

Corpus-based Approaches to Constructional Change

Martin Hilpert

25.1. Introduction

The use of corpus data is in no way unique to work in Construction Grammar, let alone to the study of constructional change. Modern and historical corpora are used across the entire spectrum of linguistics, with very diverse methods and still more diverse goals (Lüdeling and Kytö 2008). The present chapter focuses on a growing body of research that uses the methods of historical corpus linguistics in a way that is consistent with a constructional approach to language change (Fried, this volume). Not all of this work explicitly aligns itself with a particular brand of Construction Grammar, but nonetheless an implicit commitment to the constructionist agenda is there: the focus of investigation lies on form-meaning pairings and developments with regard to their structure, function, and frequency. Showcasing some of this work, it is discussed (a) why diachronic corpora and similar textual resources such as the *Oxford English Dictionary* are a particularly apt source of data for the study of constructional change, (b) how these resources are used for that purpose, and (c) how the constructional perspective sets this work apart from other approaches to language change. The case studies that are chosen also illustrate the rich variety of processes that can be grouped together under the heading of constructional change, showing in each case how the phenomenon in question can be fruitfully investigated and how these investigations advance the theory of Construction Grammar.

In line with most current definitions of the term, constructions are understood here as symbolic pairings of a form and a meaning that display structural or semantic/pragmatic idiosyncrasies or, even without such idiosyncrasies, a high level of entrenchment (Croft 2001; Goldberg 2006a; Langacker 2005). This definition is deliberately very broad, including abstract syntactic patterns such as pseudo-clefts and the Ditransitive construction, concrete lexical items such as the English noun *dog*, and idiomatic

expressions such as *blow the whistle on NP*. While these forms differ in terms of how abstract and schematic they are, each of them connects a particular form with a particular meaning, thus forming a symbolic unit. A broad definition of the term ‘construction’ is needed because the sum total of linguistic knowledge is taken to be a hierarchically ordered network of symbolic units, in which more schematic constructions share some or all of their characteristics with more specific constructions (Broccias, this volume). Each linguistic form that is mentally stored by the speaker (because of either its idiosyncrasies or its frequency) represents a node in this constructional network. Diachronically, single nodes or groups of nodes in the network may change. How these processes of change can be studied on the basis of corpus data is the main concern of this chapter.

In language use, both form and meaning of a construction are subject to variation and change. Changes in form pertain to the phonology or morphosyntax of a construction, whereas meaning changes affect its semantic and discourse-pragmatic characteristics. Each of these types of change is amenable to corpus analysis. Changes in form and meaning can be studied through frequency measurements of variant forms. One variant of a construction may become more frequent over time, thus altering the prototype of that construction. Even in the absence of structural change, a construction may undergo changes in absolute frequency, relative frequency, or type frequency. These changes are no less indicative of development than changes in the substance of a construction. In order to establish that an observed trend constitutes reliable evidence, frequency measurements in quantitative corpus linguistics are submitted to inferential statistics. A number of techniques particularly geared toward the analysis of diachronic corpus data have been developed recently (Hilpert and Gries 2008 and references therein). It is important to point out that frequency measurements are not merely used to document differences between historical stages. The art of approaching constructional change with a corpus-based methodology is to formulate theoretical questions in such a way that frequencies drawn from corpus data can illuminate them. More specifically, hypotheses have to be constructed so that observed frequencies in corpora provide evidence for or against them.

Before going into a number of individual case studies, it is necessary to specify the notion of constructional change in some more detail. If the term ‘construction’ were to denote only grammatical constructions, the term ‘constructional change’ would be largely synonymous with grammaticalization, that is, “the change whereby lexical terms and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions” (Hopper and Traugott 2003: xv). However, constructional change is more encompassing. As outlined above, lexical items are also constructions, such that lexicalization (Brinton and Traugott 2005) and all manners of lexical semantic change fall under the purview of constructional change as well. In addition, many definitions of grammaticalization exclude word order changes (Hopper and Traugott 2003: 24), some of which instantiate constructional change. The diachronic branch of Construction Grammar is thus a broader enterprise than the study of grammaticalization, even though there is substantial overlap in both subject matter and theoretical orientation (Noël 2007a). Given the broad definition of constructions, it is fair to ask whether the notion of constructional change is not in fact coextensive with language change, which would render the term rather meaningless. This is not the case. Many processes of language change are not specific to individual constructions, and are hence best described as affecting a language or a language variety as a whole, across many different constructions. One example would be a phonological change such as the Great Vowel Shift. On the syllabic level, the development of a coda condition would not be a constructional change. Also a general loss of inflectional morphology due to language contact exemplifies a change that is not construction-specific, but that seizes the constructional network of a language as a whole. Word order reorganization from OV to VO commonly does not only change transitive clauses, but affects the order of head and modifier in other syntactic constructions. Likewise, phenomena such as the development of diglossia, the formation of supra-regional standards, and language death are not constructional changes. Many more examples could be adduced. For the purposes of the present chapter, constructional change will be defined as follows:

Constructional change selectively seizes a conventionalized form-meaning pair of a language, altering it in terms of its form, its function, its frequency, its distribution in the linguistic community, or any combination of these.

