7

Corpus Linguistics and the Cognitive/Constructional Endeavor

Stefan Th. Gries

7.1 Data in Linguistics and Corpus Data in Construction Grammar

Data in linguistics can be classified along at least three different dimensions (based on Gries 2013a), each of which could, for simplicity's sake, be heuristically divided into different points/ranges:

(1) How natural does the speaker perceive his (experimental) **setting**?
   a. *most natural*, for example, speakers who know each other talk to each other in unprompted authentic dialog;
   b. *intermediately natural*, for example, a speaker describes pictures handed to him by an experimenter;
   c. *least natural*, for example, speaker lies in an fMRI unit undergoing a brain activity scan while having to press one of three buttons in responses to digitally presented black-and-white pictorial stimuli.

(2) What (linguistic) **stimulus** does/did the speaker act on?
   a. *most natural*, for example, speakers are presented with natural utterances and turns in authentic dialog;
   b. *intermediately natural*, for example, speakers are presented isolated words by an experimenter in an association task;
   c. *least natural*, for example, speakers are presented with isolated vowel phones.

(3) What (linguistic) **units/responses** does/did the subject produce?
   a. *most natural*, for example, subjects produce natural and unconstrained responses to questions;
   b. *intermediately natural*, for example, speakers respond with isolated words (e.g., to a definition);
   c. *least natural*, for example, speakers respond with a phone out of context.

The present chapter is concerned with corpus-linguistic approaches in Construction Grammar (CxG), that is, with approaches that tend towards the more/most natural part of each of these dimensions. The notion of a *corpus* can be considered a prototype category with the prototype being a collection of machine-readable files that contain text and/or transcribed speech that are supposed to be representative of a certain language, dialect, variety, etc. and were produced in a communicative setting. That means that at least the prototypical corpus scores *most natural* on each of the above three dimensions. Often, corpus files are stored in Unicode encodings (so that all sorts of different orthographies can be appropriately represented) and come with some form of markup (e.g., information about the source of the text) as well as annotation (e.g., linguistic information such as part-of-speech tagging, lemmatization, etc. added to the text, often in the form of XML annotation). However, there are many corpora that differ from the above prototype along one or more of the above dimensions and of course corpora also vary wildly in terms of their size, annotation, ease of access, and processability, etc. Accordingly, the prototypical corpus contains data that are a kind of good-news-bad-news situation. The 'good news' is that corpus data often have a very high degree of ecological validity precisely because the production data they contain are not tainted by any artificiality. But that is also the 'bad news': Data that are not tainted by any artificiality is just another expression for 'noisy and unbalanced', which is one major reason why, as we will see below, the analysis of corpus data in CxG has become more and more heavily statistical – simply to deal with the multifactorial, noisy, and redundant mess that corpus data often are.

Corpus data did not play a big role in CxG historically. It is probably fair to say that CxG is now a little more than thirty years old since the 'founding' publications are probably Lakoff (1987), Langacker (1987), and Fillmore et al. (1988), to be followed by Goldberg (1991, 1995) and Kay and Fillmore (1999). But while much of this earliest work was mostly theoretical in nature and did not rely much either on experimental or on observational corpus data, today that situation has drastically changed. To use language from usage-/exemplar-based linguistics: When I 'grew up academically' in the mid to late 1990s, learning about Cognitive Linguistics and CxG on the one hand and about corpus linguistics and Pattern Grammar (Hunston & Francis 2000) on the other, there were very few tokens of studies that, in some multi-dimensional exemplar space, would have scored highly both on the CxG and the corpus-linguistics dimensions; back in the 1990s I certainly did not form a productive category of 'corpus-based CxG'. But then in the early aughts that all changed and CxG – in particular usage-based/cognitive CxG – has evolved at what seems like a breathtaking pace into a field of study in which we have moved

- from works with virtually no corpus data (or that used corpora as a mere repository from which to pick fitting examples) to studies with systematic data retrieval and annotation processes often involving thousands of data points; and

- from works that presented isolated examples as evidence for what is *possible* to studies with complex quantitative methods that show what's *likely* and that involve, for instance, multi-factorial or multivariate statistical analyses, 'more traditional' machine-learning or fancier deep-learning or construction-induction methods, or network analyses.

Much of that move really only happened within the last fifteen or so years. In 2013, I published an overview article "Data in Construction Grammar" (in Hoffmann & Trousdale's *Oxford Handbook of Construction Grammar*), which had a mere five to six pages on corpus-based and/or computational (machine-learning) studies; this time around, even just sampling papers from leading journals publishing studies relevant to this overview (e.g., *Cognitive Linguistics*, *Constructions and Frames*, and *Corpus Linguistics and Linguistic Theory*) had to be restricted to a small number of recent years so as to avoid drowning in an unmanageable number of interesting and methodologically extremely diverse studies. The purpose of this chapter is to give an overview of the different applications of corpus-linguistic data and methods to linguistic phenomena from a CxG perspective. While the overview is unlikely to be truly representative of the field (along what dimensions anyway?), care was taken to represent studies that differ along a variety of essential parameters, including:

- the *language(s)* studied;
- the *kind of language(s)* studied: L1/native speaker data, L2/FL non-native speaker/learner data, indigenized-variety speaker data, . . .;
- the *resolution*: individual speakers vs. variation between individuals vs. (dialectal) speech communities, . . .;
- the *temporal kind of study*: synchronic vs. diachronic/longitudinal;
- the (ranges and kinds of) *corpora* used;
- the *use* to which corpora were put: a collection of examples vs. fine-grained (semi-manual) annotation vs. bottom-up/inductive processing vs. correlation with additional experimental results, . . .;
- the *question* the study is trying to answer and, related to that, the *'scientific goal'* of the study: description vs. hypothesis testing vs. exploration, . . .; and
- the *statistical methods* used for the analysis of the corpus data: none/qualitative only vs. frequencies/probabilities vs. association measures vs. multi-factorial (predictive) modeling vs. exploratory and/or machine-/deep-learning kinds of methods, . . .

The overview will be structured according to the latter two criteria because (i) the two criteria are of course often very much related to each other and (ii) for many researchers it will be interesting to see which kinds of CxG questions corpus-linguistic data, their (typically) qualitative annotation, and their statistical analysis can help address. Also, it is particularly in the interplay of the last two criteria that corpus-based CxG has maybe most developed. Put differently, while the field is of course still concerned with definitional matters, questions of learnability, abstraction, and/or representation (both mental and formal), corpus-linguistic approaches have been and are now also targeting specific subsets of questions that in turn naturally come with specific degrees of quantitative methods. I will therefore proceed by discussing

- raw/normalized frequency-based approaches;
- studies involving associations and their strengths between different constructions and/or their parts; and
- statistical modeling, machine-learning, and exploratory/inductive bottom approaches.

In each of these sections, I will try and highlight topical clusters, that is, areas/questions that appear to be targeted particularly frequently; Section 7.3 will conclude.

7.2 Corpus-Based Applications in CxG

7.2.1 Largely Qualitative Corpus Approaches

As mentioned above, the initial uses of corpus or corpus-like data in CxG papers were largely only presentative in nature and served to make some theoretical point(s) by means of authentic examples, but often without the kind of systematic feature annotation that is characteristic of much contemporary work. This pointing out a lack of multivariate annotation is not meant as a criticism, given the different goals of papers at the time, but what is perhaps a bit more critical is that some such literature often did not clarify whether examples provided were made up or attested (and, if they were attested, what the source was). For example, Fillmore et al. (1988: 519) discuss hundreds of example sentences but usually provide no information on them, let alone on their source. One time they state "we have come across incontrovertible cases of attested utterances of non-negative *let alone* sentences that seem perfectly natural and which there is no apparent justification to ignore as performance errors" and proceed to discuss their examples (71) and (72) by stipulating (admittedly likely) contexts in which they may have been uttered. Kay and Fillmore (1999) proceed in a similar way: We don't learn much about where examples are from etc., and the same is true of many other studies such as Smith (1994), Kemmer and Verhagen (1994), Dancygier and Sweetser (1997), Morgan (1997), Gutzmann and Henderson (2019), and many others, which were all introspection-based and, if they used the word *data*, typically used it to refer to introspective judgments and/or example sentences.

