Published in Constructions and Frames 8, issue 1, 66-85, 2016,
which should be used for any reference to this work

Change in modal meanings

Another look at the shifting collocates of *may*

Martin Hilpert

This paper discusses how modal auxiliaries fit into a constructional view of language and how this view allows us to think in new ways about diachronic meaning change in modal auxiliaries. These issues will be illustrated on the basis of a diachronic corpus-based study of the modal auxiliary *may*, specifically on changes in its collocational preferences during the past 200 years. The main point of this paper is the claim that a constructional view needs to take account of the mutual associations between modal auxiliaries and the lexical elements with which they occur. Changes in these mutual associations are usefully under-stood as change in a complex network of constructions.

Keywords: modality, modal constructions, collocational change, semantic change, semantic vector space modeling

1. Introduction

The main aim of this special issue is an examination of the question how Construction Grammar can deepen our understanding of modal auxiliaries, that is, their meanings, their morpho-syntactic behavior, and the way they change over time. Specifically in the case of the English modal auxiliaries, this is a tall order, given that there are few topics in the grammar of English that have received similar amounts of attention. Even if we restrict our view to works with a broadly functionalist background, it is difficult to draw up a list of essential readings (although Plank 1984; Traugott 1989; Bybee & Fleischman 1995; Fischer 2007; and Bybee 2010, amongst others, would be included in many such lists), let alone give a coherent summary of current work. This paper will therefore adopt a fairly narrow scope. What will be discussed is how modal auxiliaries fit into a constructional view of language, and how this view allows us to think in new ways about diachronic meaning change in modal auxiliaries. These issues

will be illustrated on the basis of a corpus-based study that draws on data from the Corpus of Historical American English (Davies 2010) and the Corpus of Contemporary American English (Davies 2007). A particular focus will be on the modal auxiliary *may*, specifically on changes in its collocational preferences that have been going on over the past 200 years. The main point of this paper will be the claim that a constructional view needs to pay close attention to the mutual associations between modal auxiliaries and the lexical elements with which they occur. Changes in these mutual associations are usefully understood as change in a complex network of constructions. This general idea is of course not new, in fact it has already given rise to a sizable body of constructional research. A well-known tool for the empirical investigation of associations between constructions and lexical items is the suite of collostructional methods that has been developed by Stefanowitsch & Gries (2003) and Gries & Stefanowitsch (2004a, 2004b). These methods have been applied to auxiliaries and change in their collocational profiles, specifically the types of lexical verbs that tend to occur with a given auxiliary, in Hilpert (2008). This paper will build on that work, but will pursue a different, somewhat more general approach that draws on ideas that have been developed in Sagi et al. (2011). In order to study the development of *may*, a semantic vector space model (Turney & Pantel 2010) will be constructed on the basis of its verbal collocates and the frequencies of those collocates over different periods of time. It will be shown that this approach can shed light on semantic changes that *may* has undergone during the last two centuries. This will give rise to the argument that modal auxiliaries are profitably viewed as constructions in the technical sense (Goldberg 1995, 2006). What follows from this is that probabilistic associative relations between constructions and lexical elements play a central role in the network architecture of linguistic knowledge that is proposed in cognitive and usage-based Construction Grammar.

The remainder of this paper is structured in the following way. Section 2 lays the theoretical groundwork by outlining a general idea of how modal auxiliaries could be understood from a constructional perspective. Section 3 briefly reports on an earlier study of collocational shifts in modal auxiliaries (Hilpert 2013b) that provides the relevant context for the empirical work that is presented in this paper. Section 4 presents the case study of *may* and thus connects the theoretical ideas from Section 2 to corpus-based observations. Section 5 concludes the paper, summarizes the main findings, and suggests further ways in which constructional research might investigate modal auxiliaries.

