Jiyoung Yoon & Stefan Th. Gries (eds.). (2016) *Corpus-based Approaches to Construction Grammar*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Reviewed by Dirk Pijpops (Research Foundation Flanders (FWO), University of Leuven)

Corpus-based Approaches to Construction Grammar presents eight studies which aim to advance Construction Grammar theory by applying a range of corpus techniques. One of its keywords could be 'coherent variation'. The book offers variation in the languages and phenomena studied; there are studies on Dutch, English, Italian and Spanish, adult and child language, synchronic and diachronic data. In addition there is variation in the granularity of the methods used, which range from the use of simple corpus examples and frequencies over association measures to various more advanced multivariate techniques. Finally, language variation itself of course presents a recurrent theme. Meanwhile, the book manages to maintain coherence in several ways. First, all studies subscribe to the framework of Construction Grammar, and the questions posed and answers given by all studies contribute to theoretical progress within this framework. Second, there is coherence in the type of data source, with all studies drawing on corpora. Finally, there is coherence in coupling each research question to the appropriate choice of corpus-based method, which may very well constitute the hardest task of a corpus linguist. Striking this balance between variation and coherence, i.e. providing enough diversity to be inspiring to the reader and to be representative for the entire field of corpus-based approaches to Construction Grammar, while not losing all sense of cohesion, is hard, yet this book succeeds in it.

The editors Jiyoung Yoon and Stefan Th. Gries shortly introduce the studies and congratulate usage-based linguistics in general and Construction Grammar in particular on actuating the quantitative turn in a discipline long dominated by the introspective grammaticality judgments of generative grammar:

> Usage-based linguistics in general and Construction Grammar in particular are now brimming with experimental and observational studies (…) to the point that there are now publications discussion the quantitative turn in cognitive linguistics (Janda 2013). (Yoon & Gries 2016: 1)

The fastest-growing new approach in this quantitative turn, it is argued, involves the use of corpora. It is of this approach that the book intends to take stock, not to

introduce entirely novel techniques, but to provide examples of the wide range of possibilities that corpus research offers constructional linguists.

Maaike Beliën offers a fresh perspective on a long-standing discussion in Dutch linguistics about whether to analyze *in* in (1) as part of a postposition constituent with *de garage*, or as a particle of a separable complex verb *inrijden* ('to drive into'). Whereas most previous studies focused on syntactic arguments based on fabricated sentences, Beliën argues for the second view by means of a method to determine consitituency based on semantics, which she applies to usage instances from the Internet, and from the 38 million word corpus of the Institute for Dutch Lexicology. This method consists of three steps:

> (i) a semantic analysis of the construction under study (…); (ii) a semantic comparison of that construction with relevant other constructions, i.e. constructions whose constituency may be similar (…); and (iii) an account of semantic differences and similarities in terms of constituency.   (Beliën 2016: 13)

Apart from supplying a very welcome addition – perhaps even closure – to the Dutch constituency discussion, Beliën's contribution is also useful in providing a concrete example of how to determine conceptual constituency in a systematic way (see Langacker 1995; Langacker 1997). In doing so, she makes a strong case for the significance of semantics in issues of constituency. The method proposed by Beliën definitely deserves to be copied in other discussions on such issues. The absence of a diachronic perspective is perhaps a missed chance, yet one can only do so much in a single article. It also remains somewhat unclear how ambiguous instances such as (2) should be analyzed. However, neither of these minor issues detracts from the value of the contribution.

(1)   *Hij rijdt   de garage in.*
      He drives the garage in
      'He is driving into the garage.'

(2)   *Hij rijdt   de truck in.*
      He drives the truck in
      'He is driving into the truck.' or 'He runs in the truck.'

Valeria Quochi addresses light or support verbs, as in (3), a set of verbs in Italian that has thus far received sparse attention in Construction Grammar. She draws data from the CHILDES databases, and analyses transcripts of in total eleven Italian-speaking children from age 16 to age 40 months. Based on this analysis, it is argued that children acquire a pivot scheme for Light Verb Constructions (LVC), which they employ to name events or activities for which they have yet to acquire a full lexical verb. In addition, Quochi proposes that LVCs form a cluster of constructions tied together by family resemblance, with the Perform Intransitive

Action construction as in (3) at its center. This view is supported both by a qualitative analysis of usage instances and by a descriptive, quantitative analysis.[1]

