Corpus linguistics meets historical linguistics and Construction Grammar:
How far have we come, and where do we go from here?

Abstract

This paper aims to give an overview of corpus-based research that investigates processes of language change from the theoretical perspective of Construction Grammar. Starting in the early 2000s, a dynamic community of researchers has come together in order to contribute to this effort. Among the different lines of work that have characterized this enterprise, this paper discusses the respective roles of qualitative approaches, diachronic collostructional analysis, multivariate techniques, distributional semantic models, and analyses of network structure. The paper tries to contextualize these approaches and to offer pointers for future research.

1 Introduction

When the first issue of *Corpus Linguistics and Linguistic Theory* appeared in 2005, few people would have predicted that twenty years later, there would be a sizable community of researchers using corpus-linguistic methods in order to investigate processes of language change from the theoretical perspective of Construction Grammar. Today, diachronic corpus work with a constructional outlook has become so prolific that it is not straight-forward to lay out a full and balanced overview of the field in a short paper like this one. Fortunately, the editors of this special issue have given me permission to offer my subjective views on the matter. I will try to cover the points that I see as central, but it goes without saying that there will be inadvertent omissions. There will be an overrepresentation of work on the history of English, which is the field that I am most familiar with, and there will also be opinions that are open to debate. I encourage readers to let myself and others know where and how they disagree with my views, which notions deserve more attention than I have given them, and where there might be connections that I have failed to point out. I hope that these disagreements can lead to a fruitful exchange of ideas that will take the field further ahead, which is the overall goal of this paper.

To get the discussion started, why was it that the cross-pollination of research on corpus linguistics, historical linguistics, and Construction Grammar intensified in the early 2000s? With the benefit of hindsight, four key factors can be identified. First, there have been important pioneering studies that can be considered as practicing corpus-based Diachronic Construction Grammar *avant la lettre*. For example, Israel (1996) offers an analysis of the development of the English *way*-construction, using data from the Oxford English Dictionary. He observed that the *way*-construction changes over time in terms of its verbal collocates and that these developments reflect semantic change. This idea has influenced a large number of subsequent studies of constructions and their diachronic changes. In a study with a similar methodological approach, the Oxford English Dictionary was used as a corpus by Anshen and Aronoff (1999), who investigated changes in the productivity of the suffixes *-ment* and *-ity*, and who explicitly tied observations from historical corpus data to cognitive processes in the mental lexicon. Change in the productivity of constructions has become a central line of research in Diachronic Construction Grammar (Barðdal 2008, Perek 2016, 2018). The link between diachronic corpora, constructions, and cognition is also apparent in Tabor (1995), who discussed English *be going to* as a construction that exhibits prototype structure. On the basis of a corpus of literary works from different historical periods, Tabor (1995: 242) argued that the changes that can be observed in the *be going to* construction correspond to extrapolations along salient dimensions of the constructional prototype. He further noted (1995: 246) that subtle changes in relative frequency may pave the way for category change, which is an idea that is echoed by many current studies in Diachronic Construction Grammar. Yet another contribution that can be seen as a forerunner of Diachronic Construction Grammar is Verhagen's (2002) discussion of the Dutch *way*-construction and its development, which prefigures current theorizing about multiple inheritance in constructionalization (Trousdale 2013, Torrent 2015). Examples like these illustrate that certain ideas about constructions and their development informed discussions that were underway, and even if there was not yet an explicit vocabulary and a common framework to connect these ideas, the ground for Diachronic Construction Grammar had been prepared.

A second key factor has been that the early 2000s were an unusually productive period for the development of new corpus-based techniques. This took place in different research centers with cognitive and constructional orientations. A central figure is Dirk

Geeraerts, who established the QLVL (Quantitative Lexicology and Variational Linguistics) research group at the KU Leuven. The group pioneered various quantitative techniques of corpus analysis, including regression techniques (Grondelaers et al. 2002, Tummers et al. 2004), multi-dimensional scaling and clustering (Speelman et al. 2003), and more recently distributional semantic methods (Peirsman et al. 2008, Heylen et al. 2015). Besides firmly establishing the variationist brand of cognitive corpus linguistics that continues to be their trademark, the QLVL group has been very effective in inspiring others to take up new methods, including researchers with an interest in constructions and diachrony. The early 2000s also saw the development of collostructional analysis (Stefanowitsch and Gries 2003, Gries and Stefanowitsch 2004), which applied the corpus-linguistic concept of collocational analysis to the study of associations between constructions and lexical elements, thereby building a conceptual bridge from corpus linguistics to Construction Grammar. This connection brought together a community of researchers who were coming from different backgrounds, but who shared the same interests. In other influential work, Stefan Gries has promoted the use of multivariate techniques in corpus analyses (Gries 1999, 2003) that anticipated the later popularity of methods such as binary logistic regression. Further work that is relevant in this context is research in the British tradition of corpus linguistics. For example, Hunston and Francis (2000) offer a corpus-based view on grammar that emphasizes the notion of lexico-grammatical patterns, and which is thus highly compatible with constructional approaches. Similarly, Stubbs (2002) approaches phraseology from a perspective that includes collocational measures and that is close in spirit to collostructional analysis. The list of examples in this paragraph could easily be extended, the essential point is that the early 2000s in corpus linguistics laid important methodological foundations, which were then ready to support new ideas for further developments.

The third key factor that merits discussion in this context is that by the 2000s, a strong corpus-linguistic tradition with a focus on change in grammatical constructions had established itself, especially in grammaticalization studies. Groundbreaking work in this area includes Brinton (1996) on pragmatic markers, Krug (2000) on modals and semi-modal constructions, Mair (2002) on verb complementation, and Fanego (2004) on verbal gerunds, but the literature is much too vast to give an accurate summary here. Lindquist and Mair (2004) is a volume that captures the pulse of the time, Mair (2011) offers an insightful overview. As pointed out by Noël (2007), many studies of this kind informally referred to

constructions in order to describe the phenomena under investigation, and constructions famously found their way into the definition of grammaticalization that Hopper and Traugott (2003) offer in the second edition of their textbook on the subject. What further increased the mutual attraction of corpus-based research into grammaticalization and work on Construction Grammar was a broadly shared commitment to cognitive-functional principles (Noël 2007: 178), such as the importance of iconicity, schematization, entrenchment, and categorization. The community of researchers working on these topics thus provided the fledgling field of Diachronic Construction Grammar with ideas, empirical insights, and not least, a large forum for discussion.