2. What, if anything, are modal constructions?

Considering the large amount of research that has addressed modal auxiliaries, it is perhaps remarkable that relatively little of that work has explicitly adopted a constructional perspective — studies such as Boogaart (2009), Bergs (2010), or Goldberg and van der Auwera (2012) remain the exceptions. A possible explanation for this is that constructional work, in the wake of influential studies such as Fillmore et al. (1988), has put great emphasis on the study of non-canonical constructions. In comparison to constructions such as *let alone* (Fillmore et al. 1988), the *way*-construction (Goldberg 1995), or *the X-er the Y-er* (Kay & Fillmore 1999), modal auxiliary constructions like *will* plus infinitive appear a lot less idiosyncratic, and more amenable to a description in terms of general syntactic generalizations. The combination of *will* and an infinitive furthermore fails several tests that are commonly used as criteria for the identification of constructions. In Goldberg (1995: 4), non-compositionality of meaning and non-predictability of formal aspects are suggested as definitional criteria for constructions. Even though Goldberg (2006: 5) revises that definition and downgrades non-predictability to the status of a sufficient but not necessary criterion, formal and functional non-predictability remains an important idea that conceptually motivates a constructional approach to linguistic knowledge (cf. Hilpert 2014: 9–13). Beyond that, non-predictability allows the analyst to identify constructions as such. Examples include the idiom *face the music*, which exhibits non-compositional meaning, or the so called 'big mess' construction (*How big a boat are we talking about?*), which shows non-canonical and hence unpredictable syntactic structures. In the case of *will* plus infinitive or other modal auxiliaries, it is hard to argue for construction status on the basis of these criteria. The examples of modal auxiliary constructions in (1) seem to be both semantically transparent and syntactically regular:

(1) a. We will arrive in a few minutes on platform number five.
    b. You can borrow my umbrella.
    c. Should I leave the door open?
    d. Would you mind if I opened the window?

The respective meanings of auxiliary and lexical verb (*will* and *arrive*, *can* and *borrow*, etc.) appear to add up compositionally in these examples. As regards the morpho-syntactic behavior of *will*, *can*, and other modals, the general combination of an auxiliary verb with a lexical verb in the infinitive form qualifies as a canonical grammatical pattern that manifests itself in identical form with all core modal auxiliaries, and in very similar form in other auxiliaries, as for instance *dare*. The same observation holds for the inversion of auxiliary and subject in polar questions, which are shown in (1c) and (1d). This means that two central

definitional criteria for constructions (pace Goldberg 1995) fail in the case of modal auxiliaries.

Another commonly invoked criterion for constructionhood is the presence of irreducible constraints on the use of a linguistic pattern. Even if, at first glance, a pattern may appear to conform to general grammatical regularities, such constraints may betray its status as a construction. For example, the sentence *Mary is a smarter lawyer than John* seems a perfectly regular instantiation of general grammatical patterns of English: The sentence is a predicative construction (*X is Y*) that includes a name (*Mary*) and a noun phrase (*a smarter lawyer than John*). Yet, this analysis overlooks the fact that speakers of English do not produce sentences like **Mary is the smarter lawyer than John*, which differs from the earlier example only in the definiteness of the determiner *the*. Constraints like the restriction of a construction with regard to definiteness, number, tense, or any other grammatical category show that speakers must have internalized a separate generalization beyond the general patterns that underlie the construction in question. However, also the criterion of unpredictable constraints does not unambiguously identify modal auxiliaries as constructions. Auxiliaries such as *will*, *could*, or *can* do not exhibit clearly recognizable restrictions, for instance with regard to the kinds of subject pronouns that they can take, or the kinds of lexical verbs that are used with them. Hence also this criterion for constructionhood leaves it open whether or not modal auxiliaries are to be viewed as constructions.

A criterion for constructionhood that actually does provide an argument for viewing a pattern such as *will* plus infinitive as a construction is the argument from frequency, which Goldberg (2006: 5) introduces in her revised definition of constructions. If a linguistic form is used often enough so that it becomes entrenched in speakers' minds, then it is to be viewed as a construction, even when its formal and functional properties are fully predictable. For a modal auxiliary such as *will*, the argument for constructionhood can thus be made in the following way: Even though *will* may be combined with virtually any lexical verb of the English language, some combinations are used much more often than would be expected, whereas others are used much less often than expected. In other words, *will* has a collocational profile that reflects an attraction towards certain types of lexical verbs, and repulsion of other types of verbs. Collocational profiles of this kind can be determined through collostructional analysis (Stefanowitsch & Gries 2003), which operates on the basis of frequency counts of constructions and the lexical elements that occur within them. Using collostructional techniques, Gries & Stefanowitsch (2004a: 114) establish that *will* differs significantly in its preferences from *be going to*. Hilpert (2008: 101) finds that the lexical verbs that are most strongly attracted to *will* include *come*, *need*, *continue*, *depend*, and *find*. These verbs share a number of semantic features, for instance a low degree of transitivity

and dynamicity, and the fact that they do not require intentional agents. Crucially, these collocational preferences are not predictable from any other knowledge of language that speakers of English can be assumed to have, yet, the mutual associations between *will* and verbs such as *come* and *need* must be seen as an integral part of speakers' linguistic knowledge. As Taylor (2012: 100) points out, linguistic knowledge cannot be reduced to the distinction of grammatical and ungrammatical utterances in one's native language. Rather, a central aspect of knowing one's native language is the ability to speak idiomatically, including the habit of pairing a given modal auxiliary with the lexical verbs that are typically used with it.