Crucially, this is not just some complaint from a quantitative corpus linguist who wants corpus examples for the sake of corpus examples; the point is that what seem like clear-cut judgments from native speakers on made-up or even attested examples can look very different once one looks at (larger) quantities of data – as Sinclair (1991: 100) said, "Language looks very different when you look at a lot of it at once." For example, it is likely that traditional linguists would consider a sentence such as *He [$_{VP}$ donated [$_{REC}$ her] [$_{PAT}$ transplant* *money]]* ungrammatical, since it is widely held that the verb *donate* cannot be used ditransitively (even though its meaning is so similar to that of the prototypical ditransitive verb *give*). However, Stefanowitsch (2006: 69) shows that even the British National Corpus – a great but by today's standards not particularly large corpus – already contains at least one example exactly like this (in a maybe atypical newspaper headline, admittedly), and Stefanowitsch (2007: 65) lists ten examples of *donate* used ditransitively from a variety of internet pages from .uk domains, all of which "do not conform to what we might think of as the default DONATE frame"; instead, they appear to instantiate a frame that Stefanowitsch describes as follows:

> A donor transfers some of his/her money to a recipient. The recipient is an official organization who uses the money to advance some public or charitable cause or to pay for its own expenses in doing so. The donor is an individual who gives the money because s/he believes in the cause, and without expecting to profit personally. There is no personal relationship between the donor and the recipient.

Thus, while linguistics in general and CxG in particular have benefited a lot even from papers that did not feature corpus data or analyses, linguists clearly have no unbiased and axiomatically correct view of what is possible (i.e., what can or cannot be said; see Labov 1975), let alone what is likely. Even theoretical works without any kind of quantification might have turned out a bit different if corpora or corpus-like data had been consulted systematically, and I think it is fair to say that usage-based linguistics and CxG have evolved precisely in this direction. For instance, Hamunen (2017) is not the least bit quantitative but still not only bases its diachronic exploration of the Finnish Colorative Construction mostly on 1,741 examples from three different corpora/corpus-like databases (viz. the Finnish Syntax archive, the Digital Morphology Archive, and the Digital Dictionary of Finnish Dialects), but also highlights all made-up examples as such. Beliën (2016) explicitly points out this methodological turn –

> the method is applied to corpus data, because they show what types of structures are actually produced by speakers, and in which contexts. Earlier studies, on the other hand, relied on isolated, constructed sentences, with diverging grammaticality (or acceptability) judgments as a result. The authentic data presented here were collected from the 38 million-word corpus of the Institute for Dutch Lexicology . . . and from the Internet. *(2016: 13)*

– before discussing the failure of traditional syntactic constituency tests regarding the analysis of Dutch particle constructions. However, it seems to me that most more recent and contemporary studies based on corpus data do involve at least some kind of quantification and I think that there are very few questions, if any, that cannot or should not be studied quantitatively *as a matter of principle* (but of course, there may be situations where, for example, data sparsity may rule out the use of certain statistical methods); see Jenset and

McGillivray (2017: section 3.7); Gries (2019b: 25–29), or Gries (2021 [2009]: section 1.2) for more on this question. We now turn to the simplest kind of quantification: frequencies of (co-)occurrence and (conditional) probabilities.

7.2.2 Frequencies of (Co-)occurrence and Conditional Probabilities

In spite of the statistical simplicity of frequencies and probabilities, if they are applied in the right kind of research context, they can be instructive, as is evidenced by a variety of studies having to do with issues of frequency as a mechanism driving, affecting, or at least correlating with entrenchment, learning/acquisition, language change, and productivity. For instance, in the area of language acquisition/learning, by now classic studies such as Goldberg's (1999) analysis of L1 acquisition data from CHILDES (to determine how highly frequent semantically light verbs facilitate the acquisition of semantically similar argument structure constructions) or Ellis and Ferreira-Junior's (2009a, 2009b) longitudinal study of L2 acquisition of verb-argument constructions in the European Science Foundation corpus were among the first to empirically highlight the importance of frequency of occurrence (of constructions) and frequency of co-occurrence (words in constructional slots) for language acquisition/learning or for the ubiquity of Zipfian distributions of constructions or for material within slots of constructions. Another hugely influential application of conditional probabilities – as cue validity – is Goldberg et al. (2004), which shows that certain patterns (e.g., V-Obj-Loc) have very high cue validities for certain meanings (e.g., caused motion), which reinforces the notion of constructions as pairings of form and meaning reliable enough to facilitate acquisition based on recognizing association patterns and chunking.

Quantitatively similar applications can also be found in other areas. An example of how corpus frequencies can inform theoretical argumentation is Boas (2004), who challenges a Minimalist Program account of *wanna* contraction in English. He shows that less than 1 percent of the examples of *wanna* contraction in the Switchboard corpus are instances within WH-clauses, which is interesting because most analyses put a lot of emphasis on *wanna* contraction in WH-clauses even though *wanna* contraction is actually more frequent than *want to* in relative clauses. As Boas (2004: 482) argues, if a theory of language claims not only to be descriptively but also explanatorily adequate, the question for Ausín's (2001) minimalist program analysis is how it may account for these differences in distribution.

Another study that is based on statistically very down-to-earth percentage data but uses them to make valuable theoretical contributions is Gaeta and Zeldes (2017). They use DeWaC, a 1.6 billion-word corpus of web-based German to study *-er* compounds (with agent noun heads) from a Construction Morphology perspective. On the basis of type, token, and hapax counts, they explore with which frequencies different combinations and orders of compounds are attested and the direction in which prototypical instances are generalized and argue that Construction Morphology's flexibility (in terms of permitting different derivational pathways of compounds) makes it an approach that supersedes purely syntactic or purely morphological approaches.

Quantitatively similar work – using type and token frequencies – is also found in Quochi (2016), a paper on a radial-category family of Italian light-verb constructions and their acquisition in L1 data from the CHILDES database. Approximately 2,100 instances of *fare* ('do') + noun constructions from children and adults are investigated in terms of the nouns/noun categories they occur with and the type–token ratios of verb-related nouns. Tracking new types over time she finds, among other things, that *fare* + nouns derived from verbs by suffixation appear to be rote-learned rather than instances of creative production. The general time course of acquisition, Quochi observes, is one where children first pick up on the most frequent uses, then develop a more abstract schema, which becomes generalized to intransitive actions, a development that is compatible with usage-based approaches to language acquisition of the kind outlined by Tomasello (2003), among others.

Let us finally look at a couple of statistically simple yet interesting applications that also bridge the gap to studies that involve higher degrees of statistical complexity. One of these is Vázquez Rozas and Miglio's (2016) study of which linguistic features are associated with Spanish and Italian speakers' choices of experiencer-as-subject (ES) and experiencer-as-object (EO) constructions. They look at clauses with an experiencer and a stimulus, where some such clauses construe the experiencer as Subject and the stimulus as Object while others have experiencers coded as dative/accusative Objects and stimuli as Subjects. For Spanish, they rely on the ARTHUS corpus of American and Peninsular Spanish; for Italian, they combine several databases to approximate a similar (and similarly sized) corpus (La Repubblica, C-ORAL, and the BAdIP database). Both corpora were searched for two-argument clauses with active-voice feeling verbs (excluding volition verbs). The main body of their paper reports a variety of frequency/percentage results for many different features of the clauses, including experiencer animacy, person, number, syntactic category, as well as stimulus animacy and syntactic class, and register/genre. Specifically, they point out correlations between ES vs. EO choice and experiencer and stimulus characteristics. However, they go beyond these monofactorial explorations by also subjecting the data to a multi-factorial analysis using a conditional inference tree, which is much more able to identify complex relations and interactions in the data, in particular how discourse-related factors can interact with syntactic form and semantic structure of the clause. Their paper therefore bridges the gap from frequency/percentages-only studies to the kind of multi-factorial work that seems to be the state of the art today and will be discussed more below.