The main point of that definition is that constructional change starts by affecting a single node in the constructional network. Over time, such a change may spread toward other nodes, thus affecting groups of constructions. This means that the distinction between constructional change and language change is one of degree—even highly general processes of language change may have originated by affecting a single construction first. With this provisional definition in place, we can illustrate the notion of constructional change with several concrete examples and discuss corpus-based methods that have been used for their analysis.

25.2. Studying Constructional Change Through Corpora

This chapter discusses three different types of constructional change on the basis of illustrating corpus-based studies. These types represent changes in frequency, form, and function. The discussion begins with frequency, not only because frequency measurements are the mainstay of corpus linguistic methodology but also because corpus-based analyses of formal and functional change typically also rely on frequencies. The first sections outline four kinds of frequency change that do not necessarily involve structural change, but that pertain to the question how often a construction appears in a given context. The discussion then turns to changes at the formal pole of a construction, ranging from the morphophonemic to the syntactic. The last sections highlight different meaning changes that constructions may undergo. Naturally, formal, functional, and frequency changes are anything but unrelated. Many of the studies that are discussed here illustrate multiple types of change. Still, each of these types deserves to be highlighted separately. Table 25.1 gives an overview of the changes and corresponding studies that are examined in the following sections.

[INSERT TABLE 25.1 HERE]

25.2.1 Change in Text Frequency

Of all changes discussed here, change in text frequency is probably the most widely used measure of constructional change. It is well known that an increase in frequency is a common corollary of grammaticalization (Bybee 2007), even though it remains disputed whether increasing frequency is a cause, an effect, or a mere optional concomitant of grammaticalization. Supporting evidence for the latter view comes from grammaticalized forms that remain highly infrequent throughout their history (Hoffmann 2005). In order to interpret an observed frequency change, it is useful to check whether such a development coincides with formal or functional changes, as discussed in later sections. However, a look at raw frequencies is a common starting point for analyses of constructional change.

The easiest documentation of a development is the comparison of a construction across two corpora that represent different periods of time. The BROWN family of corpora affords precisely such comparisons, offering representations of British and American English from the 1960s and 1990s, respectively.¹ Mair (2006: 113) uses these resources to show that the English *get*-passive increased in frequency in both varieties, thus establishing itself as a new grammatical construction in the domain of the passive voice. Diachronic corpus analyses can be meaningfully complemented by studies of apparent time, in which the synchronic usage of speakers from different age groups is compared. Mair (2006: 115) adduces data from the spoken part² of the British National Corpus (BNC),² which shows that speakers between 14 and 24 years of age use the *get*-passive at a significantly higher rate than their older contemporaries. Diachronic and synchronic corpora can thus be combined to test and refine hypotheses.

Given a resource with a temporally continuous coverage of language use, frequency developments can be analyzed in a more fine-grained fashion. A very fine level of detail is offered by the quotations in the *Oxford English Dictionary*, which indicate the exact year in which a particular utterance was produced. Historical studies based on the *OED* typically divide the data into bins of half-centuries or similar time slices. Using twenty-five-year increments, Mair (2004: 129) shows that the English *be going to* construction underwent a frequency development that describes a “delayed increase.” Despite being a

¹ BROWN family corpora, <http://khnt.aksis.uib.no/icame/manuals/index.htm>.

grammaticalized construction by the end of the sixteenth century, *be going to* only gently increased in frequency until the dawn of the twentieth century, after which it showed rapid spread. This observation raises the theoretical question whether the delayed increase pattern generalizes to the frequency developments of other grammaticalizing constructions.

Often it is instructive to compare the frequency developments of two related constructions. In a study of the English voice system, Hundt (2004) discusses the emergence of the progressive passive (*The house is being built*) in the early nineteenth century and its relation to the semantically similar passival construction (*The house is building*). The latter survives in modern expressions such as *Flight 780 is now boarding*, but is rare otherwise. Does this allow the conclusion that the progressive passive ousted the passival construction? Data from the ARCHER corpus³ shows that by the time the progressive passive appeared, the decay of the passival was already underway, which casts doubt on a causal connection (Hundt 2004: 112). This interpretation is in line with Hopper and Traugott (2003: 126), who argue against the idea that grammatical constructions develop to fill functional gaps. Quite to the contrary, similar constructions tend to cluster together in various grammatical domains such as deontic modality or future time reference. Hundt's corpus analysis corroborates this idea.

25.2.2 Change in Relative Frequency

In many cases, constructional developments manifest themselves more clearly in changing relative frequencies, rather than in changing overall text frequencies. A slight increase in the frequency of one construction may not appear significant if viewed in isolation, but if simultaneously a similar construction decreases in frequency, the two developments may add up to a phenomenon worthy of investigation. This section discusses two types of scenarios in which attention to changing relative frequencies is particularly useful. Both scenarios relate to principles that Hopper (1991) has identified as characteristic for incipient grammaticalization. The first scenario is tied to the principle of divergence; the second to the principle of layering.