Without the fourth and final key factor, it is unlikely that Diachronic Construction Grammar would have developed in the way it has. That last piece of the puzzle is of course the increased availability of large diachronic corpora. Resources such as the Helsinki corpus (Kytö 1991), the ARCHER corpus (Biber et al. 1994) were complemented by the Penn Parsed Corpora of Historical English (Kroch and Taylor 2000, Kroch et al. 2004), the Corpus of Late Modern English Texts (De Smet 2005), the Parsed Corpus of Early English Correspondence (Nurmi et al. 2006), the Corpus of Historical American English (Davies 2010), and the Old Bailey Corpus (Huber et al. 2016), amongst many others. Davidse and De Smet (2021) offer a comprehensive discussion of diachronic corpora that puts these developments into context. The increased size of newer diachronic corpora allowed researchers to study phenomena that were less frequent, and it became possible to use quantitative methods that were more demanding in terms of the amounts of data that were required. This gave the field the opportunity to ask new questions and to investigate them in new ways, which turned out to have a stimulating effect.

With the four key factors in place, it is fair to say that corpus-based, constructional research was an idea whose time had come. Importantly, this idea was not associated with any single leading figure, but it was a shared effort, driven by a community. Given that the researchers who participated in the development of the field came from different theoretical and methodological backgrounds, it is no wonder that lively discussions emerged around central concepts, such as for example the notions of *constructionalization* and *constructional change* (Traugott and Trousdale 2013). With regard to the corpus-based methods that have been proposed for diachronic constructional analyses, the field continues to be characterized by variety rather than homogeneity, which can be considered a strength.

The following sections of this paper aim to capture some of that variety by surveying different lines of research that have left their mark on the field and that continue to be developed. Section 2 discusses corpus-based work that is qualitative in nature. Section 3 turns to diachronic applications of collostructional analysis. Section 4 addresses multifactorial methods. The topic of Section 5 is the use of distributional semantic approaches. In Section 6, the focus shifts to analyses of network structure. The chapter concludes with Section 7, which takes stock and provides an outlook over possible future research.

2 Qualitative analyses

It is evident that linguistics in general has undergone a quantitative turn (Janda 2013, Kortmann 2021) and that the field has changed drastically with regard to its conceptualization and treatment of linguistic data. Quantitative techniques are a central element of current PhD training, and methodological literacy has become a necessary precondition for keeping up with many new developments in the field. That said, it is important to acknowledge the continued importance of qualitative approaches that use corpus data as a means to investigate theoretical issues. Quantitative approaches, powerful as they are, are specifically geared towards the analysis of research questions that find their answers in measurable characteristics of corpus data, such as frequency, dispersion, or structural complexity. As a consequence, researchers who are using a quantitative approach have to operationalize theoretical concepts in order to make them measurable. Qualitative approaches have a different set of limitations, but in principle they allow the formulation of questions that are only constrained by the researcher's capacity of imagination. This makes them highly valuable, especially to researchers who engage in theory-building.

This potential is illustrated by Trousdale (2012), who offers a discussion of theory and data in Diachronic Construction Grammar that is based on a corpus-based study of the English *what with* construction. The construction, which is exemplified in (1), consists of an absolute clause that is introduced by *what with* and that is accompanied by a main clause.

(1) a. What with the gown, the limos, and all the rest, you're probably looking at

about a hundred grand. (COCA)

b. What with the boyfriend coming back and all the confusion of the paramedics and neighbours, they couldn't find anything. (COCA)

c. The Deloche woman was going to have one heck of a time getting rid of the place, what with the economy the way it was in Florida. (COCA)

With regard to the form of the construction, the constituent introduced by *what with* may take the shape of a (typically coordinated) noun phrase, a participial clause, or a verbless small clause. Functionally, the construction conveys a reason for the proposition that is expressed in the main clause, and which tends to be negative. Trousdale (2012: 587) outlines the development of the construction on the basis of historical corpora and observes two major changes. As for the first, Middle English data suggests the existence of a more general construction in which *what* paired with several different prepositions in addition to *with*, for example *for*, *from*, and *through*. Over time, the construction specialized so that *with* gradually ousted its competitors. The collocation *what with* acquired the status of a holistic, entrenched unit and thus decreased in compositionality. The second change concerns the syntactic complement types that follow *what with*. Historically, the most frequent complement type was a coordinated noun phrase (*what with NP and NP*). In Late Modern English, examples of the *what with* construction with nominal complements still greatly outnumber those with clausal complements, but clausal complement types are increasing, giving rise to the syntactic variety that can be observed in Present-Day American and British English. Trousdale (2012: 596) shows that even examples with finite clausal complements are sporadically attested. This increase in syntactic productivity is accompanied by semantic broadening. The construction has shed some of its negative semantic prosody, so that positive propositions in the main clause are regularly found. Regarding the connection between theory and data, Trousdale (2012: 598) argues that a corpus-based investigation of piecemeal diachronic changes in the semantic and syntactic behavior of constructions can broaden our understanding of the processes that are involved in constructionalization.

The qualitative analysis of corpus data for the purpose of developing theories of language change is characteristic for the influential work done by Elizabeth C. Traugott. In a recent contribution, Traugott (2020) addresses pragmatic markers such as *by the way*, which speakers use for the management of discourse and specifically the expression of how an

utterance relates to the preceding conversation. The marker *by the way* can thus serve to flag a digression from the current topic under discussion. Traugott (2020: 3) is interested in the relation of these markers to the concept of stance, which comprises the speaker's epistemic and evaluative positioning towards an object as well as the speaker's alignment with the hearer. Using corpora from different historical periods of English, Traugott is able to identify early uses of the pragmatic markers that shed light on their functional development (2020: 5). In early data, *by the way* commonly collocates with verbs such as *say* or *mention*, which invites the interpretation that a digressive comment is made during the course of an argument. In Late Modern English, uses of *by the way* in relative clauses pair the digressive meaning of the marker with a syntactic structure that is in itself parenthetical to the main clause, so that the function of the marker is motivated by iconicity. From 1650 onward, *by the way* regularly appears in clause-initial position and with a meaning that indicates a topical shift. In the 19th century, *by the way* increasingly occurs in contexts that encode sensitive and negatively evaluated topics. In present-day data, Traugott (2020: 9) further observes a particularity of uses in which *oh* precedes *by the way*. In examples such as the ones shown in (2), *oh by the way* introduces enacted speech of a third party that the speaker portrays as inappropriate.

(2) a. The whole affair was carried off in the most offhand manner, as if the Senate were remarking: "Oh, by the way, of course we deserve more pay." (TIME)

b. But boom, she drops the hammer on him and says, oh, by the way, I'm in love with my dance partner. (COCA)

c. You don't just dump these (seniors) and say, 'Oh, by the way, you've got to be out'... This is absolutely just tragic. (COCA)

These uses illustrate how pragmatic markers can acquire meanings that capture the stance of the speaker. The semantic extension from a topical shift to a critical stance is motivated in the following way. Since *by the way* conventionally marks a casual digression, using it to introduce bad news implies open disregard for the hearer and is thus strongly face-threatening. A speaker who broaches a sensitive topic without due consideration for the hearer can be held accountable for that behavior. It is this critical stance that has become conventionally associated with the form *oh by the way*. Traugott's (2020) analysis manages

to identify individual corpus examples that illustrate relevant steps in the semantic and syntactic development of *by the way*, and her observations are put into the service of general hypotheses concerning the evolution of pragmatic markers in the constructional network.