(3) *Fare un ballo.*
     do   a   dance
     'to dance'

Victoria Vazquez Rozas and Viola G. Miglio study the use of the Experiencer Object and Experiencer Subject Construction with Spanish and Italian psych verbs. The source of the data are the ARTHUS corpus for Spanish, and the BADIP and C-ORAL databases, as well as excerpts from the *La Repubblica* newspaper for Italian. The goal is more than strictly comparative, however. The authors aim not simply to contrast both languages, but also intend to assess the influence of the text type and of properties of the participants such as animacy, person and syntactic class. To do so, they study the distribution of, e.g., animate stimuli in experiencer-subject constructions, relative to experiencer-object constructions. While this set-up is interesting and the sheer amount of data presented is admirable, there are two minor shortcomings that somewhat hinder a straightforward interpretation. First, it is unfortunate that no theoretically-informed predictions are formulated before the analysis, which could then be tested against the data. Instead, the results are mostly interpreted in a post-hoc fashion. Second, the language user does not appear to have a free choice of argument construction for the psych verbs under scrutiny. The choice rather seems to be strongly constrained by the verb (Vazquez Rozas and Miglio 2016: 82–84). This would mean that what is actually being measured is not a direct influence of, e.g., the animacy of the stimulus on the choice of argument construction. Rather, it is (i) the statistical tendencies of verbs to appear with, e.g., animate stimuli, and in turn, (ii) the likelihood that these verbs exhibit either an experiencer-subject or an experiencer-object argument realization. However, the intermediate role of the verb is not taken into account in the discussion of factors such as the animacy or syntactic class of the stimulus. This approach makes it hard to get a clear picture of the data, and is strikingly at odds with the two subsequent contributions, both of which specifically highlight the pivotal role of the verb in Spanish argument realization. Still, there is no doubt that the present article is a valuable addition to the literature on psych verbs.

   Johan Pedersen investigates the argument realization of Spanish verbs in constructions of telic motion, in order to answer questions regarding the typological status of Spanish. Data are drawn from the Corpus del Español. The quantitative inquiry is primarily based on a collostructional analysis which measures the

---

1. The graphs which provide the primary interpretative tool in the descriptive analyses are hard to interpret however, since the y-axes lack numbers.

association of a verb with the construction [SUBJ V a NP], which expresses telic motion. Pedersen convincingly shows that verbs that lexically express manner of motion also regularly appear in such environments in Spanish, which runs counter to the traditional view of Spanish as a verb-framed language. Instead of mappings of form and universal meaning components, as in the Talmian tradition (Talmy 1991, 2000), Pedersen argues that a fundamental typological parameter steering the argument realization of a language is the prevalence of schematic versus lexical constructions. In this regard, a language like English, where verbs combine rather freely with various schematic argument constructions, contrasts with Spanish, where argument encoding is much more stringently determined by the verb. The attention for intra-lingual variation in typological research is laudable, and conversely, the undertaking to answer typological questions using corpus data and a Construction Grammar framework offers a refreshing perspective. Ideally, an application of the same technique on data from a Germanic language would have constituted an interesting comparison, but this is understandably left for future research.

Jiyoung Yoon and Stefanie Wulff explore the preferences of Spanish verbs for either infinitival or sentential complement clauses. They extract data from the Spanish AnCora corpus and execute a Distinctive Collexeme Analysis on the verb slot, with the choice between infinitival or sentential complements as alternating constructional environments.[2] In addition, they discuss usage instances from the RAE database and the Internet. It was found that verbs of desire prefer infinitival complements, whereas verbs of mental activity and, to a lesser extent, verbs of communication, are more compatible with sentential complements. To explain the preference of mental activity verbs, it is hypothesized that these verbs are undergoing a grammaticalization process from full matrix clause to epistemic marker, as in (4). The question whether this truly represents language change in progress or a more stable situation, is left for future research, as is the question what motivates speakers, given a particular verb, to choose between either complementation option.

(4) *Es   muy importante, **creo,**   tener como contrapeso   de los Parlamentos*
     (it)is very important   I think have as   counterbalance of the parliament
     políticos una asamblea de culturas europeas.
     political a   meeting of cultures European
     'It is very important, I think, to have a meeting of European cultures as counterbalance of the political parliament.'

---

2. The tables that present the results of this analysis do not mention raw frequencies, which would have been helpful, but this is a minor remark.