Another interesting application is Chen (2017), a diachronic CxG study based on (i) contemporary Mandarin Chinese data from the Academia Sinica Balanced Corpus of Modern Chinese and (ii) diachronic data for Old

Chinese, Middle Chinese, and Early Mandarin from the Academia Sinica Ancient Chinese Corpus. She tracks the frequency of senses and what they co-occur with to explore how diachronic realignment processes gave rise to a synchronic polysemy network of 'one'-phrases in Mandarin involving counting/quantifying senses, but also meanings involving a negative-polarity sense and an attenuating positive polarity sense. As Chen concludes, "The associations [between 'one'-phrases and already established constructions] have been shaped by the environments where the 'one'-phrases frequently occur. The combination inherits syntactic, semantic, and pragmatic properties from the higher-level constructions, leading to new constructs" (2017: 97). This makes for a perfect transition to one of the, if not *the*, most widely used statistical methods in corpus-based CxG, the measurement of association strength and its implications for acquisition/learning, use, and change, which is the topic of the next section.

7.2.3 Association Strengths

Another frequent statistical method in corpus-based CxG involves a class of measures called association measures, that is, measures that are ultimately based on frequencies but then quantify the degree to which (typically two) elements from any level of the constructicon like or dislike to co-occur with each other; or, put differently, the degree to which the presence of one element makes the presence of another element more likely. This is a central issue for many questions, from as seemingly minute as the preference of words to occur with particular inflectional morphemes via the preference of words to occur in syntactic/argument structure constructions to, most fundamentally, actually any association of form and meaning (e.g., as when children determine from co-occurrence patterns that certain verbs have certain meaning and like to occur in certain constructions). The maybe most widely used statistical application in this context involves quantifying the degree of association between words and (slots of) constructions. The four papers by Stefanowitsch and Gries (2003, 2005) and Gries and Stefanowitsch (2004a, 2004b) develop a family of methods referred to as 'collostructional' analysis (see also Chapter 6), a blend of *collocation* and *construction*:

- collexeme analysis: the quantification of how much words are attracted to, or repelled by, a syntactically defined slot in a construction (e.g., the verb slot in the ditransitive construction or the noun slot in the N-*waiting-to-happen* construction);
- (multiple) distinctive collexeme analysis: how much a word (dis)prefers to occur in a certain slot of two or more functionally similar constructions (e.g., the verb slot in the two constructions making up the dative alternation); and
- two variants of covarying collexeme analysis: how much elements in two slots of one construction (dis)like to co-occur (e.g., the two verb slots in the *into*-causative, i.e., in V DO_{NP} *into* V-*ing*).

Table 7.1 *A schematic co-occurrence table underlying nearly all association measures*

	Element 2	Not element 2	Sum
Element 1	*a*	*b*	*a*+*b*
Not element 1	c	*d*	c+*d*
Sum	*a*+c	*b*+*d*	*N*

Most applications of either of these methods have been based on 2×2 co-occurrence tables such as Table 7.1, in which the elements' and cell frequencies' meanings depend on which analysis one conducts:

- for a collexeme analysis of the ditransitive,
  - element 1 might be one verb in the ditransitive (e.g., *give*) and element 2 would be the ditransitive construction;
  - *a*+*b* would be *give*'s frequency in the corpus, *a*+*c* would be the ditransitive's frequency in the corpus, and *a* would be the frequency of *give* in the ditransitive;
- for a distinctive collexeme analysis of the dative alternation,
  - element 1 might be one verb in the ditransitive or the prepositional dative (e.g., *give*), element 2 might be the ditransitive construction, and 'not element 2' would be the prepositional dative;
  - *a*+*b* would be *give*'s frequency in the corpus, *a*+*c* would be the ditransitive's frequency in the corpus, and *b*+*d* would be the frequency of the prepositional dative;
- for a covaring collexeme analysis of the *into*-causative,
  - element 1 might be one verb_1 in the *into*-causative (e.g., *trick*) and element 2 would then be a verb_2 in the *into*-causative (e.g., *believe*);
  - *a*+*b* would be *trick*'s frequency in the verb_1 slot of the *into*-causative, *a* +*c* would be *believe*'s frequency in the verb_2 slot of the *into*-causative, and *a* would be the frequency of *trick* DO_{NP} *into believing* in the *into*-causative.

Each of these applications follows a very similar four-step template, which is identical to the same decades-old approach in collocation studies in non-CxG corpus linguistics:

(1) one retrieves (ideally) all instances of a construction of interest *C*;
(2) for the element(s) of interest (e.g., a verb in a slot of *C*) one computes (a) measure(s) of association that is/are (usually) based on the relevant 2×2 tables of the above kind;
(3) one sorts the elements of interest according to that association measure; and
(4) one analyzes the top *x* elements of interest (often called collexemes) in terms of their structural, semantic, or other functional characteristics.

This family of methods was already relatively widespread ten years ago, when it was already used in studies on near-synonymous constructions (alternations), where, for instance, the method was precise enough to discover the iconicity difference (Thompson & Koide 1987) between the ditransitive (small distances between recipient and patient) and the prepositional dative (larger distances between recipient and patient) and many other domains: for example, in the study of priming effects (Gries 2005; Szmrecsanyi 2006), L1/L2 acquisition and learning of constructions (Gries & Wulff 2005, 2009; Ellis & Ferreira-Junior 2009a, 2009b; Wulff & Gries 2011, and especially the extremely comprehensive Ellis et al. 2016), constructional change over time (Hilpert 2006, 2008), etc. In addition, the approach has received some experimental support (Gries et al. 2005, 2010) and has stimulated research that combined it with other methods. Backus and Mos (2011), for instance, explore the productivity and similarity of two Dutch potentiality constructions – a derivational morpheme (*-baar*) and a copula construction (SUBJ $\text{COP}_{\text{finite}}$ *te* INF) – and combine association measures with acceptability judgments. They report the results of a distinctive collexeme analysis to determine which verbs prefer which of the two constructions in the Corpus of Spoken Dutch, and follow this result up with a judgment experiment to probe more deeply into seemingly productive uses of the constructions. The authors find converging evidence such that acceptability is often correlated with corpus frequencies and lexical preferences (see the chapters in Schönefeld 2011 for more examples of converging evidence and more on frequency vs. acceptability below).

More recent applications have broadened the scope even more, have used suggested improvements, and/or even extended the method and, thereby, have added to the theory of CxG. For example, Hoffmann et al. (2019) extend collostructional analysis by exploring the elements in slots on a more schematic level and the correlations between what happens in a construction's slots on that more abstract level. They study 1409 tokens of the comparative correlative constructions (e.g., *the more, the merrier*) from the 2015 part of the Corpus of Contemporary American English (COCA) in terms of several of the construction's characteristics: the grammatical/phrasal filler type (of either comparative), the lexical filler, and the presence/absence of different kinds of deletion. They first apply a covarying collexeme analysis using the unidirectional association measure Delta P (Ellis & Ferreira-Junior 2009b; Gries 2013b) rather than the usual bidirectional measures, and not only explore the words in the slots per se, but also in relation to the more schematic characteristics. Among other things, they find that the only filler types significantly attracted to each other are pairs of the same filler type, indicating that one's account of the construction should not attempt to treat the construction's slots as independent, an observation that can only be made when corpus data meet statistical methods in the analysis. Another example of the use of the unidirectional Delta P approach to collostructions is Rastelli's (2020) analysis of lexical aspect in L2 Italian, which is not a study of argument structure (the 'usual suspect' in this type of analysis) but of lexeme–morpheme associations.