To sum up this section, qualitative approaches represent an essential part of corpus-based Diachronic Construction Grammar. The observations that these approaches provide typically lay the groundwork for subsequent quantitative testing, which may then assess whether a reliable generalization can be made. It can be considered a strength of the field that qualitative and quantitative techniques are often applied side by side in the same studies, and that there is a mutual recognition of the advantages that each of the two approaches can provide (Traugott and Trousdale 2013: 238).

3 Diachronic collostructional analysis

As was mentioned in the introductory section of this paper, the development of collostructional analysis (Stefanowitsch and Gries 2003, Gries and Stefanowitsch 2004) had a particularly strong impact on research in Construction Grammar. Whereas collostructional analysis was primarily conceived for the purpose of analyzing and contrasting constructions in the synchronic use of a given variety of a language, the potential for various extensions was immediately apparent, so that researchers used it to analyze L2 learner data (Gries and Wulff 2005), to compare constructions across varieties (Wulff et al. 2007), and to link it to behavioral data from psycholinguistic experiments (Wiechmann 2008). Connecting collostructional analysis with the analysis of language change (Hilpert 2006) was another piece of the puzzle. In particular the logic of distinctive collexeme analysis (Gries and Stefanowitsch 2004), which contrasts the collocational preferences of two or more constructions, was easy to extend to contrasts of the same construction across different historical time periods. In Hilpert (2008), diachronic corpora of Dutch, English, German, and Swedish were used in order to study collocational shifts in constructions with future time reference. The results indicated that these constructions show developments in their collocational behavior that can be linked to ongoing grammaticalization. Hilpert (2012) offers a discussion of two factors that may confound the results of a diachronic distinctive

collexeme analysis. The first concerns the periodization of the corpus data that is used for the analysis, which can be alleviated by partitioning the data with variability-based neighbor clustering (Gries and Hilpert 2008). The second factor is the possibility of sampling error due to the uneven representation of different text genres in corpus data. Comparisons between genre-specific changes can test whether a construction undergoes a general development or whether its use diversifies across different text types.

Diachronic collostructional analyses have been applied for a range of different purposes, but a central objective has been to better understand the process of collocational expansion. With this aim in mind, Coussé (2014) investigates collocational shifts in Dutch perfect constructions with auxiliary verbs of possession and existence. Between the 13th and the 18th century, the analyzed constructions incrementally widen their collocational profiles. This process does not unfold randomly, but rather in a way that suggests gradual expansion through related verb classes that are increasingly removed from a prototype, so that the overall process creates a radial network.

Diachronic distinctive collexeme analysis has also been applied in studies that track the diachronic development of syntactic alternations. Li et al. (2023) offer a study that addresses four Mandarin Chinese constructions in which the theme and the recipient alternate in their respective positions, as in the English dative alternation. The study is based on corpus data from the 14th century up to the 20th century. Li et al. (2023) first use multivariate techniques in order to determine how the conditioning factors of the alternation changed over time, finding that end weight remains highly stable throughout, compared to shifting effects of definiteness and animacy. The collostructional analysis complements these findings by revealing construction-specific collocational shifts. The constructions that participate in the alternation thus follow their own developmental paths. Li et al. (2023: 28) further observe a transition from prototypical senses of transfer (‘give’, ‘borrow’) to a more diverse set of meanings.

A third context in which diachronic distinctive collexeme analysis are commonly used is the study of grammaticalization. Garachana and Sansiñena (2023) use the CORDE (Corpus diacrónico del español) in order to compare the Spanish auxiliary verbs *dejar* ‘cease’ and *parar* ‘stop’ with regard to their productivity and their collocational behavior, aiming to find out how strongly the two auxiliaries are grammaticalized. The analysis indicates that *parar* is more productive than *dejar* throughout the entire period under investigation. The

collostructional analysis further shows that *dejar* is attracted to stative verbs, and that from the 17th century onward, it develops an increasing preference for a few formulaic combinations that encode pragmatic meanings (Garachana and Sansiñena 2023: 19). By contrast, *parar* exhibits collocational expansion and thus shows the exact opposite tendency. Garachana and Sansiñena (2023: 20) conclude that *parar* functions as a fully grammaticalized aspectual marker, whereas *dejar* evolved along a cline of pragmaticalization and thus retreated into a niche of specialized meaning.

It is probably no coincidence that the studies by Li et al. (2023) and Garachana and Sansiñena (2023) both use diachronic distinctive collexeme analysis as a complement to other quantitative methods. In both cases, the analysis begins with an examination of general factors, such as end weight, productivity, or the role of a category such as definiteness. The collostructional analysis is then used to provide a more fine-grained view on the changing constructional meanings that are at play. Future applications of diachronic distinctive collexeme analysis stand to benefit from current suggestions made by Gries (2023) that provide collostructional analysis with conceptual and technical updates, such as added confidence intervals, the inclusion of dispersion measures, and a consideration of polysemy in the senses of words and constructions.

4 Multivariate techniques

The increasing popularity of multivariate statistical methods can be identified as one of the major trends that has shaped corpus linguistics in the last twenty years. What these methods provide is a way to examine how the use of a linguistic structure is influenced by a set of conditioning factors. These factors may pertain to the sound, morpho-syntactic form, and meaning of the structure, as well as to extra-linguistic variables that relate to the speaker, the situational context, or textual characteristics. Multivariate techniques can distinguish the factors that have an impact from those that do not, and the former can be ranked in terms of how strong their respective impacts are. This makes these methods especially useful for the analysis of phenomena that are characterized by probabilistic variation.

Whereas multivariate techniques such as binary logistic regression had been a standard tool in variationist sociolinguistics for a long time (Cedergren and Sankoff 1974), and whereas Biber (1988, 1995) had pioneered multidimensional analysis as a way of investigating variation between text types, the application of multivariate methods to the analysis of constructions and how speakers choose between them only emerged in corpus linguistics in the 1990s. Leech et al. (1994) offer an early study on variation in English genitive constructions, Gries (1999) examines the factors that condition English particle placement. The availability of historical corpora made it possible to apply multivariate techniques in diachronic analyses. The BROWN corpora were used by Hinrichs and Szmrecsanyi (2007) in a study of English *s*-genitives and *of*-genitives. Gries and Hilpert (2010) study change in the third-person present tense in Early Modern English. Wolk et al. (2013) investigate how dative and genitive constructions change in Late Modern English. All of these studies pay particular attention to interaction effects between the passage of time and other factors that influence speaker behavior. If an analysis reveals an interaction effect between time and a second variable, this means that the impact of the latter changes over time, so that it either becomes stronger or weaker. New conditioning factors may emerge, and older ones may become obsolete.