To come to an answer to the question that was asked in the title of this section, this paper affirms the idea that there are modal auxiliary constructions, and it proposes a view of such constructions that includes the following aspects. First of all, a modal construction such as *will* plus infinitive is viewed as a complex symbol that connects a form with a meaning. Its form is partly schematic, as it includes both the fixed form *will* and a schematic slot for a lexical verb in the infinitive. Also the meaning pole of the construction is complex. As is well-documented (cf. Coates 1983; Bybee & Pagliuca 1987, amongst many others), *will* serves a number of different semantic functions, which include future time reference and epistemic modality, but which are not restricted to these two. Modal auxiliary constructions are thus profitably seen as polysemous, such that the formal pole is linked to several historically and synchronically interrelated meanings that constitute the complex meaning pole of the construction. The formal pole of the construction furthermore contains information on how *will* plus infinitive can be integrated into larger syntactic structures, specifying for instance that it can instantiate the predicate in the more general subject-predicate construction. Thus far, this description of *will* plus infinitive as a construction represents a standard view that, for example, matches statements about constructions made in Fillmore et al. (1988), and general ideas about symbolic units developed in Langacker (1987). Where the present proposal goes beyond that view is in the following: Taking the findings of usage-based linguistics (Tomasello 2003; Bybee 2010) and collostructional studies (Stefanowitsch & Gries 2003) seriously, it is argued here that knowledge of a construction includes probabilistic knowledge of how that construction is associated with lexical elements. On the view that is taken in this paper, association patterns between syntactic constructions and lexical items are part and parcel of linguistic knowledge. Phrased in more general terms, this view boils down to the very simple claim that if a speaker knows a construction, that speaker has knowledge of links in a constructional network, and of the respective strengths of those links. She or he knows how (and how strongly) that construction is connected to other constructions, including lexical ones.

A fundamental consequence of this view is that language change may manifest itself not only in changes of meaning and changes of syntactic, morphological, or phonological form, but also in changes that affect only the relative strength of connections in the constructional network. A modal auxiliary construction such as *will* plus infinitive may remain formally unchanged, and even unchanged with regard to the spectrum of semantic functions that it can convey, but speakers' knowledge of the construction can still undergo change when the construction undergoes a shift in its collocational profile. Certain verbs may become more strongly attracted to *will*, others may decrease in the strength of their mutual attraction. Hilpert (2008: 79–84) uses collostructional analysis to track the shifting verbal collocates of *will* from the 16th to the 20th century and determines the verbs that are most distinctive for the respective time periods. One result of that study is that there is an asymmetry between earlier and later periods with regard to the semantic feature of intentionality. The verbs that are distinctive for later periods feature a greater number of verbs that denote involuntary actions (*cry*, *understand*, etc.) rather than intentional actions. This observation is consonant with the grammaticalization scenario for future constructions that was proposed by Bybee et al. (1991), in which verbs with the meaning of volition first turn into markers of intentionality before they become markers of future time reference. To be sure, all of these meanings are co-present already in the 16th century (cf. Visser 1969: 1692), but the shifts in their relative importance reveals itself in the changing collocational profile of *will*. In a nutshell, then, collocational change in the short term may proceed almost invisibly, but in the long term, this kind of change tends to result in more tangible change in the meaning pole of a construction.