A generally similar analytical approach, which also explores co-occurrences at multiple levels of generality, is pursued in Abdulrahim (2019), who studies 'go'-constructions with three types of verbs in Modern Standard Arabic and their association to a variety of lexico-syntactic features. Abdulrahim uses a multi-dimensional extension of collostructions, so to speak, Hierarchical Configural Frequency Analysis (HCFA, Gries 2021 [2009]; Stefanowitsch & Gries 2005), a method that tries to identify over- and under-represented cells in multi-dimensional frequency tables.

None of the above should imply that collostructional analysis has not also been criticized, but much of the critique was based on a variety of misunderstandings with regard to both the method's goals and their implementation. For instance, with regard to the former, Bybee (2010: chapter 5) criticized collostructional analysis for its lack of considering semantics (especially on the *in*put side of the analysis) when in fact the whole point of collostructional methods is to be able to infer semantic (or other functional patterns) from its *out*put. Similarly, Bybee criticized the collostructional approach for a lack of discriminability in her results, but did not actually perform a full-fledged analysis herself: Rather than using the method to all words in a certain construction and as described in the four steps above, she restricted her input to extremely low-frequency items that collostructional methods were not developed for and then performed only step 2 of the above four. Schmid and Küchenhoff (2013) suffers from similar problems. For instance, they misunderstood how software handles extremely small values (e.g., $<10^{-320}$) and falsely claim that one needs more powerful computers for collostructional computations (when all that is needed is a specific software package, which would allow any normal computer to handle such numbers); also, they object to how most collostructional applications compute association strength (using the *p*-value of a Fisher–Yates exact test), but at least some of their argumentation is self-contradictory: For instance, they criticize p_{FYE} for, among other things, being bidirectional, but devote quite some space to discussing an alternative they prefer, the odds ratio, which is *also* bidirectional (for the specifics of this debate, see Bybee 2010: chapter 5 and Gries 2012 for a rebuttal, as well as Schmid & Küchenhoff 2013 and Gries 2015a for another rebuttal). Gries (2019a) is an attempt to place collostructional analysis on a new statistical foundation, by encouraging the use of many more and independent dimensions of information that an analyst should consider, namely frequency, association (independently of frequency and potentially bidirectionally), dispersion, entropy, and possibly others.

In some recent research, collostructional methods are now more often combined with other kinds of data and methods (see the discussion in Ellis et al. 2016; Sommerer & Baumann 2021; Chen 2022), and collostructional results are now sometimes included as predictors or control variables, given how they can help bring item-specific (e.g., verb-specific) variability under statistical control. This may also help validate/critique the approach, but of course much remains to be done and by now many such attempts are underway. For example, Bernolet and Colleman (2016) raise the bar for just about all collostructional studies in that they take polysemy more seriously than nearly all others and incorporate sense information into the analysis. Gries (2015b) is a first step to try to disentangle the correlations between directions of attraction and experimental data in the *as*-predicative. Flach (2020b) revisits the frequency vs. association issue with data on *gonna*/*wanna*/*gotta* contraction and shows that contingency/association measures consistently outperform string frequency. Finally, Herbst (2020) is an interesting new proposal to change one's perspective on co-occurrence away from a view of items-attracted-to-constructions (as in all collostructional studies, e.g., verbs in the verb slot of an argument construction) to a view of items-in-constructions.

The above should not also imply that collostructional studies are the only examples of association measures in corpus-based CxG. As an example of a different kind of application, Cappelle et al. (2019) retrieve *n*-grams involving necessity modal verb lemmas from the BNC that meet a frequency and an association strength threshold (50 and MI ≥ 3 respectively). Adopting a perspective of "contexts as constructions," they then cluster the modal verb lemmas on the basis of the contexts they share; they find a hierarchical cluster structure that can be represented as (([*have to*, *need to*], *must*), *should*) (with parentheses and square brackets indicating less and more robust clusters respectively). While not much is done with that specific quantitative result, Cappelle et al. proceed with some qualitative discussion of how the modals' functions are reflected in terms of the preferred *n*-grams.

The second most widespread quantitative treatment of corpus data in CxG involves various kinds of modeling, to which we turn now.

7.2.4 Monofactorial, Multi-factorial, and Multivariate Approaches

Corpus-based CxG studies using both mono- and multi-factorial tests have increased substantially, especially over the last ten or so years. Petré and Anthonissen (2020), for example, report results from monofactorial regressions on individual variation in diachronic data, finding, among other things, excellent fits of (i) first attestations of motionless *be-going-to* INF in the sixteenth to seventeenth centuries with time (the expected logistic *s*-curves) and (ii) within-individual uses of Nominativus-cum-Infinitivo and prepositional passive constructions. However, the 'standard' by now are multi-factorial/multivariate approaches. While I am splitting this section up into 'inferential' and 'exploratory' approaches, it needs to be pointed out that the dividing line is often more tenuous than one might think. Many studies use inferential tools, such as regression modeling, but incorporate a certain degree of exploration because their modeling involves model selection; similarly, a method like HCFA as in Abdulrahim (2019) also combines inferential and exploratory aspects. In addition, the notion of exploratory I am using is rather broad and intended to cover all methods covered in traditional statistics textbooks such as different kinds of cluster analyses, principal component/factor analysis, correspondence analysis, and multi-dimensional scaling, but also unsupervised machine-learning methods such as vector spaces and deep learning.

Inferential/Statistical Approaches

It seems as if the vast majority of multi-factorial corpus-based CxG studies uses some kind of *regression modeling*, that is, the application of statistical tools that are extensions of simple correlational statistics to situations where the behavior of one response variable (often the effect of a hypothesized cause–effect scenario) is explored with regard to how it varies as a function of multiple predictor variables (often the causes in that hypothesized cause–effect scenario). The range of applications of such methods is huge because they are useful for really any kind of correlational hypothesis and, at least as a proxy, for any kind of causal hypothesis that can be 'translated' into a correlational effect or pattern of effects.

As an example in the areas of individual variation and productivity, De Smet (2020) studies constructional morphological productivity (*-ly* and *-ness* derivation) based on hapaxes across individuals in the NY Times and Hansard corpora in order to tease apart effects of token and type frequency (when controlling for several other factors in a series of linear models); interestingly, he finds an interaction effect between the frequency types that supports "a view of entrenchment as both a conservative and creative force in language" but also notes that "some variation remains irreducibly individual" (De Smet 2020: 251).

A big topic is alternation research on various phenomena and, by now, for various languages, and the field has come a long way since some of the earliest multi-factorial studies in Cognitive CxG (e.g., Gries 2003), surpassing those in sample sizes and sophistication. De Vaere et al. (2021) is a case in point. They study German *geben* 'give' in two alternating ditransitive constructions based on 1301 occurrences from the DeReKo corpus, which were annotated for twenty morphosyntactic, semantic, and pragmatic factors and submitted to a logistic regression model. Intriguingly, in some ways, they go much beyond the current standard:

- Most existing studies assume (usually implicitly) that the effect of numeric predictors can be modeled with a straight line (i.e., a linear trend), which is surprising given that very many cognitive phenomena do not apply linearly: Learning, forgetting, priming, language change, etc. all involve curved trends. Laudably, De Vaere et al. accommodate this fact by allowing their numeric predictors to be curved.
- Many existing studies run the risk of what is called overfitting, that is, the risk that a model that is fit on a certain data set fits that data set so well that it does not also generalize well to other data sets. De Vaere et al. use a statistical method called penalization, whose details are not relevant in the present context, to

protect their analysis against that risk. In addition, they also use a technique called bootstrapping to make sure their model quality statistics do not exaggerate the model's quality.