³ BNC (The British National Corpus), <http://www.natcorp.ox.ac.uk/>, <http://corpus.byu.edu/bnc/x.asp>.

Divergence describes the phenomenon that a grammaticalizing form and its coexisting lexical source undergo separate developments. A well-worn example of this is that only the grammaticalizing construction *be going to* underwent attrition to *gonna*, whereas the lexical verb *go* retained its full phonological form. Similarly, grammaticalizing forms undergo frequency developments that set them apart from their lexical sources, which typically do not undergo any dramatic frequency changes. Hence, contrasting the frequencies of a putative candidate for grammaticalization and its lexical source can substantiate that an apparent development is actually underway.

An illustrating example from English is the verb *keep*, which has a number of lexical senses relating to the idea of retaining a state (*keep quiet*), object (*keep the money*), or location (*keep off*). In conjunction with a gerund complement, *keep* has come to express the grammatical category of continuative aspect, as in *He keeps telling me about his problems*. Hilpert and Gries (2009) use the TIME corpus⁴ to investigate whether the ratio of examples in which *keep* takes a gerund complement has increased over time, showing that this is indeed the case. Whereas the form *keep V-ing* only accounts for 4% of all uses of *keep* in the 1920s, its relative frequency compared to other uses of *keep* rises continually, ranging between 12% and 15% in the 1990s and 2000s. This relative frequency change suggests that either the construction is becoming more popular in journalistic writing, or that the construction is becoming more widely applicable, and hence more grammatical. In either case, the relative frequencies reveal a change in progress.

The second phenomenon to be discussed here is relative frequency change in a scenario of layering. Layering describes the coexistence of different constructions for the expression of similar grammatical meanings. For instance, the grammar of English contrasts present and past time reference either phonologically through ablaut (*I write* vs. *I wrote*), morphologically through a dental suffix (*I work* vs. *I worked*), or periphrastically through the Perfect construction (*I work* vs. *I have worked*). The metaphor of ‘layering’ indicates that these alternatives represent different historical stages; the

³ ARCHER (A Representative Corpus of Historical English Registers), no website.

periphrastic strategy is newer than the two others (Hopper 1991: 24). Change in the relative frequencies of such alternatives may reveal that one construction is becoming the default choice, while another is falling out of usage. Alternatively, it may become apparent that each construction has found its grammatical niche and that the relative frequencies stay constant.

Tagliamonte (2004) studies the grammatical domain of deontic modality in northern British English, which contains several constructions for the expression of obligation, such as *must*, *have to*, *have got to*, and contracted variants. Rather than analyzing diachronic corpus data, Tagliamonte relies on synchronic data and works with an apparent time design. Using the YORK corpus,⁴ she compares the relative frequencies of four layered constructions across three different age groups. The analysis reveals that *must* is used less by younger speakers, whereas *have to* emerges as the default choice for the youngest group. Bare *got to* or *gotta* is used more by younger speakers, but remains a very infrequent choice across all age groups. The differing relative frequencies of these constructions point to ongoing transitions in the domain of deontic modality. Tagliamonte takes this observation as a starting point for a multivariate analysis that predicts the choice of a given construction, taking into account not only age, but also sex, education, and different linguistic variables. Modeling constructional change through simultaneous measurements of multiple interacting factors of frequency, form, and function is a fruitful avenue of research that is further discussed in section 25.2.5.

25.2.3 Change in Productivity

Frequency developments may not only pertain to token frequency (i.e., the question of how often a construction occurs in a given corpus). Developments may also take place at the level of type frequency. To illustrate, a construction may widen its range of co-occurring elements by host-class expansion (i.e., the integration of different syntactic categories) (section 25.2.8). A construction may also change by accepting a greater number of different elements from the same syntactic category, as is discussed in the sections on analogical extension (25.2.11) and collocational change (25.2.12). All of these phenomena

⁴ TIME corpus, <http://corpus.byu.edu/time/>.

reflect the type frequency, and hence the productivity of a construction. To illustrate, the English suffix *-able* allows the formation of adjectives through attachment to a verbal stem, yielding forms such as *acceptable*, *enjoyable*, or *predictable*. From a diachronic perspective, the *V-able* construction has undergone several changes regarding its acceptance of verbal stems: entering the English language as a French import during the Middle English period, it first occurred with Romance verbal bases. As the construction generalized, it came to be used with Germanic stems in forms such as *breakable* or *wearable*. The construction remains productive in Present-Day English, as is evidenced by *clickable*, *faxable*, or *cut-and-pasteable*. A diachronic corpus analysis can determine when speakers of English began forming adjectives with Germanic verbal stems, whether there were periods of enhanced productivity, and whether productivity remains strong in Present-Day English or has in fact stalled.

In a study of this kind, Bauer (2001: 8) uses the *OED* to analyze diachronic variation in the productivity of the suffix *-ment*, which attaches to verbal stems forming nouns such as *engagement* or *settlement*. In Present-Day English, the construction is no longer productive. Bauer divides the *OED* into fifty-year periods and determines how many new formations with *-ment* are found in each period. He shows that while the overall rate of productivity diachronically varies, new forms keep entering the English language until the nineteenth century, after which the construction becomes markedly less productive. From the perspective of Construction Grammar, the loss of productivity is an interesting phenomenon because it suggests that a general constructional schema (*V-ment*) has been lost, while of course its lexicalized forms (*engagement*, *settlement*, etc.) remain in use.