With this general idea of modal auxiliary constructions in mind, it is now time to examine the question how that idea may be used in actual empirical investigations in order to gain new insights. The present analysis takes its cues from research into the visualization of semantic change (Sagi et al. 2011) and meaning relationships (Brezina et al. 2015). Both of these strands of research are based on the insight that co-occurrence patterns in corpora reflect meaning, which has a double relevance in this context. The first relevant aspect is that differences in collocational behavior allow the researcher to contrast alternative forms, such as *may* and *might*, or to compare the uses of one single form across different historical periods. Sagi et al. (2011) adopt the latter approach and visualize the historical processes of semantic broadening and narrowing for different lexical words of English. As these elements change in meaning, they also change in their collocational behavior, that is, the variety of co-occurring words and the strength of association between an element and its words are changing. The second crucial aspect is that the collocates of a linguistic element can be analyzed as a semantic space that reflects the meaning potential of that linguistic element. Brezina et al. (2015)

develop a software for the visualization of collocational networks that allows the user to explore the semantic relations between an element and its collocates. They apply this tool in a case study of the word *swearing* in a corpus of 17th and 18th century texts, which reveals different semantic classes of collocates (2015: 157), such as words that encode general negative concepts (*false*, *contemptuous*) as opposed to words that have specific religious associations (*sin*, *conscience*, *temptation*). To understand word meaning in terms of such collocational networks means that speakers have knowledge of the associations that a word such as *swearing* may evoke in different contexts. The next two sections draw on these ideas and report on two diachronic, corpus-based studies that track historical semantic change in English modal auxiliaries. The analyses make use of the COCA (Davies 2008) and COHA (Davies 2010). The choice of these corpora is motivated in part by a general problem of diachronic corpus linguistics, namely that any changes across different historical periods might be due to sampling error, rather than actual change in language use. The risk of sampling effects is smaller in corpora that are sufficiently large and that strive for a well-balanced composition that is the same for all periods. The above-mentioned corpora fulfill these characteristics.

3. Collocational shifts in the English core modals

Hilpert (2013b) uses the COHA corpus to investigate diachronic collocational shifts in the nine English core modals *can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, and *would*. For each of these modal auxiliaries, all instances that are followed by a lexical verb in the infinitive are retrieved from the corpus. The frequencies of those lexical verbs are aggregated for each of the fifteen decades between the 1860s and the 2000s, so that each decade is represented by a table of data that lists for all nine modals the collocation frequencies of each lexical verb. Expectably, all nine modals are frequently used with general high-frequency verbs such as *be*, *have*, and *do*. Yet, in the full range of lexical verbs that are found with the modal auxiliaries, there are verbs that do not occur with all of the modals in each decade. Some verbs exhibit clear frequency asymmetries, so that they do not occur to the same extent with all of the nine modals during all periods. The collocate frequencies are thus different for each modal auxiliary in each decade, and these differences can be exploited for a diachronic analysis of how the modal auxiliaries have changed.

Whereas data of this kind lends itself in principle to a diachronic collostructional analysis, this particular dataset is analyzed through a different technique that is described in Hilpert (2011): With the help of multi-dimensional scaling (Wheeler 2005), differences in the collocational profiles of the modals are visualized as two-dimensional graphs, in which auxiliaries with similar profiles are

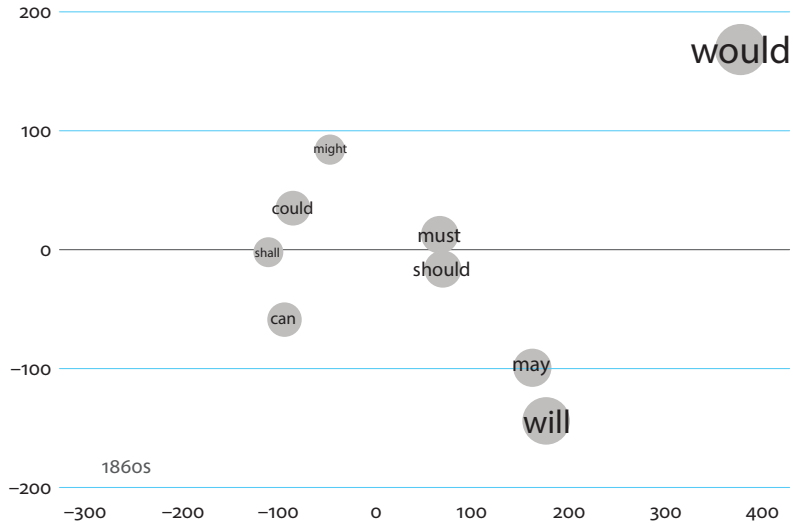

Figure 1. An MDS analysis of the English modals in the 1860s, based on collocate frequencies

placed in close proximity, whereas auxiliaries with very different profiles are placed at a greater distance. Figure 1 offers an illustration of this technique and shows the mutual similarities of the nine core modals in the COHA period of the 1860s.