- Many analyses of observational data suffer from the fact that linguistic predictors are often highly correlated with each other, a phenomenon called (multi-) collinearity. For example, NPs referring to discourse-given referents are often not just given but also short, definite, pronominal, etc. De Vaere et al. report collinearity diagnostics so that readers can contextualize their findings better.

They interpret their findings as providing evidence for the main meaning of *geben* being not so much literal transfer from one person to another (as in *give* or *hand*) but a more general transfer meaning, and highlight the fact that one of the constructions is often associated with the passive voice; this echoes Gries et al. (2005) and points to a more general need to include voice as a variable in collostructional and/or alternation studies (see also Pijpops et al. 2018 for another application of logistic regression on constructional contamination).

Even more frequent than fixed-effects regressions are currently mixed-effects models, which in various ways take into consideration speaker-/file-specific effects (are there systematic individual differences between speakers or files?) as well as item-specific effects (are there systematic effects?). The following is just a small overview of the published work:

- In non-native speaker/L2 research, Wulff and Gries (2019, 2021), Gries and Wulff (2021), and Azazil (2020). The study by Azazil is noteworthy for combining multiple predictive modeling methods (mixed-effects models and random forests) and for showing how such studies support the notion of frequency-based entrenchment of item-specific information.
- In native speaker alternation research, Schäfer (2018) studies the measure NP alternation in the 21 billion-word German DECOW14A corpus and, on a theoretical level, concludes that speakers' choices require mechanisms from both prototype and exemplar models, which makes an important contribution to corpus-based studies on (degrees of) mental representation and abstraction. Flach (2020b) is another relevant study mentioned in Section 7.2.3.
- In work bringing together corpus and experimental data beyond that already mentioned above, Flach (2020a) explores the frequency–acceptability mismatch – the fact that corpus frequencies are often not a good predictor of acceptability ratings. She combines corpus data from COCA (collostructional results and the results of a correspondence analysis on *go*/*come*-V in nine different syntactic contexts) with the results of an acceptability-judgment experiment to explore, with mixed-effects modeling, what resolution of frequency is most related to the acceptability-judgment data. She concludes that "acceptability is a function of compatibility with a licensing schema, which accounts for the acceptability even of rare or corpus-absent patterns" (p. 636) and "acceptability patterns are better captured by complex than by simplistic measures" (p. 637); see also Gould & Michaelis (2018) or Busso et al. (2021) for additional examples of studies coupling observational and experimental data.

- In a diachronic (1300–2000) study of strong vs. weak past tense in several corpora of Old Dutch, De Smet and Van de Velde (2020) use mixed-effects modeling to show how the realization of past tense varies systematically with aspect (durative vs. punctual) and meaning (metaphorical vs. literal).

There is also a slowly growing set of studies that deal with 'curvature' in the structure between predictors and responses. For instance, apart from the above-mentioned De Vaere et al. (2021), Wulff and Gries (2019, 2021) incorporate polynomial predictors in mixed-effects models for learner data, and Lorenz and Tizón-Couto (2019) use generalized additive mixed models in their study on the role of corpus frequency on phonological reduction.

Another recent development in much of linguistics and also in corpus-based CxG is the use of 'tree-based methods' such as classification trees and random forests, that is, machine-learning methods that often appear to be an attractive plan B when the nature of the data seems to not license regression modeling. Tree-based methods try to identify structure in the relation(s) between a response and multiple predictors by determining how the data set can be split up repeatedly into successively smaller groups (based on the values of the predictors); to simplify a bit, each split increases the tree's/forest's ability to predict the response variable (which can be numeric, but is more often categorical, such as one of several constructional choices). For instance, Fonteyn and Nini (2020) use both tree-based methods in a diachronic analysis of the gerund alternation (e.g., *the eating of meat* vs. *eating meat*) that included language-internal and -external factors and identify similarities and differences between different speakers in the 90 million-word EMMA corpus. Soares da Silva et al. (2021) use a conditional inference tree to model, among other things, the alternation of overt and null *se* constructions in Brazilian and European Portuguese from two decades and find language-internal factors (the construal of the change of state or voice) as well as language-external factors (register) to be relevant. Finally, there is work that combines corpus and experimental data as well as mixed-effects modeling and tree-based statistics, for example, Azazil's (2020) study of frequency effects in the L2 acquisition of the catenative verb construction by German learners of English (following up on Gries & Wulff 2009 and Martinez-Garcia & Wulff 2012).

Finally, other kinds of computational modeling are also found: Liu and Ambridge (2021) is a study of four two-argument constructions involving actives and passives from the CCL corpus that uses Bayesian mixed-effects modeling but also naive discriminative learning (Baayen 2011), a computational learner without the hidden layers characteristic of many connectionist/neural network learners, that has been argued to "enjoy psychological plausibility" (Liu & Ambridge 2021). Their results reflect how speakers balance information-structural and semantic constraints and suggest that competing constructions are retained because they offer speakers choices to express both topicalization and other implications at the same time. At the same time, their findings tell a cautionary tale as to the psychological reality of such learners because the computational learner improved when a cue that humans are sure to use – the specific lexical item – was removed from the learner. Nevertheless, such studies are interesting additions to the inventory of multi-factorial/multivariate methods that have taken corpus-based CxG by storm.

Exploratory/Computational Approaches

While there is of course the major body of work on Fluid Construction Grammar – see, for example, the special issue of *Constructions and Frames* (2017, Vol. 9, Issue 2), also Chapter 10 – there is now also much more computational-linguistic work in CxG than even eight to ten years ago. At the risk of some simplification, we can distinguish two main kinds of exploratory studies. First, there are those that are largely descriptive in nature; in such studies the starting point is one or more constructions and the goal is to see what we can learn about their function pole(s) from the results of exploratory tools applied to their distribution. Second, there are those exploratory studies whose focus is on identifying construction types and tokens in corpora in a bottom-up way; thus, in such studies, the starting point is not a construction whose distributional behavior is explored – instead, the starting is a corpus and constructions extracted from it in an automated way are the endpoint/goal (see also Chapter 23). Over the last ten years or so, both kinds of studies have become noticeably more frequent.

As a first example of the former kind of exploratory studies, Flach (2021) uses a technique called variability-based neighbor clustering – a method to identify clusters in temporal data (e.g., acquisition or historical corpus data) that respects the temporal ordering of the data (Gries & Hilpert 2008) – to identify temporal stages in the way the *into*-causative construction slots have become more lexically diverse over the last 200 years. She further shows how this change is accompanied by a subtle change in the construction's semantics.

