how co-occurrence of words actually shapes meaning" (p. 1). While there are studies venturing into either one of these approaches, many studies now combine the two and consider contextual knowledge. According to the editors, the contributions in the edited collection provide up-to-date examples of corpus resources and software, findings about frequency, applications of corpus findings, and new theories and concepts in lexis and semantics.

The first chapter, "Modelling Lexical Structures in the Oxford English Dictionary" by **Weiner**, introduces and discusses the current structure of the Oxford English Dictionary (OED). Crucially, the author highlights that there are "numerous structural information networks implicit in the Dictionary, both within the individual entry and (especially) crossing the boundaries from one entry to another" (p. 7). After a comparison of the CD-ROM version of the OED and the online version, Weiner presents the current structure of the OED and key aspects of its organisation: the headword lemma, variants of entries, pronunciation of words, etymology, featured quotations, and sense sections. While the dictionary is in a good stage already, the author feels that the many existing interconnections should lead the way in further developing the dictionary; this approach would reflect how words are structured in the mind of the reader.

The second contribution to the collection, "Investigating the Circumstances of Coinage" by **Renouf**, considers the signalling of coinages in a 1.5 billion word corpus, consisting of news texts from the *Independent* and the *Guardian* from 1984 until 2014. Coinage, in this study, is understood as the creation of a new word featuring an already existing word and one or more affixes. Based on the fact that writers "sometimes use a lexical signal to characterise a word in text as being new or possibly new coinage" (p. 44), Renouf differentiates no signalling, tentative signalling, performative signalling, and past signalling. Two very interesting kinds are tentative and performative signalling: in tentative signalling, the existence of a coinage is put into question; that is, its addition to the lexicon is questioned. Performative signalling, in turn, announces a coinage in the text. The typology of signalling outlined in the study may be used to investigate how coinages come into being and are contextualised in news writing.

In their contribution "Synonym Selection as a Strategy of Stress Clash Avoidance", **Schlüter** and **Knappe** analyse the effect of the adjacency of two stressed syllables on replacing a word with a (near) synonym. Since speakers generally avoid having two strongly stressed syllables in a sequence, the choice of a word may be informed by the syllabic structure of adjacent words. A selection of five out of the total of 20 decades featured in the Corpus of Historical American English (COHA) serves as the database for the authors' study; the chosen synonym pairs and triplets are *rich* and *wealthy*, *glad* and *happy*, *fast*, *quick*, and *rapid*, and, finally, *shut* and *closed*. These adjectives are chosen because they are rhythmically different, semantically similar enough to each other, and easily retrievable from the corpus. An important outcome of the study is that "monosyllabic or end-stressed adjectives occur less often before stressed syllables than their non-end-stressed (near) synonyms" (p. 96). However, certain caveats exist; for instance, the degree of synonymy between the items is different.

In the fourth paper, **Aijmer** studies "Intensification with *Very*, *Really* and *So* in Selected Varieties of English". Following the belief that regional variation can be understood as a kind of social variation, the author investigates usage patterns of the intensifiers *very*, *really*, and *so* in spoken American English (AmE), British English (BrE), New Zealand English (NZE), and Singapore English (SinE). Using the International Corpus of English (ICE) for BrE, NZE, and SinE and the Santa Barbara Corpus for AmE as her data, Aijmer finds that the overall frequencies of the three intensifiers are highest in SinE and lowest in AmE. The intensifiers overwhelmingly occur together with evaluative and emotional adjectives; *very* and *really* tend to co-occur mostly with *good* and *nice*, while *so* is used mostly in conjunction with *funny*. The reasons for the different

frequencies are not entirely clear from the dataset, but Aijmer suggests that the high frequencies of the intensifiers in SinE can be attributed to the selection of them as variants over other possible intensifiers in the process of SinE becoming a stable variety with its unique set of features. The low frequencies in AmE, on the other hand, may be attributed to the preference of other intensifiers.

Continuing the discourse-pragmatic focus, **Kirk**'s contribution "The Pragmatics of *Well* as a Discourse Marker in Broadcast Discussions" zooms in on the discourse marker *well* in the broadcast discussions featured in the International Corpus of English for Ireland (SPICE-Ireland) and Great Britain (ICE-GB). Kirk follows Aijmer (2013) in differentiating the expression of coherence, involvement, and politeness as the central functions of *well*. Despite different frequencies in the two corpora, establishing coherence is the dominant function of *well* in both SPICE-Ireland and ICE-GB, whereas expressing politeness is the rarest function at less than ten per cent each. This distribution is partially attributed to connectivity as "*well*'s basic function" (p. 164). From a methodological perspective, Kirk identifies the overlap of involvement and politeness as a problem, which can be addressed by considering them as one larger category. Overall, "[*w*]*ell* thus maintains the flow of discourse as well as harmonious relationships among speakers, each of which is crucial for an interactive broadcast situation" (p. 168).