25.2.4 Change in Genre and Variety

Constructions may also change in the way they are associated with a certain linguistic genre or variety. For instance, a construction that originates in informal speech may over time find its way into elevated written registers—many such examples of colloquialization are discussed in Mair (2006). Conversely, constructions may also undergo a process in which their range of genres narrows over time.

One example of this kind is discussed by Noël (2008), who studies the English nominative and infinitive construction (NCI), which involves a subject, a verb of perception, cognition, or utterance in the passive, and a *to*-infinitive. The construction functions as a marker of evidentiality, as shown in (4a) and (4b).

- (4) a. The deal is thought to be worth £200m.
 b. Half of Canada's acid rain is said to come from US power stations.

Noël uses the Corpus of Late Modern English Texts (CLMET)⁶ and the BNC to track the frequency of the construction from 1640 to the 1990s, finding a nonlinear pattern. Throughout the four periods of the CLMET (1640–1920), the construction continually decreases in frequency, only to be represented much more strongly in the BNC. Noël suggests that the U-shaped development finds an explanation in genre differences between the CLMET and the BNC. Whereas the latter is a balanced corpus, the CLMET is limited to works of literature. Complementing evidence from the ARCHER corpus shows that historically, the NCI became increasingly more frequent in journalistic and scientific writing (i.e., genres in which the linkage of evidence to a source is of great importance), whereas other genres replicate the decreasing trend that can be observed in the CLMET. Closer examination of the BNC data confirms that newspapers and scientific texts in Present-Day English show a higher rate of NCI examples than other genres. Hence, the development of the NCI can be seen as a process of specialization into a particular genre. Noël's analysis further reveals that the NCI has developed different preferences for particular verbal patterns in journalistic and scientific writing, respectively. Frequent patterns in newspaper writing are *be said to*, *be reported to*, *be understood to*, and *be alleged to*, frequent patterns in science texts are, among others, *be shown to*, *be found to*, *be known to*, and *be estimated to*. The study thus demonstrates how the specialization process of the NCI proceeds even beyond particular genres at the level of individual collocations. The role of collocations in constructional change is the main subject of section 25.2.12 below.

⁶ CLMET (Corpus of Late Modern English Texts), <https://perswww.kuleuven.be/~u0044428/>.

25.2.5 Morphophonemic Change

Regular sound change represents a prime example of language change that does *not* instantiate constructional change. If the lexicon as a whole is affected, the fact that a single lexical construction used to have an initial stop but now has an initial fricative falls out of a general change and does not need to be stated separately. Still, there are changes involving alternations of sounds that are confined to specific constructions and that hence merit inclusion in this overview.

Gries and Hilpert (2010) study change in the English present tense. Between 1450 and 1680, the interdental third person singular suffix, as in *he doth*, is replaced by a suffix with an alveolar fricative, as in *he does*. This development is not paralleled in lexical items; the fricatives in *breath* or *youth* do not change. A corpus-based analysis can take measurements of relative frequencies (section 25.2.2) to determine when this constructional change happened and for how long there was variation between the two forms. Beyond that, corpora afford an analysis of the linguistic and nonlinguistic factors that correlate with the choices that speakers make. Technically speaking, the outcome of a dependent variable (*-th* vs. *-s*) can be predicted on the basis of several independent variables, which concern frequency, form, function, and extra-linguistic factors. It can thus be investigated, which structural contexts favor the use of the new form, which speakers were driving the change, and how linguistic and extra-linguistic factors interact.

Using the Corpus of Early English Correspondence (CEEC),⁷ a corpus of letters that is annotated for various social variables, Gries and Hilpert extract all relevant verb forms. Each example is annotated for the dependent variable (*-th* vs. *-s*) and for independent variables including (1) which verb is used, (2) whether that verb is lexical (*sing*) or grammatical (*do*, *have*), (3) whether the verb stem ends in a sibilant (*wish*), (4) whether the following word begins with an *s* or *th*, and (5) which variant of the third person singular suffix was used last. Nonlinguistic factors include (6) the successive time periods, (7) the identity of the author, and (8) the gender of author and intended recipient. To model the development from *doth* to

⁷ CEEC (Corpus of Early English Correspondence), <http://www.helsinki.fi/varieng/domains/CEEC.html>.

does, Gries and Hilpert fit a logistic regression. The analysis reveals first of all which of the above factors do not influence the choice. For instance, the gender of the intended recipient does not play a role. By contrast, the gender of the author, the grammaticality of the verb, and the previously used form each show a significant influence. Crucially, some influences do not stay constant over time. A final sibilant matters to the choice only during one early period, after which the effect disappears.