Sarah Bernolet and Timothy Colleman present the results of both a corpus study and a priming experiment on the Dutch dative alternation. For each of 15 alternating verbs, they distinguish between a concrete and figurative sense, taken from the Van Dale Groot Woordenboek der Nederlandse Taal (den Boon & Geeraerts 2005). In the corpus study based on the ConDiv corpus, a Distinctive Collexeme Analysis measures the biases of these 30 senses for the Double Object (DO) or Prepositional Dative Construction (PD). For a number of verbs, large discrepancies in preference are found between both senses. It is shown that neither concrete nor figurative senses have a consistent preference for either construction. Instead, the relevant semantic distinction appears to be the (lack of) affectedness of the receiver. These results are taken to indicate that language users store usage frequencies at the sense-level. To further test this, and to investigate whether such storage occurs in addition to or instead of storage at lexeme-level, a priming experiment is conducted. Participants are shown drawings of transfer events and are asked to describe these. Before each picture, they are primed with DO or PD sentences containing the same verb, in either the same or the other sense. While the priming effects were found to be weak, the results still indicated that usage frequencies are stored at the sense-level, but are also accessible at lexeme-level. The contribution of Bernolet and Colleman is a prime example of how corpus research and psycholinguistic experiments can complement one another. Presenting data from various sources in the same article is admirable, and sadly still an exception. With their combind approach, Bernolet and Colleman show how corpus research cannot only be used to describe how language functions in a community of speakers, but can also provide insights in how language may operate in the minds of individual language users.

   Christopher Shank, Koen Plevoets and Julie Van Bogaert study the diachronic development of the *that/zero* alternation in English subordinate clauses with the matrix verbs *think, guess* and *understand*. Data are drawn from a large selection of English corpora and annotated for 11 variables proposed in the literature. All of these variables, as well as their two-way interactions with the additional variables *period, verb* and *mode* (spoken vs. written), are entered into a stepwise variable selection procedure to compose a regression model. Contrary to the expectations in the literature, it is found that only *guess* exhibits a diachronic increase in *zero* use, while an outspoken rise in *that* use is observed for *understand*, and a slight rise for *think*. Furthermore, their analysis allows them to precisely assess the influence of each variable in each period and for each verb. Shank, Plevoets and Van Bogaert provide an excellent overview of the extensive literature and a much-needed corpus-based description of what is going on with *that/zero*. They show raw counts and discuss the relevant effect plots from the regression model, which makes for a very lucid presentation of the results. As various factors are shown to obscure

the actual diachronic trends, their multivariate analysis uncovers the dangers of relying solely on bivariate tests, which still presents the more customary approach in historical research. In addition, they reveal the influence of several variables to be dependent on the matrix verb, which is something not typically considered in earlier research.

Natalia Levshina scrutinizes the semantic structure of the Dutch causative construction with the verb *laten* ('to let'), using a method similar to Behavioral Profiles (Divjak and Gries 2006; Gries 2006). 731 instances of causative *laten* are taken from face-to-face dialogues, newspapers and online discussion groups from several sources, and coded for 35 contextual variables. Based on these variables a vector is created for each instance, from which a distance matrix is calculated. Next, the dimensions of this distance matrix are reduced to two by applying Multi-Dimensional Scaling in order to allow the relative position of all 731 instances to be visualized in a single graph. The shape that is revealed is that of a doughnut. The doughnut-shape indicates that a conceptual centre is lacking, which means that there is no single central sense or prototype of causative *laten* ('let') to be discovered. This finding is taken as a warning against the a priori assumption of a prototypical semantic structure. Three clusters can be distinguished within the doughnut, which roughly correspond to (i) lexicalized expressions of mental caused events, e.g. *laten weten/horen/zien* ('let know/hear/see'), (ii) a lexically more diverse group of non-mental caused events, such as *laten vallen/liggen* ('let fall/lie'), and (iii) 'delegated' causation as in (5) and middle-voice events. These clusters with intermediate transitional zones constitute neither clearly delineated senses, nor a shapeless continuum. Levshina presents an approach that holds promise of yielding many more interesting results when applied to other constructions. It sparks enthusiasm to see how creative corpus methods can inform even intricate theoretical discussions on semantic structure that are often assumed impervious to empirical testing.

(5) *De makkelijkste manier van beleggen is om het iemand   anders te laten*
     the easiest   way   of investing is to it someone else   to let
     doen.
     do
     'The easiest way to invest is to have someone else do it.'