The graph should be read in the following way. First of all, the modals are represented by circles that differ in size. Circle size indicates text frequency. As can be seen, *would* is the most frequent auxiliary whereas *shall* is the least frequent one. The relative positions of the circles are a consequence of the collocational profiles that the modals exhibit. The MDS algorithm quantifies the difference between each possible pair of modals and then tries to preserve those differences as accurately as possible by placing all nine modals on a two-dimensional coordinate system. The resulting graph thus visualizes the frequency data in such a way that similarities and differences between the collocational profiles of the modals can be inspected. For instance, it is readily apparent that *must* and *should* have rather similar profiles, whereas *would* and *shall* differ considerably. This kind of visualization can be used to analyze the modal auxiliaries with regard to their respective meanings. According to the so-called distributional hypothesis (Turney & Pantel 2010: 153), words that occur in similar contexts tend to be semantically related. If we adopt this hypothesis and assume that relative proximity on the graph translates into a close semantic relation, we could for instance point out that *must* and *should* both encode obligations and that *could* and *might* both encode possibilities. Also, the graph suggests that *would* is semantically distinct from the rest of the modal auxiliaries. Hilpert (2013b: 74–77) proposes that the y-axis of Figure 1

can be interpreted as a cline from deontic modality (in the lower half) to epistemic modality (in the upper half), and that the x-axis maps onto a continuum that ranges from informational (on the left) to interpersonal (on the right). This continuum has been identified by Biber (1988: 115) as one important dimension of textual variability, and it seems to contribute to the differences in collocational profiles that distinguish the different modal auxiliaries.

The COHA data does not only allow the analyst to make comparisons of the nine modals during a given decade, but crucially it also affords a study of how the interrelations of the modals have changed over time. Do the relative similarities between the modal auxiliaries stay the same between the 1860s and the 2000s? A diachronic analysis of the COHA data does not only compare each modal against all eight other modals, but also each modal against itself in all fifteen decades. In this way, *could* from the 1860s is compared against *could* from the 1870s, the 1880s, and so on. The result of such an analysis is a series of two-dimensional plots that can be viewed in sequence, as a dynamic visualization of language change. Figure 2 highlights one development in the changes that the modals have undergone between the 1860s and the 2000s that is particularly noteworthy. This development concerns the modal *may*, which displays a systematic trajectory of collocational change that takes it from the lower right quadrant of the graph into the upper left quadrant. In Figure 2, this trajectory up to the 2000s is visualized through increasingly darker shadings of the circles that represent *may*. The other eight modals are only shown in the positions and frequencies that they have in the 2000s. A comparison with Figure 1 thus reveals that *will* and *would* have gravitated

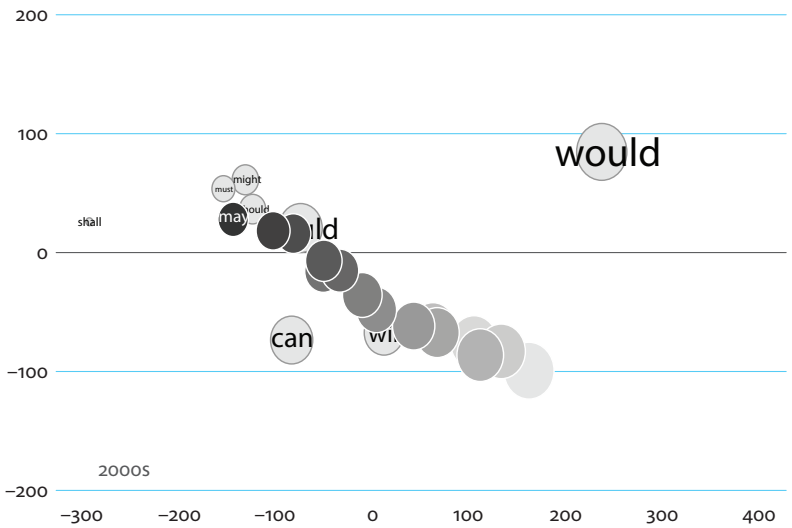

Figure 2. The English modals in the 2000s and the diachronic change of *may*

towards the center, while *shall* has moved to the left periphery of the configuration. In the 2000s, *may* finds itself in a cluster that also includes *must*, *might*, *should*, and *could*.