Next, consider the body of work by Hilpert and colleagues on modal constructions. For example, Hilpert (2016: 70) explicitly extends the theory by arguing that "knowledge of a construction includes probabilistic knowledge of how that construction is associated with lexical elements" and, accordingly, combines the logic underlying frequencies and association measures with the use of multivariate exploratory methods. Using data from COCA and COHA, he explores the similarities and diachronic development of the collocational profiles of a variety of English modal verbs. For instance, multi-dimensional scaling of modals based on collocate frequencies reveals, among other things, clines from informational to interpersonal uses and from deontic to epistemic modality. That kind of analysis is then extended to the diachronic data, reflecting how the location of *may*, for instance, in this 'modal space' changes over time. Then, Hilpert follows up on an earlier collostructional analysis with a diachronic semantic vector space analysis, whose results show in an unprecedented bird's-eye view how the distribution of *may*'s collocates changes over time with regard to the dimensions of abstractness and volitionality/physicality. In a related paper, Hilpert and Flach (2020) contrast *may* and *might* by identifying and comparing their second-order collocates, using such a vector space method, and validate the accuracy of the collocational differences by (i) reducing the collocational space using multidimensional scaling and (ii) using a binary logistic regression to determine the classificatory power of the collocates for modal choice. While the obtained classification accuracies are only moderate, Hilpert and Flach (2020: 13) argue that second-order collocates "provide a statistical signal that facilitates the discrimination of deontic and epistemic modal meaning," which in turn supports the notion of "linguistic knowledge as a network of symbolic units that are mutually interconnected at different levels of schematicity" (see also Hilpert & Saavedra 2020 for a more general characterization of their methodology).

Apart from the increasing interest in vector space approaches, network-based approaches are also slowly garnering more attention. One particularly prominent example is maybe Ellis et al. (2016), who develop semantic networks for the verb-argument constructions they study (e.g., the V *about* N construction, the V *across* N construction, etc.), then derive a variety of statistics from those (e.g., betweenness and degree centrality, density, and others), and, most interestingly, apply a community-detection algorithm to them in order to identify a variety of semantically related coherent groups of verbs in these constructions; these in turn shed light on the polysemy of constructions and the prototypical members of semantic groups of constructions. Another example of a network study is Chen's (2022) network of Mandarin Chinese space particles in the constructional schema *zai* + NP + space particle in the 10 million-word POS-tagged Sinica corpus. Approximately 26,000 pairs of nouns and particles from these constructions were analyzed with a network approach based on three inputs: (i) collostruction strengths between nouns and particles from a covarying collexeme analysis, (ii) similarities between the nouns from a word2vec model, and (iii) cosine similarities between the particles. Chen shows that the network exhibits a scale-free structure, meaning that only a few nodes are frequently connected to other units and that most other nodes are relatively unconnected – a striking emergence of the well-known Zipfian distribution of words in constructional slots at the level of a constructional network. Also, the network indicates that experientially and interactionally more prominent particles exhibit higher degrees of local clustering and, thus, more semantic homogeneity. These kinds of observations – and others, for example, about prototypicality within the network – would be extremely hard to make on the basis of just qualitative analysis and, therefore, testify to the power of these more advanced types of methods.

As for the kind of inductive construction-identification studies that constitute the second major area of exploratory/computational CxG work, one example is Martí et al. (2019), whose DISCOver algorithm is "an unsupervised methodology for the automatic identification and extraction of lexico-syntactic patterns that are candidates for consideration as constructions." This, too, is essentially a vector space method that involves identifying dimensions in co-occurrence data for lemmas and syntactic dependency relations in their contexts, specifically "tuples involving two lexical items (lemmas) related both by a dependency direction and a dependency label." Their method, while tested on 15,000 lemmas from one specific corpus (the 94 million-word Diana-Araknion corpus of Spanish), is applicable to any corpus with POS and syntactic dependency annotation from which one can construct clusters of lemmas that are related by their preference for a set of lexico-syntactic contexts. Interestingly, the approach makes it possible to identify construction candidates that are actually attested in the data as well as unattested-but-likely construction candidates that merit scrutiny by the human analyst.

A somewhat similar approach is Dunn (2018), who first runs a CxG induction approach (C2XG) on the ukWaC corpus and then uses the grammar learned from that to measure the similarity between inner- and outer-circle varieties of English (from the ICE project and the Leipzig corpora collection). The first part, the induction algorithm, requires as input three different levels of information for each 'word':

- a lexical level consisting of whitespace-separated 'words';
- a morphosyntactic level consisting of part-of-speech tags assigned to those words; and
- a semantic level, which approximates the semantic/conceptual pole of a word with a distributional-semantics-based vector representation.

Thus, each word is represented as combination of information of these levels in *n*-dimensional space, which can then be clustered (e.g., using *k*-means analysis, a kind of cluster analysis where data points are grouped into a user-defined number of clusters).

The second part of the analysis is a classification task that attempts to determine how well a machine-learning algorithm can predict English varieties from the (relative) frequencies of the construction candidates arrived in the first step; in other words, the question is whether English varieties exhibit distinctive behavioral profile-like distributions of constructions. On a meta-theoretical level, this kind of work – cognitive sociolinguistic work that models many speakers of a variety as a whole – can provide the regionally-dialectally-motivated counterpoint to studies of individual variation.

7.3 Concluding Remarks

Given all of the above, what *is* the state of the art in corpus-based CxG? I think it is fair to say that, after the field's Big Bang in the late 1980s, the field is still exhibiting a rapid but healthy expansion. Compared to the relative (!) paucity of corpus studies discussed in Gries (2013a), there is now a multitude of studies covering all the parameters mentioned at the beginning of this overview: (kinds of) languages studied; temporal orientation; range of corpora, registers, and genres; resolution (individual(s), speech communities); scientific goals (description, theoretical development, computational simulation); and statistical methods. Even from the highly selective review offered above, it seems as if nearly every combination of choices from these features is now a lively field of inquiry advanced by the continued development, application, and – by now often – combination of quantitative methods to constructional corpus data.

However, corpus-linguistic methods and analysis have not only simply become more frequent (to the point of being mainstream), they have also helped to advance the theory itself. From Goldberg's revision of her definition of a construction from Goldberg (1995) to Goldberg (2006) (which did away with non-compositionality as a necessary condition but added sufficient frequency as criterion) and Hilpert's (2016) addition of probabilistic knowledge of how a construction is used to constructional knowledge, to Cappelle et al.'s (2019) perspective of context-as-construction and Flach's (2020a) determination of the level of granularity constructional co-occurrence matches best one of the oldest linguistic methods (acceptability judgments), there are many ways in which corpus-based CxG has made valuable contributions (even if many may need to be revised later). In addition, one cannot help but feel that the overall quality of the field has increased as well. I would like to think this is not only a subjective impression but an assessment that can also be supported by looking at a recent critical review of cognitive-linguistic work, of which much of CxG is probably a part, namely Dąbrowska (2016). She catalogued seven deadly sins of cognitive linguistics, which I would summarize as follows:

(1) excessive reliance on introspection;
(2) not treating the Cognitive Commitment seriously;
(3) not enough serious hypothesis testing;
(4) ignoring individual differences;
(5) neglecting the social aspect of language;
(6) assuming that we can deduce mental representations from patterns of use; and
(7) assuming that distribution equals meaning.

While I 'only' agree with most of the points Dąbrowska is making, it does seem to me as if much of the CxG work summarized above addresses (mostly implicitly) many of these issues superbly. For instance:

- regarding 1 and 2, we see much less reliance on introspection in general, but also the combination of corpus data with various kinds of experimental work, computational simulation, interrater reliability, etc.
- regarding 3, we see a *lot* of hypothesis testing now, with a wide range of sophisticated statistical/machine-learning models and networks; and
- regarding 4 and 5, we see more work on both these aspects.

And this does not even count the spread of CxG-inspired work into areas I have not discussed at all, for example, work on constructions and their preferences and alternations in indigenized varieties of World Englishes (as in Mukherjee & Gries 2009; Gries & Mukherjee 2010; Bernaisch et al.

2014; Heller et al. 2017; Röthlisberger et al. 2017; Rautionaho & Deshors 2018; Brunner & Hoffmann 2020; Hoffmann 2020b, etc.). Thus, to my inevitably biased mind, the field can take a certain degree of pride in these developments which, in spite of the high degree of inertia of academia, have happened in a rather short period of time. That does not mean it is time to rest on our laurels (see Hoffmann 2020a for a recent call to include more psycholinguistic and neurolinguistic data to CxG's arsenal), but it does inspire hope for high standards of, and interesting findings from, future research. It is a good time to be a corpus-based Construction Grammarian; here's to the next ten to twenty years!