Multivariate analyses of phonological changes already have a long history in quantitative sociolinguistics, and recent years have seen the development of corpus-based historical sociolinguistics (Nevalainen and Raumolin-Brunberg 2003). Corpus-based approaches to constructional change stand to benefit a lot from adopting and refining these techniques. Especially as Construction Grammar endorses the idea that constructions may change at any level of structure, meaning, or use, the possibility to monitor these levels simultaneously over time holds a great potential to deepen our understanding of constructional change.

25.2.6 Morphosyntactic Change

Perhaps the most salient characteristic of constructional change is the continuous morphosyntactic reduction of grammaticalizing forms. Stock examples of this are constructions such as English *gonna* from *be going to*, and the French morphological future *je ferai* ‘I will do’ from the periphrastic Latin form *facere habeo* ‘I have to do’ (Hopper and Traugott 2003).

An example from German studied by Nübling (2005) concerns the fusion of prepositions with definite articles, from *auf das* ‘onto the’ to *aufs*, from *bei dem* ‘at the’ to *beim*. This process, which is attested since Old High German times, is not fully productive, as some prepositions and articles do not fuse. Forms such as **gegenüßerm* ‘across the’ are not attested, necessitating a piecemeal account of individual constructional developments. Fusion is obligatory in some usages (*im Himmel* ‘in heaven,’ **in dem Himmel*), highly frequent with other forms, and rare with still others. Orthographical variation (*für*’s/*fürs* ‘for the’) suggests that constructional change is in progress, but interestingly, modern varieties of German have lost fused forms that were attested in earlier dialects; especially fused forms with genitive

articles. The phenomenon under investigation is thus a cluster of constructional changes that represent developmental stages of similar, but nonidentical processes.

Nübling uses the IDS corpora of modern written German as a resource to study this process from a synchronic perspective. In current usage, approximately twenty-five preposition-article pairs alternate between the fused form and the analytical variant. The forms *am* ‘at the,’ *zum* ‘to the,’ and *im* ‘in the’ vastly outnumber their nonfused counterparts; *hinters* ‘behind the,’ *ans* ‘to the,’ and *ums* ‘around the’ display relatively balanced frequencies; the forms *unterm* ‘under the,’ *vorm* ‘in front of the,’ and *überm* ‘over the’ are relatively rare. How can these figures be explained? Nübling suggests several motivating factors. First, fusion is more frequent with prepositions that govern the dative, rather than the accusative or genitive. Also the co-occurrence frequency of preposition and article explains some of the observed variance. Developmentally old and short prepositions (*an*, *in*, *zu*) are more likely to fuse than more recently grammaticalized prepositions (*anstatt*, *wegen*, *dank*). Beyond that there is much dialectal and genre-specific variation. In determining the relevant parameters of an ongoing change, Nübling’s analysis of present-day corpus data lays the groundwork for the kind of diachronic multivariate analysis that was described in the preceding section. Her analysis underscores that the study of constructional change requires simultaneous attention to form, frequency, and a multitude of extra-linguistic variables. The work also raises questions that show the limits of current corpus-linguistic resources: at present there are no diachronic corpora of German that would allow the researcher to control the variables of genre, variety, and several structural factors while tracing the development of a construction over time. Corpus-based approaches to constructional change thus have to find ways to ask questions that currently available resources can illuminate.

25.2.7 Change in Argument Structure

The development of a construction may bring along a change in the valency of an element that forms part of it; commonly this concerns the argument structure of verbs. New arguments may become obligatory,

required arguments may be subject to change, or even loss. Diachronic corpus data can show when and how new argument structure patterns developed.

Hilpert and Koops (2008) use data from the Swedish SPRÅKBANKEN resource⁸ to document a change in the argument structure of a grammaticalizing complex verb construction. Specifically, they discuss the development of the Swedish Pseudo-coordination construction (SPC) with the posture verb *sitta* ‘sit.’ In this construction, the verb *sitta* is coordinated with a second verb and takes on an aspectual meaning of durativity. (For a rough comparison, consider the English sentence *We sat and talked*.) Diachronically, Swedish *sitta* underwent a change in its argument structure as its lexical meaning gave way to grammatical meaning: the more *sitta* came to be used as an aspectual marker, the less relevant its locative meaning became. This is reflected in the degree to which the verb is used with prepositional phrases or adverbials that specify where the sitting takes place. While lexical *sitta* is often elaborated spatially (7a), such elaboration is typically absent in the SPC (7b). Nevertheless, lexical *sitta* can occur without spatial elaboration (7c), and coordinated *sitta* may occur with it (7d).

- (7) a. Jag har suttit vid skrivbordet nästan hela dagen.
I have sat at desk.the almost all day
‘I’ve sat at the desk almost all day.’
- b. Jag har suttit och läst hela dagen.
I have sat and read all day
‘I have been reading all day.’
- c. Han satt en stund igen.
he sat a while again
‘He sat (somewhere) for a while again.’
- d. Jag har suttit vid skrivbordet och läst hela dagen.
I have sat at desk.the and read all day

⁸ SPRÅKBANKEN (various Swedish corpora), <http://spraakbanken.gu.se/>.

‘I’ve sat at the desk and read all day.’