Recent diachronic studies that employ multivariate techniques have focused on issues that relate to different theoretical aspects of Construction Grammar, asking for example where in the network of constructions a given change is happening. To illustrate, Zehentner (2024) investigates variation in English time adjuncts. As is shown in (3), speakers may choose between adjuncts that have the form of either a nominal or a prepositional phrase.

(3) a. That day I saw them for three hours. (COCA)
b. On that day, I did not go with them. (COCA)

From Middle English onward, the nominal variant has faced increasing competition from the nominal variant. Zehentner (2024: 2) examines the variables that have been at play in that process between Middle and Late Modern English. The analysis includes lexical preferences of fourteen different time nouns, the position of the time adjunct relative to the verb, morphosyntactic complexity of the adjunct, its length in words, the distance between the

adjunct and the main predicate, and collocational strength between the time noun and the main predicate. Zehentner (2024: 2) hypothesizes that the prepositional variant, which offers speakers a more explicit and transparent option, will be preferred in contexts that involve greater cognitive complexity. The results indicate significant main effects for position, length, and distance, as well as significant interaction effects between these three factors and time. Specifically, it is observed that the prepositional variant favors the pre-verbal position and that this preference strengthens over time. The analysis also reveals that a greater distance between the time adjunct and the verb increases the likelihood of the prepositional variant. This effect however wanes over time. It is further shown that longer time adjuncts preferentially occur in the nominal variant, but also this effect gradually weakens. Importantly, Zehentner (2024: 7) observes that the impact of lexical preferences outweighs the effects of the other variables. Whereas time nouns such as *time* and *winter* favor the prepositional variant, *day*, *night*, and *week* exhibit the opposite preference. Zehentner (2024: 8) thus cautions against interpreting the observed variation in terms of an alternation between two abstract constructions. Instead, she argues for the importance of low-level representations that capture lexically specific constructions.

Besides the importance of lexical specificities, multivariate studies have also addressed the issue of individual variation in the use of constructions. Adopting this perspective, Fonteyn and Nini (2020) study variation in the use of English gerund constructions with the help of conditional inference trees. The examples in (4), taken from the EMMA corpus (Petré et al. 2019), illustrate uses of gerunds that are complemented by either prepositional phrases (*doing of some miracle*) or nominals (*doing miracles*).

(4) a. … the temptation was hot upon me to try if I had Faith by doing of some miracle

b. … to engage Men to obey those of God, By doing miracles so numerous and great

c. So that the leaving of the word, Altar, out of the Common Prayer booke last established, …

d. The leaving out one word in a Will, may marr the estate and disapoint all a mans hopes

Historically, the prepositional variant preceded the nominal variant, which gained ground from Middle English onward and spread to different syntactic environments. The variables that influence the choice between the two variants include the presence or absence of a determiner, the grammatical relation that is instantiated by the gerund (subject, object, subject complement, prepositional complement), the type of verb (possessive *have*, light verbs, lexical verbs), as well as language-external factors that pertain to the age and the generation of the writer as well as the textual register. The analysis shows that determiner usage is by far the most influential variable, followed by individual differences between writers and the grammatical relation of the gerund (Fonteyn and Nini 2020: 295). The remaining factors show only minor effects. Conditional inference tree analyses of individual authors reveal that the same conditioning factors may be organized in different ways that are specific to the writer in question and that override extra-linguistic factors such as age or the generation of the writer. Based on this finding, Fonteyn and Nini (2020: 303) argue that the writers that are represented in the EMMA corpus have idiosyncratic mental representations of the variation that affects gerund constructions in Early Modern English.

The studies by Zehentner (2024) and Fonteyn and Nini (2020) illustrate how multivariate methods serve to identify generalizations and, at the same time, pockets of variability that resist explanations in terms of broad tendencies. For Construction Grammar, a precise understanding of the dynamics between generalizations and idiosyncrasies allows researchers to pinpoint what phenomena are to be seen as constructions and at what level of abstraction these constructions are mentally represented by the speakers. With regard to diachrony, multivariate methods have the potential of examining to what extent factors such as structural complexity, end weight, or priming, which are thought to relate to general cognitive processes, are actually subject to change and thus affect speaker behavior differently at different points in time. In summary, multivariate techniques have brought Diachronic Construction Grammar both empirical depth and precise tools for crafting theoretical arguments. Going forward, an exciting prospect for constructional research will be to adopt proposals by Winter and Wieling (2016), who advocate for the use of mixed models, Growth Curve Analysis and Generalized Additive Modeling in analyses of language change.

5 Distributional semantic methods

Another quantitative approach that has had a profound impact on diachronic, corpus-based research on constructions in recent years is the use of distributional semantic methods. The fundamental idea that unites these methods is that linguistic units, such as words or constructions, can be characterized in terms of semantic vector representations. In the literature, count models are distinguished from predictive models (Baroni et al. 2014). In the former, semantic vectors capture the frequencies of linguistic elements that co-occur with a target word in corpus data. In simple terms, a word such as the English noun *sail* is often found close to words such as *boat*, *mast*, and *sea*. Preferences of this kind are represented by a semantic vector that contains a large set of collocates of *sail* and that provides transformed frequency measures that reflect degrees of mutual attraction between *sail* and these collocates. In a semantic vector for *sail*, the collocates *boat*, *mast*, and *sea* would have high values, whereas semantically unrelated collocates such as *desert*, *microphone*, or *roof* would have low values. Comparisons between semantic vectors allow assessments of how similar two linguistic units are with regard to their meanings. For example, the vectors of the words *sail* and *keel* would be relatively similar, while those of *sail* and *violin* would be very different. Count models and predictive models both use semantic vectors, but they differ with regard to how these vectors are created. In predictive models, machine learning techniques are used in order to construct semantic vectors that are maximally predictive of a target word. Both of the two types of models are used in current research, but predictive models have been shown to outperform count models in many contexts (Baroni et al. 2014: 238), so that they have become more popular.

Diachronic applications of distributional semantic methods provide an empirical perspective on meaning change. Shifts in the meaning of a linguistic unit are analyzed through comparisons between semantic vector representations of the same unit at different points in time. For example, if the collocates of the word *sail* are analyzed in a diachronic corpus, present-day data would reveal newly emerging collocations such as *solar sail*, which reflects a recent meaning extension. Studies that have paved the way for current applications of distributional semantic methods in research on language change include Sagi et al. (2011), who discuss semantic narrowing and broadening, Jenset (2013), who tracks the semantic development of existential *there* in English, and Mitra et al. (2014), who identify

the birth of new word senses. Hamilton et al. (2016) observe that polysemous words have a higher potential for meaning change, Perek (2016) offers a study of how productivity changes over time. A comprehensive overview of the current state of the art is provided by Fonteyn et al. (2022), who also discuss the distinction between type-based and token-based models. In a type-based model, semantic vectors represent the meaning of a target word on the basis of many instances of use. If the meaning of the word *sail* is captured by a semantic vector that is based on all occurrences of the word in a given corpus, the vector represents a word type, rather than an individual instance. By contrast, token-based models (Heylen et al. 2015) allow comparisons between uses of the same word across different instances. To stay with the example of the English noun *sail*, the OED distinguishes between the basic sense of a 'textile canvas', as in *a white sail*, and a metonymically derived sense that conveys 'an act of sailing', as in *a three-day sail*. The two senses exhibit differences in their respective collocational preferences. To capture these differences, token-based models construct semantic vector representations for individual concordance lines. This is accomplished by averaging semantic vectors of the words that appear in a given concordance line. Token-based models can be used for tasks such as word sense disambiguation on the basis of synchronic corpus data. In studies that address diachronic issues, token-based models can determine how polysemy networks evolve, or how uses of grammaticalized forms differ from their lexical counterparts (Fonteyn 2020, Hilpert and Correia Saavedra 2020), among other things.