In the chapter "Between Lexis and Discourse: A Cross-register Study of Connectors of Contrast", **Dupont** provides an analysis of adverbial connectors (referred to as conjuncts in Quirk et al. 1985 and as linking adverbials in Biber et al. 1999). The author embeds the analysis of these connectors in the framework of systemic functional linguistics, which has been made popular in English linguistics predominantly by Halliday (see Halliday and Matthiessen 2004). In particular, Dupont suggests that in the Theme-Rheme distinction, the Rheme has been treated simplistically; rather than treating it as one entity, a more fine-grained approach is necessary. By means of a corpus of 5.5 million words containing newspaper editorials, research articles in the humanities, and minutes of debates in the European Parliament, the frequency and placement of *however*, *instead*, *nevertheless*, *though*, *on the other* (*hand*), *still*, and *in contrast* in English produced by native and expert users is analysed. Connectors of contrast are used most frequently in academic texts and least frequently in debates; both lexis and register have an impact on connector placement.

The next chapter by **Almela** and **Cantos**, entitled "Towards a Model of Co-collocation Analysis: Theory, Methodology, and Preliminary Results", takes into focus inter-collocational dependency, also referred to as 'mediated collocation' by the authors. Based on data from the large corpus enTenTen2013 (ca. 20

billion words), the authors propose a more complex analysis of collocation by investigating mediated collocation. The corpus has been selected because it features many collocates not present in smaller corpora such as the British National Corpus (BNC). Studying mediated collocation means analysing co-collocates outside of the collocational pair, for instance, *mitigate* and *minimise* as co-collocates of *consequence* and *unintended*. Thus, moving away from simpler analyses of a node and its collocate to more detailed analyses of a node, its collocate(s), but also potential co-collocates, promises intriguing insights into lexical association.

Shifting the focus to pedagogy, "The Lexicogrammar of BE *Interested*: Description and Pedagogy" by **Gabrielatos** studies lexicogrammatical patterns of BE *interested* in spoken and written L1 (represented by data from the BNC) and L2 (represented by the ICLE and LINDSEI). Patterns of relevance are BE *interested* with a Noun Phrase, an -*ing* participle clause, a noun (*wh*-) clause, a *to*-infinitive clause, and zero complementation. Furthermore, the frequency of other copular verbs with adjectival *interested* as well as modalised BE are investigated. In addition to the corpus analysis, the author also includes a close study of pedagogical materials in order to evaluate the coverage of BE *interested* in learning materials geared towards learners at the CEFR levels of B2 to C2. The L1 data features lexicogrammatical patterns not included in the pedagogical or reference material, while the L2 data generally correlates with the treatment of BE *interested* in the pedagogical material. However, all consulted sources are described as incomprehensive, which means for learners that studying a selection of grammars and dictionaries is recommendable to obtain a fuller picture of how collocations with BE *interested* work.

Keeping the focus on learner English, **Bestgen** and **Granger**'s chapter "Tracking L2 Writers' Phraseological Development Using Collgrams: Evidence from a Longitudinal EFL Corpus" investigates the phraseological unit 'collgrams'. Collgrams combine collocations and lexical bundles and represent a useful tool to investigate the idiomaticity of learner's writing. For their study, the authors employ a subsection from the Longitudinal Database of Learner English (LONGDALE), consisting of 178 essays (ca. 100,000 words) written by 89 French-speaking English undergraduates. The corpus features essays written by the students in their first year and their third year, which allows longitudinal comparisons. In order to assess idiomaticity, n-grams in the learner corpus are identified and then compared to their presence and frequency in the BNC. The bigram *vicious circle*, for instance, occurs in LONGDALE and is also frequent in the BNC; it therefore receives a high score and counts as idiomatic. The bigram *accept to* in LONGDALE, on the other hand, receives a negative score,

since its occurrence in the BNC is lower than chance would predict. The method is considered ideal for tracking the learning process, since it focuses on "degrees of appropriateness" (p. 293).

Overall, the book can be considered a great success. The breadth of corpora and software tools employed is impressive, and the chapters are meaningfully linked to areas as diverse as word-formation, lexicography, discourse pragmatics, pedagogy, and systemic functional linguistics. The first chapter by Weiner is a great addition to the more linguistically-inclined papers that constitute the rest of the volume, and its relevance reaches beyond a specialist audience. In addition, many of the contributions can be used in undergraduate, some in graduate classes, on topics such as word-formation, pragmatics, semantics, lexicology, and lexicography. However, the breadth of topics covered could also be criticized: the title of the book itself is fairly broad, and strong topical coherence is, apart from what is suggested by the title, not clearly evident by reading the contributions. This is not to take away from the excellence of the volume as a whole or the individual chapters, however: scholars interested in the study of lexis on the basis of corpus data will find plenty of innovation in the collection. Thus, the common aim of all nine contributions could be described as innovating a field with a longer tradition, the study of words and their meanings, enriched with the possibilities of, at times cutting-edge, technology, statistics, and methodological finesse.

References

Aijmer, Karin. 2013. *Understanding pragmatic markers*. Edinburgh: Edinburgh University Press.

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

Halliday, M.A.K. and Christian Matthiessen. 2004. *An introduction to Functional Grammar*. London: Routledge.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.