In order to demonstrate that *sitta* underwent a change in its argument structure, Hilpert and Koops (2008: 255) measure the relative frequency of spatial elaboration with lexical *sitta* and the SPC over different periods of time, finding that the rate of spatial elaboration stays constant with lexical *sitta*, but decreases significantly within the SPC. The observed change is thus a construction-specific change in argument structure.

25.2.8 Host-class Expansion

Himmelmann (2004) suggests the notion of host-class expansion as a useful parameter of constructional change in grammaticalization. A syntactic head that initially only occurs with a specific type of complement may increase its range of possible complements to include more and different syntactic categories. Diachronic corpus analysis can bring to light when and how this extension took place, and how the different complement types changed in frequency.

An illustrating example is offered by Patten (2010), who studies the development of *it*-clefts in the Penn Parsed Corpora of Historical English (PPCME, PPCEME).⁹ The English *it*-cleft consists of a focal element, most frequently a noun phrase, which is followed by a relative clause expressing presupposed information. An example of this structural type is given in (8a). In Present-Day English, *it*-clefts allow several different categories as complements of the copula, for instance, prepositional and adverbial phrases (8b, 8c) or *that*-clauses (8d).

- (8) a. It was the therapist_{NP} that killed her.
 b. It is in December_{PP} that she’s coming.
 c. It was here_{AdvP} that we met.
 d. It’s that he’s so annoying_{THAT-CL} that bothers me.

⁹ PPCME2 (The Penn Parsed Corpus of Middle English), <http://www.ling.upenn.edu/hist-corpora/>.

PPCEME (The Penn Parsed Corpus of Early Modern English), <http://www.ling.upenn.edu/hist-corpora/>.

Patten demonstrates through corpus analysis that the *It*-cleft construction has undergone host-class expansion over time. Prepositional phrases appear as focal elements from the fourteenth century onward, adverbial and clausal elements are an innovation of the Early Modern English period. The construction as such gains in overall frequency; noun phrases remain the most common focal elements throughout the entire corpus. The gradual process of host-class expansion that Patten observes allows her to argue against an alternative account (Ball 1994), which derives *it*-clefts with non-nominal focal elements from an impersonal construction.

25.2.9 Metaphor and Metonymy

Metaphorical and metonymic extension has been studied extensively in the context of polysemy (Croft and Cruse 2004). In particular the polysemy of prepositions has received much attention, as many spatial meanings of prepositions are metaphorically extended into temporal meanings or other semantically more abstract domains. Meanwhile, also syntactically complex constructions display synchronic polysemy, as they are subject to the same diachronic processes of extension. The English Ditransitive construction (Goldberg 1995) commonly encodes an intended transfer of an object, but is being used metaphorically in examples such as *Thyme gives the soup a nice flavor*, which denotes a causal relationship rather than a transfer. The semantic extension of constructions can be studied through diachronic corpus analysis, in particular the qualitative analysis of ambiguous examples that represent bridging contexts between two different interpretations.

Petré and Cuyckens (2008) study the development the verbal prefix *be-* from Old English onward. Old English had a variety of verbal prefixes indicating spatial directions, as in *a-drifan* ‘drive out/away,’ *be-rowan* ‘row round,’ or *to-brecan* ‘break asunder.’ The loss of this prefixation pattern can be seen as a leveling process resulting from the OV to VO shift that English underwent. Of the OE prefixes, only *be-* remains sporadically productive in expressions such as *be-trained* ‘wearing running shoes’ or *be-wigged* ‘wearing a wig.’ It is thus of interest why a single construction resisted the tide of a syntactic sea change, and it is this question that Petré and Cuyckens investigate. Based on corpus data from the

YCOE¹⁰ and YPC,¹¹ they argue that the dominant sense of the construction in OE was the spatial meaning of enclosure, as in *besiege*. This sense served as a prototype from which other, more abstract meanings develop. The semantic extensions with the highest frequencies are labeled ‘Extensive Coverage,’ ‘Total Affectedness,’ and ‘Furnishing.’ The latter is the most recent extension and also the most productive in modern usage (cf. *be-trained*).

The meaning extension from enclosure to extensive coverage is motivated by metonymy, specifically a part–whole relationship. Some uses of the *be*-construction extended the idea of enclosure to the closely related idea of complete coverage. Examples such as (9a) illustrate this extension (Petre and Cuyckens 2008: 152).

- (9) a. Se apostol hine begeat mid ðam wætere
 the apostle him be-poured with the water
 ‘The apostle doused him with the water.’

The next extension abstracts away from the domain of space and projects the idea of coverage onto affectedness. Petre and Cuyckens argue that an event of covering an entity will invite the inference that the entity in question is also strongly affected. Over time, the spatial meaning of the *be*-construction became secondary, leading to the possibility of examples such as (9b), which makes no reference to a spatial location (2008: 154).

- (9) b. Bewyl þara meolce þridðan dæl.
 be-boil of.the milk third part
 ‘Boil away a third of the milk.’

¹⁰ YCOE (The York-Toronto-Helsinki Parsed Corpus of Old English Prose),

<http://www.ling.upenn.edu/mideng/ppcme2dir/YCOE/doc/annotation/YcoeLite.htm>.