References

Abdulrahim, D. (2019). 'Go' constructions in Modern Standard Arabic: A corpus-based study. *Constructions and Frames*, 11(1), 1–42.

Ausín, A. (2001). *On A-Movement*. PhD thesis. University of Connecticut.

Azazil, L. (2020). Frequency effects in the L2 acquisition of the catenative verb construction: Evidence from experimental and corpus data. *Cognitive Linguistics*, 31(3), 417–451.

Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, 11(2), 295–328.

Backus, A. & Mos, M. (2011). Islands of productivity in corpus data and acceptability judgments: Contrasting two potentiality constructions in Dutch. In D. Schönefeld, ed., *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*. Amsterdam & Philadelphia: John Benjamins, pp. 165–192.

Beliën, M. (2016). A constructional perspective on conceptual constituency: Dutch postpositions or particles? In J. Yoon & S. Th. Gries, eds., *Corpus-Based Approaches to Construction Grammar*. Amsterdam & Philadelphia: John Benjamins, pp. 11–37.

Bernaisch, T., Gries, S. Th., & Mukherjee, J. (2014). The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide*, 35(1), 7–31.

Bernolet, S. & Colleman, T. (2016). Sense-based and lexeme-based alternation biases in the Dutch dative alternation. In J. Yoon & S. Th. Gries, eds., *Corpus-Based Approaches to Construction Grammar*. Amsterdam & Philadelphia: John Benjamins, pp. 165–198.

Boas, H. C. (2004). You wanna consider a Constructional Approach to Wanna-Contraction? In M. Achard & S. Kemmer, eds., *Language, Culture, and Mind*. Stanford: CSLI Publications, pp. 479–491.

Brunner, T. & Hoffmann, T. (2020). The *way*-construction in World Englishes. *English World-Wide*, 41(1), 1–32.

Busso, L., Perek, F., & Lenci, A. (2021). Constructional associations trump lexical associations in processing valency coercion. *Cognitive Linguistics*, 32(2), 287–318.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Cappelle, B., Depraetere, I., & Lesuisse, M. (2019). The necessity modals have to, must, need to, and should: Using n-grams to help identify common and distinct semantic and pragmatic aspects. *Constructions and Frames*, 11(2), 220–243.

Chen, A. C.-H. (2022). Words, constructions and corpora: Network representations of constructional semantics for Mandarin space particles. *Corpus Linguistics and Linguistic Theory*, 18(2), 209–235.

Chen, I.-H. (2017). The polysemy network of Chinese 'one'-phrases in a diachronic constructional perspective. *Constructions and Frames*, 9(1), 70–100.

Dąbrowska, E. (2016). Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics*, 27(4), 479–491.

Dancygier, B. & Sweetser, E. (1997). *Then* in conditional constructions. *Cognitive Linguistics*, 8(2), 109–136.

De Smet, H. (2020). What predicts productivity? Theory meets individuals. *Cognitive Linguistics*, 31(2), 251–278.

De Smet, I. & Van de Velde, F. (2020). Semantic differences between strong and weak verb forms in Dutch. *Cognitive Linguistics*, 31(3), 393–416.

De Vaere, H., De Cuypere, L., & Willems, K. (2021). Alternating constructions with ditransitive *geben* in present-day German. *Corpus Linguistics and Linguistic Theory*, 17(1), 73–107.

Dunn, J. (2018). Finding variants for construction-based dialectometry: A corpus-based approach to regional CxGs. *Cognitive Linguistics*, 23(2), 275–311.

Ellis, N. C. & Ferreira-Junior, F. (2009a). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3), 370–385.

Ellis, N. C. & Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, 7, 187–220.

Ellis, N. C., Römer U., & O'Donnell, M. B. (2016). *Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. New York: John Wiley.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64(3), 501–538.

Flach, S. (2020a). Schemas and the frequency/acceptability mismatch: Corpus distribution predicts sentence judgments. *Cognitive Linguistics*, 31(4), 609–645.

Flach, S. (2020b). Reduction Hypothesis revisited: Frequency or association? In C. Sanchez-Stockhammer, F. Günther, & H.-J. Schmid, eds., *Language in Mind and Brain: Multimedial Proceedings of the Workshop Held at LMU, Munich, December 10–11, 2018*. Munich: LMU Open Access, pp. 16–22.

Flach, S. (2021). From movement into action to manner of causation: Changes in argument mapping in the *into*-causative. *Linguistics*, 59(1), 247–283.

Fonteyn, L. & Nini, A. (2020). Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics*, 31(2), 279–308.

Gaeta, L. & Zeldes, A. (2017). Between VP and NN: On the constructional types of German -er compounds. *Constructions and Frames*, 9(1), 1–40.

Goldberg, A. E. (1991). The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics*, 3(1), 37–74.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney, ed., *The Emergence of Language*. Mahwah: Lawrence Erlbaum Associates, pp. 197–212.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15(3), 289–316.

Gould, K. M. & Michaelis, L. A. (2018). Match, mismatch, and envisioning transfer events: How verbal constructional bias and lexical-class concord shape motor simulation effects. *Constructions and Frames*, 10(2), 234–268.

Gries, S. Th. (2003). *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London & New York: Continuum Press.

Gries, S. Th. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399.

Gries, S. Th. (2012). Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language*, 36(3), 477–410.

Gries, S. Th. (2013a). Data in Construction Grammar. In T. Hoffmann & G. Trousdale, eds., *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, pp. 93–108.

Gries, S. Th. (2013b). 50-something years of work on collocations: What is or should be next . . . *International Journal of Corpus Linguistics*, 18(1), 137–165.

Gries, S. Th. (2015a). More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff (2013). *Cognitive Linguistics*, 26(3), 505–536.

Gries, S. Th. (2015b). The role of quantitative methods in Cognitive Linguistics: Corpus and experimental data on (relative) frequency and contingency of words and constructions. In J. Daems, E. Zenner, K. Heylen, D. Speelman, & H. Cuyckens, eds., *Change of Paradigms – New Paradoxes: Recontextualizing Language and Linguistics*. Berlin & New York: De Gruyter Mouton, pp. 311–325.

Gries, S. Th. (2019a). 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics*, 24(3), 385–412.

Gries, S. Th. (2019b). *Ten Lectures on Corpus-Linguistic Approaches: Applications for Usage-Based and Psycholinguistic Research*. Leiden & Boston: Brill.

Gries, S. Th. (2021 [2009]). *Statistics for Linguistics with R. A Practical Introduction*. Berlin: De Gruyter Mouton.

Gries, S. Th. , Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics*, 16(4), 635–676.

Gries, S. Th., Hampe, B., & Schönefeld, D. (2010). Converging evidence II: More on the association of verbs and constructions. In S. Rice & J. Newman, eds., *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford: CSLI Publications, pp. 59–72.

Gries, S. Th. & Hilpert, M. (2008). The identification of stages in diachronic data: Variability-based neighbor clustering. *Corpora*, 3(1), 59–81.

Gries, S. Th. & Mukherjee, J. (2010). Lexical gravity across varieties of English: An ICE-based study of *n*-grams in Asian Englishes. *International Journal of Corpus Linguistics*, 15(4), 520–548.

Gries, S. Th. & Stefanowitsch, A. (2004a). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9(1), 97–129.

Gries, S. Th. & Stefanowitsch, A. (2004b). Co-varying collexemes in the *into*-causative. In M. Achard & S. Kemmer, eds., *Language, Culture, and Mind*. Stanford: CSLI Publications, pp. 225–236.