Whereas most efforts in diachronic applications of distributional semantic methods have targeted change in lexical semantics, a growing number of studies focuses on grammatical constructions. For instance, Budts (2020) uses a predictive token-based model in order to address periphrastic *do* in Early Modern English. The model is trained on data from the EEBO corpus. Specifically, it uses concordance lines with forms of the verb *do* as well as present and past tense modal verbs such as *can*, *will* and *would*. The training enables the model to distinguish between concordance lines that would be a probable context for forms of *do* and those in which a different verb would be expected. The analysis reveals that highly predictable uses of *do* in the 16th century are associated with the expression of universality or habituality. Besides these prototypical instances of periphrastic *do*, the analysis reveals other uses that are harder to predict. Budts (2020: 354) identifies concordance lines for which the analysis assigns similar probabilities to *do* and a competing

modal auxiliary such as *can*. This is illustrated by the examples in (5), which convey the meaning of general impossibility.

(5) a. Our bodies and souls [do / can] not make vs members of Christ, but our faith and obedience.

b. The good tree [doth / can] not bear ill fruit.

c. They which [do / can] not valiantly expose themselves to dangers, become slaves to those which assail them.

The analysis indicates that *do* and *can* are both likely candidates for these contexts. The existence of such contexts suggests that speakers of Early Modern English would have judged *do* as a potential paradigmatic alternative to *can* in certain contexts. Budts (2020: 358) identifies similar areas of distributional overlap for *do* and *may*, *do* and *shall*, and *do* and *will*. For each of these pairings, there are contexts that are equally predictive of *do* and a corresponding modal auxiliary. This motivates the hypothesis that the development of periphrastic *do* was at least in part driven by analogy with existing modal auxiliaries. Issues such as gradience between different categories have received particular attention in foundational research in Construction Grammar, and the study by Budts (2020) illustrates that current research still maintains a special sensitivity for such phenomena. Distributional semantic methods can thus inform key theoretical issues in Construction Grammar on the basis of a new empirical perspective.

Gradient changes in category status are also addressed by Desagulier (2022), who investigates the diachronic development of complex prepositions in English, focusing on *in the middle of*, *in/at the center of*, *in/at the heart of*, and *in the midst of*. In the existing literature on complex prepositions, different views exist on how the syntactic constituency of these elements is best modeled, and whether these elements belong to a common category. Desagulier (2022) aims to find out whether the four patterns that are studied can be viewed as constructions and, in a second step, how they evolve over time in relation to one another. Data from the COHA is used to retrieve the types of landmark nouns that co-occur with each of the four patterns (i.e. *in the middle of the road*, *city*, *night*, etc.). Measurements of type frequency and the number of hapax legomena reveal productive usage and thus indicate that the patterns under study are to be seen as constructions. In

order to assess how the four constructions developed semantically, Desagulier (2022: 353) turns to distributional semantic methods, specifically a predictive, type-based model, and creates a semantic vector space based on all landmark nouns that are attested with the constructions. The semantic vector space is projected onto a two-dimensional graph that visualizes similarities in the meanings of different landmark nouns (Desagulier 2022: 356). To illustrate, the graph shows clusters such as natural expanses (*field*, *forest*, *desert*), urban areas (*street*, *block*, *plaza*), and communities (*family*, *nation*, *society*), among others. Desagulier (2022: 359) then investigates which areas in the semantic vector space are associated with each of the four constructions, and how these associations change over time. The analysis reveals several semantic developments. For example, in the early decades of the COHA, *in the heart of* exhibits a collocational preference for geographic and geopolitical entities such as *country* or *city*. Over time, this preference broadens towards abstract entities such as *issue*, *conflict*, or *problem*. Desagulier (2022: 360) notes that during this time, the schematization of *in the heart of* does not impinge on the meaning potential of *in the midst of*, so that there is a division of labor between the two. The results of the distributional semantic analysis allow Desagulier (2022) to model the respective semantic profiles of the four constructions. The analysis reveals where and to what extent the constructions overlap, and how their meaning potentials expand and contract over time.

As these examples illustrate, distributional semantic methods hold the promise of exciting new developments for Diachronic Construction Grammar. There are, of course, several challenges that should be mentioned (Fonteyn et al. 2022: 28). Distributional methods are data-intensive, and for many lesser-documented languages, the existing corpus resources may prove to be insufficient in size. Another problem relates to the wide variety of methods that are available and to the number of decisions that enter into the construction of semantic vector spaces. Researchers need to keep a close watch on what goes into an analysis in order to assure the replicability of studies, the interpretability of results, and comparability between different analyses. Also, and as is always the case in historical linguistics, a caveat on the representativity of the data is necessary. The results allow statements about the data that goes into an analysis, extrapolations need to be viewed with caution. If the field manages to converge on a set of good practices, that will assure that these difficulties are kept in check.

6 Analyzing change in constructional networks

A defining characteristic of Construction Grammar is that it adopts a network model of language. In recent years, the field has increasingly focused on questions that concern the architecture of the constructional network and the different types of connections that give shape to the network. Comprehensive accounts are offered by Diessel (2019) and Schmid (2020). With regard to diachrony, a growing body of research has tried to analyze processes of language change in terms of changes in the nodes and connections that make up the constructional network (Barðdal et al. 2015, Sommerer and Smirnova 2020, Hilpert 2021, Noël and Colleman 2021, Herbst and Huber 2022). This section reviews some of the ideas that have come to the fore in this context.