¹¹ YPC (The York-Helsinki Parsed Corpus of Old English Poetry), <http://www-users.york.ac.uk/~lang18/pcorpus.html>.

A further extension of the *be*-construction targets the idea of furnishing an entity with an object or an attribute. This process involves host-class expansion (cf. section 25.2.8), as the prefix attaches to nominal and adjectival hosts (*bespouse*, *befoul*), which are thereby coerced into verbhood. The meaning of furnishing is the mainstay of the *be*-construction in Present-Day English, where it occurs primarily in the form of participial adjectives like *bespectacled*.

Petré and Cuyckens argue that the successive semantic extensions of the *be*-construction rendered it more salient than other OE prefixes. Because of this salience, *be*- resisted the process by which the other prefixes semantically bleached and subsequently fell out of usage. The study thus demonstrates the importance of close, qualitative investigations of constructional developments, as these may go against the grain of language changes on a larger scale.

25.2.10 Analogical Extension

In Construction Grammar, grammatical knowledge is thought to be a hierarchical network of symbolic units that display varying degrees of schematicity. A construction that is highly specific, for instance, an idiom such as *not give a damn*, may sprout offshoots such as *not give a hoot* or *not give a monkey's*, in which speakers replace one part of the idiom with an analogous element. Repeated analogical extensions may over time lead to the emergence of a general schema *not give a NP*, which invites further additions to the range of expressions occurring in this now partly schematic idiom. Stepwise analogical extension may offer a diachronic explanation for constraints on the synchronic use of a construction, motivating for instance, why certain lexical items cannot occur with it. Through the use of diachronic corpora, the notion of analogical extension can be operationalized as an increasing range of elements that occur within a given constructional slot (cf. section 25.2.3 on productivity).

Taking such an approach, Israel (1996) uses the *OED* to study the English *Way* construction, which expresses the creation of a (commonly metaphorical) path through a terrain with substantial obstacles. This constructional meaning is reflected in two types of main verbs that the *Way* construction accommodates in modern usage: verbs denoting the manner of a movement, such as *crawl*, *wind*, or *plod*,

and verbs denoting the means of creating a path, such as *dig*, *cut*, or *force*. The focus will be on the latter type here. Israel reports that by 1650, the *OED* contains examples with means verbs such as *pave*, *smooth*, and *cut*, which directly represent the domain of road building and path clearing (10a). A century later, the *OED* attests more verbs from the same semantic domains, but additionally also verbs from the domain of fighting (10b). Since road building requires force, Israel argues that the original meaning has undergone analogical extension to the use of force in order to make progress on a path. Example (10c) demonstrates that the construction has become even more schematized by the nineteenth century, during which an even greater range of co-occurring verbs is attested. Semantically, the use of force to make progress has been extended to any purposive, goal-directed activity.

- (10) a. Arminius paved his way first by aspersing and sugillating the fame and authority of Calvin. (1647)
- b. Every step that he takes he must battle his way. (1794)
- c. Not one man in five hundred could have spelled his way through a psalm. (1849)

25.2.12 Collocational Change

The preceding section discussed the idea that the occurrence of a construction with new lexical elements is indicative of analogical extension, and hence semantic change. Besides qualitative changes in the collocate set of a construction, there are also quantitative changes. Certain collocates may become more or less frequent over time, and these shifts may be no less indicative of change than the occurrence of new lexical collocates. A quantitative analysis of shifting collocational preferences can thus offer some insight into the semantic development of a grammatical construction. With this aim in mind, Hilpert (2006a; 2008) applies the method of distinctive collexeme analysis (Stefanowitsch, this volume) to diachronic data. Distinctive collexeme analysis contrasts two or more constructions with respect to their collocational preferences. Unlike the synchronic variant, the diachronic application does not contrast multiple constructions, but studies a single construction across multiple periods of time. On the basis of diachronic

corpora, the method can determine what types of co-occurring elements were typical of a given construction at different historical stages.

Using the example of English *shall*, Hilpert shows first that the construction appears to remain stable with regard to its lexical collocates. In the PPCEME, which is divided into three periods of time, *shall* occurs with remarkably similar sets of collocates between 1500 and 1710. In each of the three periods, the verbs *be*, *have*, *see*, and *find* are among the five most frequent verbs. The collostructional method abstracts away from items that are common in each period, instead highlighting those that are significantly more frequent than expected, given their overall frequencies across the three periods. Items are judged as distinctive if they occur frequently in one period but are relatively sparse in the two others. In this way, differences between the three periods are accentuated and semantic developments become visible. Technically, the method determines for each verb whether the observed occurrences deviate significantly from a chance distribution. This kind of calculation determines the most distinctive elements for each of the three periods, promoting the ranks of verbs that are maximally unevenly distributed. The ten most distinctive elements for each period are shown in Table 25.2. All verbs that are shown are significantly distinctive for their respective periods; a CollStr value larger than 1.3 signals significance at the level of $p < 0.05$.