Gries, S. Th. & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics*, 3, 182–200.

Gries, S. Th. & Wulff, S. (2009). Psycholinguistic and corpus linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics*, 7(1), 163–186.

Gries, S. Th. & Wulff, S. (2021). Examining individual variation in learner production data: A few programmatic pointers for corpus-based analyses using the example of adverbial clause ordering. *Applied Psycholinguistics*, 42(2), 279–299.

Gutzmann, D. & Henderson, R. (2019). Expressive updates, much? *Language*, 95(1), 107–135.

Hamunen, M. V. J. (2017). On the grammaticalization of Finnish Colorative Construction. *Constructions and Frames*, 9(1), 101–138.

Heller, B., Bernaisch, T., & Gries, S. Th. (2017). Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal*, 41(1), 111–144.

Herbst, Th. (2020). Constructions, generalizations, and the unpredictability of language: Moving towards colloconstruction grammar. *Constructions and Frames*, 12(1), 56–95.

Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, 2(2), 243–257.

Hilpert, M. (2008). *Germanic Future Constructions: A Usage-Based Approach to Language Change*. Amsterdam & Philadelphia: John Benjamins.

Hilpert, M. (2016). Change in modal meanings: Another look at the shifting collocates of *may*. *Constructions and Frames*, 8(1), 66–85.

Hilpert, M. & Flach, S. (2020). Disentangling modal meanings with distributional semantics. *Digital Scholarship in the Humanities*, 36(2), 307–321.

Hilpert, M. & Saavedra, D. C. (2020). Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. *Corpus Linguistics and Linguistic Theory*, 16(2), 393–424.

Hoffmann, T. (2020a). What would it take for us to abandon Construction Grammar? Falsifiability, confirmation bias and the future of the Constructionist enterprise. *Belgian Journal of Linguistics*, 34, 149–161.

Hoffmann, T. (2020b). Marginal argument structure constructions: the [V *the* $_{\text{Ntaboo-word}}$ *out of*]-construction in Post-colonial Englishes. *Linguistics Vanguard*, 6(1), https://doi.org/10.1515/lingvan-2019-0054.

Hoffmann, T, Horsch, J., & Brunner, T. (2019). The more data, the better: A usage-based account of the English comparative correlative construction. *Cognitive Linguistics*, 30(1), 1–36.

Hoffmann, T. & Trousdale, G., eds. (2013). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Hunston, S. & Francis, G. (2000). *Pattern Grammar*. Amsterdam & Philadelphia: John Benjamins.

Jenset, G. B. & McGillivray, B. (2017). *Quantitative Historical Linguistics*. Oxford: Oxford University Press.

Kay, P. & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The *What's X Doing Y?* Construction. *Language*, 75(1), 1–33.

Kemmer, S. & Verhagen, A. (1994). The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics*, 5(2), 115–156.

Labov, W. (1975). Empirical foundations of linguistic theory. In R. Austerlitz, ed., *The Scope of American Linguistics*. Lisse: Peter de Ridder Press, pp. 7–133.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford: Stanford University Press.

Liu, L. & Ambridge, B. (2021). Balancing information-structure and semantic constraints on construction choice: Building a computational model of passive and passive-like constructions in Mandarin Chinese. *Cognitive Linguistics*, 32, 349–388.

Lorenz, D. & Tizón-Couto, D. (2019). Chunking or predicting: Frequency information and reduction in the perception of multi-word sequences. *Cognitive Linguistics*, 30(4), 751–784.

Martí, M. A., Taulé, M., Kovatchev, V., & Salamó, M. (2019). DISCOver: DIStributional approach based on syntactic dependencies for discovering Constructions. *Corpus Linguistics and Linguistic Theory*, 17(2). https://doi.org/10.1515/cllt-2018-0028.

Martinez-Garcia, M. T. & Wulff, S. (2012). Not wrong, yet not quite right: Spanish ESL students' use of gerundial and infinitival complementation. *International Journal of Applied Linguistics*, 22(2), 225–244.

Morgan, P. S. (1997). Figuring out *figure out*: Metaphor and the semantics of the English verb-particle construction. *Cognitive Linguistics*, 8(4), 327–358.

Mukherjee, J. & Gries, S. Th. (2009). Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide*, 30(1), 27–51.

Petré, P. & Anthonissen, L. (2020). Individuality in complex systems: A constructionist approach. *Cognitive Linguistics*, 31(2), 185–212.

Pijpops, D., De Smet, I., & Van de Velde, F. (2018). Constructional contamination in morphology and syntax: Four case studies. *Constructions and Frames*, 10(2), 269–305.

Quochi, V. (2016). Development and representation of Italian light-fare constructions. In J. Yoon and S. Th. Gries, eds., *Corpus-Based Approaches to Construction Grammar*. Amsterdam & Philadelphia: John Benjamins, pp. 39–64.

Rastelli, S. (2020). Contingency learning and perfective morpheme productivity in L2 Italian: A study on lexeme–morpheme associations with ΔP. *Corpus Linguistics and Linguistic Theory*, 16(3), 459–486.

Rautionaho, P. & Deshors, S. C. (2018). Progressive or not progressive? Modeling constructional choices in EFL and ESL. *International Journal of Learner Corpus Research*, 4(2), 225–252.

Röthlisberger, M., Grafmiller, J., & Szmrecsanyi, B. (2017). Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics*, 28(4), 673–710.

Schäfer, R. (2018). Abstractions and exemplars: The measure noun phrase alternation in German. *Cognitive Linguistics*, 29(4), 729–771.

Schmid, H.-J. & Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24(3), 531–577.

Schönefeld, D. (ed.). (2011). *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*. Amsterdam & Philadelphia: John Benjamins.

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

Smith, M. B. (1994). Agreement and iconicity in Russian impersonal constructions. *Cognitive Linguistics*, 5(1), 5–56.

Soares da Silva, A., Afonso, S., Palú, D., & Franco, K. (2021). Null 'se' constructions in Brazilian and European Portuguese: Morphosyntactic deletion or emergence of new constructions? *Cognitive Linguistics*, 32(1), 159–193.

Sommerer, L. & Baumann, A. (2021). Of absent mothers, strong sisters and peculiar daughters: The constructional network of English NPN constructions. *Cognitive Linguistics*, 32(1), 97–131.

Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory*, 2(1), 61–77.

Stefanowitsch, A. (2007). Linguistics beyond grammaticality. *Corpus Linguistics and Linguistic Theory*, 3(1), 57–71.

Stefanowitsch, A. & Gries, S. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.

Stefanowitsch, A. & Gries, S. Th. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1), 1–43.

Szczesniak, K. (2019). Variation motivated by analogy with fixed chunks: Overlap between the reflexive and the way construction. *Constructions and Frames*, 11(1), 79–106.

Szmrecsanyi, B. (2006). *Morphosyntactic Persistence in Spoken English*. Berlin & New York: De Gruyter Mouton.

Thompson, S. A. & Koide, Y. (1987). Iconicity and 'indirect objects' in English. *Journal of Pragmatics*, 11(3), 399–406.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

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

Wulff, S. & Gries, S. Th. (2011). Corpus-driven methods for assessing accuracy in learner production. In P. Robinson, ed., *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Amsterdam & Philadelphia: John Benjamins, pp. 61–87.

Wulff, S. & Gries, S. Th. (2019). Particle placement in learner English: Measuring effects of context, first language, and individual variation. *Language Learning*, 69(4), 873–910.

Wulff, S. & Gries, S. Th. (2021). Explaining individual variation in learner corpus research: some methodological suggestions. In B. Le Bruyn & M. Paquot, eds., *Learner Corpora and Second Language Acquisition Research*. Cambridge: Cambridge University Press, pp. 191–213.