A central notion in this area of research is multiple inheritance, that is, the idea that a specific construction may exhibit features of several more abstract constructions (Hilpert 2019: 63). Trousdale (2013) discusses the relationship between multiple inheritance and constructionalization on the basis of expressions such as *give someone a talking to* 'reprimand'. The expression combines ditransitive syntax with a gerund as the theme object and thus inherits information from both the ditransitive construction and the gerund construction. Data from the OED shows that combinations of *give* with different gerunds have been productively used to convey meanings of verbal or physical assaults by the 19th century. How do inheritance and change relate to one another? Trousdale (2013: 501) argues that language users adopt inheritance as a default strategy. Utterances are thus interpreted in terms of the general constructions that they instantiate, unless there is a reason to assume that the speaker has a different meaning in mind. This is arguably the case in examples such as *give someone a talking to*, which deviate from canonical syntax in that the theme object is not nominal, and which invite the interpretation that the communication that took place involved criticism. In other words, the interpretation of the whole construction goes beyond the meaning that can be inferred through inheritance. If such usage events occur repeatedly, a new form-meaning pair conventionalizes. Trousdale (2013: 511) makes the point that new constructions often emerge through the re-use of existing parts, which has been echoed in subsequent studies on multiple inheritance and language change (Fanego 2015, Van de Velde et al. 2015, Hoffmann and Trousdale 2022).

If language is modeled in terms of a constructional network, it becomes possible to account for phenomena that would be difficult to explain in other ways. For example, Pijpops and Van de Velde (2016) discuss an effect that they call constructional contamination. The term describes a relation that can obtain between two constructions that resemble one another with regard to their morphosyntax. In such a constellation, usage frequencies of one construction may influence patterns of variation in the other one. Pijpops and Van de Velde (2016: 544) illustrate this on the basis of Dutch partitive genitive constructions such as the one shown in (6).

(6) in begin van de week iets verkeerd/s gegeten
[in beginning of the week]$_{PP}$ [something wrong]$_{NP}$ eaten
'I ate something wrong at the start of the week.'

The construction varies with regard to the presence or absence of an inflectional suffix *-s* that attaches to the adjective. Speakers may either use the s-less variant *iets verkeerd* or the suffixed alternative *iets verkeerds*. The variation is subject to various conditioning factors, including usage frequencies of a construction that is superficially isomorphic to the partitive genitive. Example (7) illustrates that construction.

(7) dat iets verkeerd geïnterpreteerd wordt?
that [something]$_{NP}$ [wrongly]$_{AdvP}$ interpreted gets
'...that something gets wrongly interpreted?'

In contrast to the partitive genitive, this construction does not allow the addition of the suffix. Pijpops and Van de Velde (2016) demonstrate that high usage frequency of the contaminating construction influences how speakers use the partitive genitive, so that the *s*-less variant receives a boost. The effect, which is confirmed for other constructions of Dutch (Pijpops et al. 2018) and English (Hilpert and Flach 2022), provides evidence for associative links between formally different constructions, and thus offers insights into the organization of the constructional network. Constructional contamination effects also have implications for Diachronic Construction Grammar. For example, Bouso (2022) uses corpus data to show that the development of the English Reaction Object Construction (e.g. *She nodded*

agreement) in the 19th century is impacted by constructions that exhibit similar meanings and forms (e.g. *She nodded in agreement*, *She gave a nod of agreement*). Overlap in the collocational profiles of these constructions and similarities in their frequency developments support the hypothesis that constructional contamination was at play.

Research in Diachronic Construction Grammar has also aimed to uncover generalizations about the ways in which the constructional network changes over time. For example, Torrent (2015) has analyzed changing infinitive constructions in Brazilian Portuguese. On the basis of his results, he works out the conditions that need to be met for new links to form in the constructional network. He further explores how the emergence of a new node in the constructional network causes downstream effects in which the network reorganizes itself. Drawing on these ideas, Sommerer and Hofmann (2021) investigate network reconfiguration in the history of English, using the determinative *sum(e)* as their point of departure. In Old English, *sum(e)* alternated with *ān* in the marking of indefinite singular nouns. While it was eventually ousted from this context, it became the default choice for the marking of indefinite plural and mass nouns in late Middle English. Using the Penn Parsed Corpora of Historical English, Sommerer and Hofmann (2021: 12) show that during Middle and Early Modern English, *sum(e)* no longer competes for the marking of indefinite singular nouns. Instead, contexts without any overt marking are increasingly replaced by *ān*. Sommerer and Hofmann (2021: 14-15) hypothesize a strengthening association of *sum(e)* with singular mass nouns, but the data only reveals a minor tendency in this direction. By contrast, the data show a strong and significant increase of uses in which *sum(e)* is used as an article before a plural noun. This effect persists even when other conditioning factors such as the formality of the text type or the presence of modification is taken into account. These developments can be interpreted as changes that affect the nodes and connections in a constructional network. Sommerer and Hofmann (2021: 23) argue that during Old English, an abstract definite noun phrase schema emerged that required the overt marking of nominals. In Middle English, this requirement started to affect also indefinite noun phrases, which entertained connections to the definite noun phrase construction and which were therefore subject to analogical pressure. Within the emerging node of an indefinite noun phrase construction, *ān* established itself in the context of singular count nouns, but it left niches such as the plural or singular mass nouns for potential

competitors, so that sum(e) could complement the English article paradigm (Sommerer and Hofmann 2021: 28).

Studies such as the ones reviewed in this section put the analysis of constructions and corpus data into the service of detecting broad generalizations about language change. Whereas constructional approaches still have a reputation of occasionally getting lost in the fine details of constructions and their idiosyncrasies, it is clear that the deeper goal of Diachronic Construction Grammar is the construction of a theoretical model that is based on general principles, grounded in domain-general cognitive processes, and that makes clear empirical predictions for unseen data. It is safe to say that the field has reached a critical mass of researchers who are interested in these issues, which is cause for some optimism regarding future insights.

7 Concluding remarks

It was the goal of this paper to offer a broad sketch of how corpus-based work and Diachronic Construction Grammar have merged into a productive and dynamic research program, and to speculate a little bit about future developments. In the introduction I had to warn readers that a full survey would be beyond my capabilities. There are several issues that I would have liked to discuss, such as for example language contact (Höder 2010), gesture (Schoonjans 2014), creativity (Hoffmann 2020), change over the lifespan (Antonissen 2019), or change in word formation (Hartmann 2016). I regret these omissions, but I encourage readers to seek out this work on their own. Likewise, the selective overview that I have given should be understood as offering various eclectic points of entry to the literature, a first orientation rather than a detailed map. What I can say with confidence is that once a reader decides to go down one of the rabbit holes that I have described, such as for example the use of distributional semantics in studies on language change, or the nature of changes in the constructional network, they will discover a wonderland of ideas, and many interesting characters. Newcomers from all areas of linguistics are always welcome, and a lot of interesting work remains to be done, so I hope to see you there soon!

References

Anshen, Frank and Mark Aronoff. 1999. Using dictionaries to study the mental lexicon. *Brain and Language* 68/1-2, 16-26.

Anthonissen. Lynn. 2019. Constructional change across the lifespan: The nominative and infinitive in early modern writers. In Kristin Bech and Ruth Möhlig-Falke (eds.), *Grammar – Discourse – Context: Grammar and Usage in Language Variation and Change*, 125-156. Berlin: De Gruyter.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Barðdal, Jóhanna, Elena Smirnova, Spike Gildea, and Lotte Sommerer (eds.) 2015. *Diachronic Construction Grammar*. Amsterdam: John Benjamins.