[INSERT TABLE 25.2 HERE]

The most distinctive verb of the first period, *fortune*, occurs nine times in the first period, but not at all elsewhere. It appears as the most distinctive element of the first period. The most distinctive verbs are therefore those that occur with some frequency, but are idiosyncratic for just one period. Changes in the most attracted verbs can be interpreted as a symptom of ongoing semantic change, and the observed tendencies can be compared against pre-existing hypotheses about the development of the construction.

A tendency that can be observed in Table 25.2 is an increase of stative verbs over the three periods. The first period lists none, the second period has *contain*, the third period lists *stay*, *stand*, *extend*, and *be*. Another tendency that emerges when actual examples with distinctive verbs are examined is that

shall in the third period typically assumes a text-structuring function, as is illustrated in the following examples.

- (11) a. I shall add a little more evidence.
- b. Some products and places provide a few exceptions to this pattern, as we shall discover in the next chapter.
- c. We shall examine this further in Chapter 7.
- d. I shall refer to this comparative study where appropriate in this report.

Whereas the auxiliary *shall* has largely fallen out of usage as a general future marker, the idiomaticity of the above examples suggests that it has found a functional niche as a text-structuring device in formal writing.

25.3. Concluding Remarks

The introduction posed the question why corpora are useful for the study of constructional change. As the previous sections have illustrated, constructions change in frequency, form, and function, typically in gradual fashion. The primary means to document change in the usage of a construction are frequency measurements in conjunction with the use of inferential statistics. Corpora hold the potential to demonstrate that there are indeed differences between, say, the usage patterns of a construction in two different historical periods, or across two different age groups in synchronic data. Especially when noncategorical changes are at stake, quantitative data from corpora may provide the only means to establish that a change is underway.

The case studies discussed above have also shown the practical side of corpus work in the study of constructional change. In some cases, a close examination of corpus examples from different historical periods already reveals that a given change has occurred. In many cases, however, analyses rely on the exhaustive retrieval of a given construction from a corpus, because the entire set of tokens is needed to arrive at generalizations about how the construction changed. Annotation of all tokens for formal and

functional features allows quantitative analyses of the factors that changed a construction, and thereby allows the researcher to decide between competing explanations of the same phenomenon.

Finally, it needs to be briefly discussed how the constructional perspective sets apart the work described above from other approaches in historical linguistics. On a theoretical level, work on constructional change sets out to document phenomena that cannot be reduced to more general patterns of language change (i.e., it aims to find processes that are construction-specific). Much as early work in synchronic Construction Grammar (e.g., Fillmore, Kay, and O'Connor 1988) determined the aspects of a given construction that were not reducible to more general grammatical patterns, work on constructional change studies those processes that do not reduce to a broader development of grammatical reorganization. In order to demonstrate that a change is indeed idiosyncratic, not general, corpus data are again of utmost importance. Only quantitative corpus data can document that, for instance, two seemingly related processes are in fact independent. While the constructional perspective thus places great importance on the particularities of language change, it is also committed to finding generalizations wherever possible. Changes in historical corpus data may yield evidence for the development of abstract constructional patterns that represent linguistic generalizations. Determining the exact level of abstractness at which a given construction needs to be posited is again only possible with recourse to corpus data.

In sum, corpus-based approaches to constructional change address a wide array of phenomena with a similarly wide array of methods, but converge on the theoretical aim of offering the best descriptions of—and explanations for—the development of form-meaning pairings in language.

Table 25.1.

Corpus-based studies of constructional change

Type		Phenomenon	Reference
Frequency	(1)	change in text frequency	Mair (2004, 2006); Hundt (2004)
	(2)	change in relative frequency	Hilpert and Gries (2009); Tagliamonte (2004)
	(3)	change in productivity	Bauer (2001)
	(4)	change in genre and variety	Noël (2008)

Form	(5)	morphophonemic change	Gries and Hilpert (2010)
	(6)	morphosyntactic change	Niöbling (2005)
	(7)	change in argument structure	Hilpert and Koops (2008)
	(8)	host class expansion	Patten (2010)
Function	(9)	metaphor and metonymy	Petré and Cuyckens (2008)
	(10)	analogical extension	Israel (1996)
	(11)	collocational change	Hilpert (2008)

Table 25.2.

Top ten distinctive collexemes of *shall* over three periods of time

1500–1570			1570–1640			1640–1710		
Verb	N	CollStr	Verb	N	CollStr	Verb	N	CollStr
fortune	9	3.80	incur	13	3.38	add	14	5.52
wage	8	3.38	assemble	7	2.68	direct	8	5.46
divide	8	2.60	feed	7	2.68	discover	5	3.41
appear	24	2.07	hear	43	2.56	examine	7	3.39
perceive	11	1.92	offend	12	2.17	stay	6	2.81
understand	26	1.81	fall	14	2.04	refer	4	2.73
beg	7	1.75	contain	7	1.97	stand	11	2.65
require	9	1.66	enjoy	5	1.92	endeavor	7	2.60
say	28	1.61	imagine	5	1.92	extend	15	2.58
attain	5	1.50	do	65	1.79	be	234	2.50