This new volume of the series *Constructional Approaches to Language* is not meant as a methodological handbook in corpus linguistics, nor does it introduce an entirely new theoretical approach to Construction Grammar. Rather, it furnishes the reader with an overview of the possibilities of corpus methods in a constructional approach to language. In doing so, it provides inspiration to linguists working in Construction Grammar on how to choose the appropriate corpus technique from

an ever-growing methodological toolbox. Meanwhile, the theoretical implications of the contributions show how we are steadily making scientific progress as a field. Corpus researchers are offered with examples of how to embed their analyses within the framework of Construction Grammar. At the end of their introduction, the editors remark that these are exciting times for a linguist. I wholeheartedly agree. It will be especially invigorating to discover what lies beyond the next corner in the path of usage-based linguistics. One possibility is the continued evolution of empirical methods into truly data-driven approaches, as in Gries, Newman and Shaoul (2011) or Levshina and Heylen (2014). Another is the complementary development of computer simulations. The use of these simulations seems to be on the rise in the so-called third wave of usage-based linguistics (Beuls and van Trijp 2016: 2). It may be possible that some years after the quantitative turn, the path of usage-based linguistics may take a new, simulational turn.

Acknowledgements

References

Beliën, M. (2016). A constructional perspective on conceptual consituency. Dutch postpositions or particles? In S. Th. Gries (Eds.), *Corpus-based Approaches to Construction Grammar* (pp. 11–37). Amsterdam: John Benjamins.

Beuls, K., & Remi van Trijp. (2016). Computational construction grammar and constructional change. *Belgian Journal of Linguistics*, 30, 1–13. doi: 10.1075/bjl.30.01beu

Boon, T. den, & Dirk Geeraerts (Eds.). (2005). *Van Dale Groot woordenboek van de Nederlandse taal*. 14th ed. Antwerpen/Utrecht: Van Dale Lexicography.

Divjak, D. and S. Th. Gries. (2006). Ways of Trying in Russian: Clustering Behavioral Profiles. *Corpus Linguistics and Linguistic Theory*, 2(1), 23–60. doi: 10.1515/CLLT.2006.002

Gries, S. Th. (2006). Corpus-based methods and cognitive semantics. The many senses of to run. In S. Th. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics. Corpus-based approaches to syntax and lexis*, 57–99. Berlin: Mouton de Gruyter. doi: 10.1515/9783110197709.57

Gries, S. Th., J. Newman and C. Shaoul. (2011). N-grams and the clustering of registers. *Empirical Language Research Journal* 5(11).

Janda, L. (Ed.). (2013). *Cognitive linguistics. the quantitative turn*. Berlin: Mouton de Gruyter. doi: 10.1515/9783110335255

Langacker, R. (1995). Conceptual grouping and constituency in cognitive grammar. In I. Lee (Ed.), *Linguistics in the Morning Calm 3* (pp. 149–172). Seoul: Hanshin.

Langacker, R. (1997). Constituency, dependency and conceptual grouping. *Cognitive Linguistics*, 8(1), 1–32. doi: 10.1515/cogl.1997.8.1.1

Levshina, N., & K. Heylen. (2014). A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In R. Boogaart, T. Colleman & G. Rutten (Eds.), *Extending the Scope of Construction Grammar* (pp. 17–46). Berlin: Mouton de Gruyter.

Talmy, L. (1991). Path to realization. A typology of event conflation. In L. Sutton, C. Johnson & R. Shields (Eds.), *Proceedings of the seventeenth annual Berkeley linguistics society* (pp. 480–519). Berkeley: Berkeley Linguistics Society.

Talmy, L. (2000). *Toward a cognitive semantics.* Cambridge, MA: MIT Press.

Vazquez Rozas, V., & V. G. Miglio. (2016). Constructions with subject vs. object experiencer in Spanish and Italian. A corpus-based approach. In J. Yoon & S. Th. Gries (Eds.), *Corpus-based Approaches to Construction Grammar* (pp. 65–101). Amsterdam: John Benjamins.

Yoon, J., & S. Th. Gries. (2016). Corpus-based approaches to Construction Grammar. Introduction. In J. Yoon & S. Th. Gries (Eds.), *Corpus-based approaches to Construction Grammar*, 1–7. Amsterdam: John Benjamins.