Baroni, Marco, Georgiana Dinu and Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd annual meeting of the association for computational linguistics* (volume 1: Long papers), 238–247.

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Biber, Douglas, Edward Finegan, and Dwight Atkinson 1994. ARCHER and its Challenges: Compiling and Exploring a Representative Corpus of Historical English Registers. In Udo Fries, Gunnel Tottie and Peter Schneider (eds.), *Creating and Using English Language Corpora*, 1-13. Amsterdam: Rodopi.

Brinton, Laurel J. 1996. *Pragmatic markers in English: grammaticalization and discourse functions*. Berlin: Mouton de Gruyter.

Bouso Rivas, Tamara. 2022. The English Reaction Object Construction: A Case of Syntactic Constructional Contamination. *Miscelánea: A Journal of English and American Studies* 65, 13–36.

Budts, Sara. 2020. A connectionist approach to analogy. On the modal meaning of periphrastic do in Early Modern English *Corpus Linguistics and Linguistic Theory* 18/2, 337-364.

Cedergren, Henrietta and David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50/2, 333-355.

Coussé, Evie. 2014. Lexical expansion in the HAVE and BE perfect in Dutch: A constructionist prototype account. *Diachronica* 31/2, 159-191.

Davidse, Kristin and Hendrik De Smet. 2020. Diachronic Corpora. In Magali Paquot and Stefan Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*, 211-233. Cham: Springer.

Davies, Mark. 2010. *The Corpus of Historical American English (COHA): 400+ Million Words, 1810–2009*, available online at http://corpus.byu.edu/coha.

Desagulier, Guillaume. 2022. Changes in the midst of a construction network: a diachronic construction grammar approach to complex prepositions denoting internal location. Cognitive Linguistics 33/2, 339-386.

De Smet, Hendrik. 2005. A corpus of Late Modern English Texts. *ICAME Journal* 29, 69–82.

Fanego, Teresa. 2004. On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica* 21/1, 5-55.

Fanego, Teresa. 2015. Multiple Sources in Language Change: The Role of Free Adjuncts and Absolutes in the Formation of English ACC-ing Gerundives. In Mikko Höglund, Paul Rickman, Juhani Rudanko and Jukka Havu (eds.), *Perspectives on Complementation: Structure, Variation and Boundaries*, 179-205. Basingstoke: Palgrave Macmillan.

Fonteyn, Lauren. 2020. What about grammar? Using BERT embeddings to explore functional-semantic shifts of semi-lexical and grammatical constructions. https://hdl.handle.net/1887/138612

Fonteyn, Lauren and Andrea Nini. 2020. Individuality in syntactic variation: An investigation of the seventeenth century gerund alternation. *Cognitive Linguistics* 31/2, 279-308.

Garachana, Mar and María Sol Sansiñena. 2023. Combinatorial Productivity of Spanish Verbal Periphrases as an Indicator of Their Degree of Grammaticalization. *Languages* 8/3: 187.

Gries, Stefan Th. 1999. Particle movement: A cognitive and functional approach. *Cognitive Linguistics* 10/2, 105-145.

Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: a study of Particle Placement*. London: Continuum Press

Gries, Stefan Th. 2023. Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics* 9/3, 351-386.

Gries, Stefan Th. and Martin Hilpert. 2008. The identification of stages in diachronic data: variability-based neighbor clustering. *Corpora* 3/1, 59-81.

Gries, Stefan T. and Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on "alternations". *International Journal of Corpus Linguistics* 9/1: 97-129.

Gries, Stefan T. and Stefanie Wulff. 2005. Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3, 182-200.

Grondelaers, Stefan, Dirk Speelman and Dirk Geeraerts. 2002. Regressing on 'er'. Statistical analysis of texts and language variation. In Anne Morin and Pascale Sébillot (eds.), *6ièmes Journées internationales d'Analyse statistique des Données Textuelles - 6th International Conference on Textual Data Statistical Analysis*, 335-346. Rennes: Institut National de Recherche en Informatique et en Automatique.

Hamilton, William L., Jure Leskovec and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *Proceedings of the 54th annual meeting of the association for computational linguistics*, 1489–1501.

Hartmann, Stefan. 2016. *Wortbildungswandel. Eine diachrone Studie zu deutschen Nominalisierungsmustern*. Berlin: De Gruyter.

Herbst, Thomas and Judith Huber. 2022. Diachronic Construction Grammar – Introductory Remarks to This Special Issue. *Zeitschrift für Anglistik und Amerikanistik* 70/3, 213-221.

Heylen, Kris, Thomas Wielfaert, Dirk Speelman and Dirk Geeraerts. 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua* 157, 153-172.

Hilpert, Martin 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory* 2/2, 243–256.

Hilpert, Martin. 2012. Diachronic collostructional analysis: How to use it and how to deal with confounding factors. In Kathryn Allan and Justyna A. Robinson (eds.), *Current Methods in Historical Semantics*, 133-160. Berlin: De Gruyter.

Hilpert, Martin. 2008. *Germanic Future Constructions. A Usage-based Approach to Language Change*. Amsterdam: John Benjamins.

Hilpert, Martin. 2019. *Construction Grammar and its Application to English*. 2nd ed. Edinburgh: Edinburgh University Press.

Hilpert, Martin. 2021. *Ten Lectures on Diachronic Construction Grammar*. Leiden: Brill.

Hilpert, Martin and Saavedra, David Correia. 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.

Hilpert, Martin and Susanne Flach. 2022. A case of constructional contamination in English: Modified noun phrases influence adverb placement in the passive. In Krawczak, Karolina, Barbara Lewandowska-Tomaszczyk and Marcin Grygiel (eds.), *Analogy and Contrast in Language: Perspectives from cognitive linguistics*. Amsterdam: John Benjamins.

Hinrichs, Lars and Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11/3, 437-474.

Höder, Steffen. 2010. *Sprachausbau im Sprachkontakt. Syntaktischer Wandel im Altschwedischen.* Heidelberg: Winter.

Hoffmann, Thomas. 2020. Construction grammar and creativity: Evolution, psychology, and cognitive science. *Cognitive Semiotics*, 13/1, 1-11.

Hoffmann, Thomas and Trousdale, Graeme. 2022. On Multiple Paths and Change in the Language Network. *Zeitschrift für Anglistik und Amerikanistik* 70/3, 359-382.

Hopper, Paul J. and Elizabeth C. Traugott. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.

Huber, Magnus, Magnus Nissel, and Karin Puga. 2016. Old Bailey Corpus 2.0. hdl:11858/00-246C-0000-0023-8CFB-2

Hunston, Susan and Gill Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Israel, Michael. 1996. The way constructions grow. In Adele Goldberg (ed.), *Conceptual Structure, Discourse and Language*, 217-230. Stanford: CSLI.

Janda, Laura A. 2013. *Cognitive Linguistics – The Quantitative Turn: The Essential Reader*, Berlin: De Gruyter.

Jenset, Gard B. 2013. Mapping meaning with distributional methods: A diachronic corpus-based study of existential *there*. *Journal of Historical Linguistics*, 3/2, 272-306.

Kortmann, Bernd. (2021) Reflecting on the quantitative turn in linguistics. *Linguistics* 59/5, 1207–1226.

Kroch, Anthony, and Taylor, Ann. 2000. Penn-Helsinki Parsed Corpus of Middle English, second edition.

Kroch, Anthony, Santorini, Beatrice, and Delfs, Lauren. 2004. The Penn-Helsinki Parsed Corpus of Early Modern English.

Krug, Manfred. 2000. *Emerging English modals: A corpus-based study of grammaticalization*. Berlin: Mouton de Gruyter.

Kytö, Merja. 1991. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*. 3rd ed. Helsinki: Department of English, University of Helsinki.

Leech, Geoffrey, Brian Francis, and Xunfeng Xu. 1994. The use of computer corpora in the textual demonstrability of gradience in linguistic categories. In Catherine Fuchs and Bernard Victorri (eds.), *Continuity in Linguistic Semantics*, 57-76. Amsterdam: John Benjamins.

Li, Yi, Szmrecsanyi, Benedikt and Zhang, Weiwei. 2023. Beyond dynasties and binary alternations: a diachronic corpus study of four-way variability in Chinese theme-recipient constructions. *Folia Linguistica*.

Lindquist Hans and Christian Mair (eds.) (2004). *Corpus Approaches to Grammaticalization in English*. Amsterdam: John Benjamins.

Mair, Christian. 2002. Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora. *English Language and Linguistics* 6/1, 105-131.

Mair, Christian. 2011. Grammaticalization and corpus linguistics. In Bernd Heine, and Heiko Narrog (eds), *The Oxford Handbook of Grammaticalization*, 239-250. Oxford: Oxford University Press.

Mitra, Sunny, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee and Pawan Goyal. 2014. That's sick dude!: Automatic identification of word sense change across different timescales. *Proceedings of the 52nd annual meeting of the association for computational linguistics* (volume 1: Long papers), 1020-1029.

Noël, Dirk. 2007. Diachronic construction grammar and grammaticalization theory. *Functions of Language* 14/2, 177–202.

Noël, Dirk and Timothy Colleman. 2021. Diachronic construction grammar. In Wen Xu and John R. Taylor (eds.), *The Routledge Handbook of Cognitive Linguistics*, 662-675. New York: Routledge.

Peirsman, Yves, Kris Heylen and Dirk Geeraerts. 2008. Size matters: Tight and loose context definitions in English word space models. In Marco Baroni, Stefan Evert and Alessandro Lenci (eds.), *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics. Bridging the gap between semantic theory and computational simulations*, 34-41. Hamburg: ESSLLI.

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54/1, 149-188.

Perek, Florent. 2018. Recent change in the productivity and schematicity of the *way*-construction: A distributional semantic analysis. *Corpus Linguistic and Linguistic Theory*, 14/1, 65–97.

Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing and Odile A.O. Strik. 2019. Early modern multiloquent authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME Journal* 43/1, 83-122.

Pijpops, Dirk, Isabeau De Smet and Freek Van de Velde. 2018. Constructional contamination in morphology and syntax. Four case studies. *Constructions and Frames* 10/2, 269-305.

Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50/2, 543-581.

Sagi, Eyal, Kaufmann, Stefan and Clark, Brady. 2011. Tracing semantic change with Latent Semantic Analysis. In Kathryn Allan and Justyna A. Robinson (eds.), *Current Methods in Historical Semantics*, 161-183. Berlin: De Gruyter.

Schoonjans, Steven. 2014. Is gesture subject to grammaticalization? *Papers of the Linguistic Society of Belgium* 8, 30-43.

Sommerer, Lotte and Klaus Hofmann. 2021. Constructional competition and network reconfiguration: investigating *sum(e)* in Old, Middle and Early Modern English. *English Language and Linguistics* 25/1, 1-33.

Sommerer, Lotte and Elena Smirnova (eds.) 2020. *Nodes and Networks in Diachronic Construction Grammar*. Amsterdam: John Benjamins.

Speelman, Dirk, Stefan Grondelaers and Dirk Geeraerts. 2003. Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37: 317-337.

Stefanowitsch, Anatol and Stefan T. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8/2: 209-43.

Stubbs, Michael. 2002. Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7/2, 215-244.

Tabor, Whitney. 1995. Lexical Change as Nonlinear Interpolation. *Proceedings of the 17th Cognitive Science Conference*, 242-247.

Torrent, Tiago. 2015. On the relation between inheritance and change: The Constructional Convergence and the Construction Network Reconfiguration Hypotheses. In Jóhanna Barðdal, Elena Smirnova, Spike Gildea, and Lotte Sommerer (eds.), *Diachronic Construction Grammar*, 175-214. Amsterdam: John Benjamins.

Traugott, Elizabeth Closs. 2020. Expressions of stance-to-text: discourse management markers as stance markers. *Language Sciences* 82, 1-14.

Traugott, Elizabeth C. and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Trousdale, Graeme. 2012. Theory and data in diachronic Construction Grammar: The case of the *what with* construction. *Studies in Language*, 36/3, 576-602.

Trousdale, Graeme. 2013. Multiple inheritance and constructional change. *Studies in Language* 37/3, 491-514.

Tummers, José, Dirk Speelman and Dirk Geeraerts. 2004. Quantifying semantic effects. The impact of lexical collocations on the inflectional variation of Dutch attributive adjectives. In Gérald Purnelle, Cédrick Fairon and Anne Dister (eds.), *Le poids des mots. Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles* 1079-1088. Louvain la Neuve: Presses Universitaires de Louvain.

Van De Velde, Freek, Hendrik De Smet, and Lobke Ghesquière (eds.). 2015. *On Multiple Source Constructions and Language Change*. Amsterdam: John Benjamins.

Verhagen, Arie. 2002. From parts to wholes and back again. *Cognitive Linguistics* 13/4, 403-439.

Wiechmann, Daniel. 2008. On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4/2, 253-290.

Winter, Bodo & Wieling, Martijn. 2016. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution* 1/1, 7-18.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach, and Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30/3, 382-419.

Wulff, Stefanie, Anatol Stefanowitsch, and Stefan T. Gries. 2007. Brutal Brits and persuasive Americans: Variety-specific meaning construction in the into-causative. In Günter Radden, Klaus-Michael Köpcke, Thomas Berg, and Peter Siemund (eds), *Aspects of Meaning Construction*, 265-281. Amsterdam: John Benjamins.

Zehentner, Eva. 2024. Alternations (at) that time: NP versus PP time adjuncts in the history of English. *Linguistic Vanguard*. https://doi.org/10.1515/lingvan-2023-0054