If the two dimensions of the graph are interpreted as representing continua of deontic vs. epistemic meaning and informational vs. involved textual functions respectively, Figure 2 suggests that over the past 150 years, *may* has become more epistemic (which is a point that has also been argued by Millar 2009) and at the same time more attached to informational types of text. This interpretation is consonant with the observations that can be gleaned from a comparison of the lexical verbs that occur with *may* in the 1860s and the 2000s. Hilpert (2013b: 76) performs a collostructional analysis that identifies verbs such as *say*, *do*, *add*, or *judge* as typical for the 1860s, and verbs such as *help*, *want*, and *need* for the 2000s. The sentences in (2) illustrate representative usage patterns for the 1860s; the examples in (3) are from the 2000s. These sets of examples differ in the expected ways along the dimensions of deontic vs. epistemic modality and informativity vs. involvedness, showing a tendency away from deontic and interpersonally involved meaning towards more epistemic and informative meaning.

(2) a. However, there are a few old roads that may be trodden with profit.
b. Without vanity, I may say that I succeeded.
c. "You may do so if you like," said Mrs Sortridge composedly.
d. Mrs. Chapman, I may add here, had a great weakness for distinctions.
e. There has been much festivity in this saloon, if I may judge by the character of its frescos.

(3) a. I may have told you that Miller bailed out.
b. If the hives are itchy, antihistamines may help.
c. The police may want to speak with you.
d. If fillets are large, you may need to cook them in two batches.
e. This is not so radical a step as it may sound.

Summing up, through corpus-based observations of diachronic collocational shifts, both in the entire set of the English modals, and in more specific contrasts of the same modal auxiliary at different points in time, the analysis in Hilpert (2013b) identifies relevant dimensions of meaning change and illustrates how *may* has changed over time. The analysis that is presented in the next section takes these results as a basis, but tries to explore the development of *may* at a finer level of detail that offers a more comprehensive look at the changing interconnections between *may* and the lexical verbs with which it occurs.

4. Another look at the shifting collocates of *may*

The present analysis is based on a part of the data set that was used in Hilpert (2013b). As was described above, that dataset is a concordance from the COHA that retrieved instances of *may* that are followed by a verb in the infinitive. The full concordance contains approximately 300,000 tokens from the entire temporal range of the corpus, these tokens instantiate 1,776 different verb types. For practical reasons, the analysis in this section focuses on only a subset of that data, namely the 250 most frequent verb types that are found with *may* (with the deliberate exceptions of the highly frequent verbs *be*, *do*, and *have*). This restriction decreases the number of tokens to about 275,000, so not too much of the data is discarded. The overarching question that guides the subsequent analysis is the following: How has the modal auxiliary *may* changed over the past 200 years with regard to the connections to its most frequent verbal collocates? The collostructional analysis in Hilpert (2013b) already provides part of an answer to that question, but that analysis only highlights the elements that have undergone the most significant changes, the tip of the iceberg so to speak. To complement those results, the study that is presented here provides a more comprehensive view of changes in the collocational network of *may*, thereby trying to bring the whole iceberg into view.

In order to pursue that goal, synchronic data from the COCA corpus (Davies 2008) was used to construct a semantic vector space (Turney & Pantel 2010) for the 250 most frequent verbal collocates of *may*. The motivation for selecting this approach is that it would be desirable to learn more about the semantic space in which *may* is situated and in which it has changed its position over time. The overall methodological procedure and its visual outcome are very similar to the analysis of the nine English modals that has been presented above in Figure 1. A number of linguistic elements are compared on the basis of frequencies of their collocates, and a scaling algorithm transforms the mutual differences into an arrangement of those elements on a two-dimensional map. However, in the present study, the data that goes into the analysis and the visualizations that come out as a result are more complex. First of all, we are dealing with a dataset of 250 different elements, not just nine. Second, whereas the nine modals were characterized just in terms of the lexical verbs that occurred with them, the present approach casts a wider net. Each verb is characterized in terms of a frequency vector that includes all of its collocates that are found within a context window of four words to the left and four words to the right. This necessitates a number of additional processing steps, such as the exclusion of stop words and the restriction to a set of collocate types with a minimum frequency. In more specific detail, the analytical steps that were taken to create the semantic vector space of the 250 verbs were the following:

1. A representative 50 million word sample of the COCA was created as the reference corpus.
2. For each of the 250 verbs that occur frequently with *may*, a concordance with four words to the left and four words to the right was retrieved from the corpus. This yielded for each verb a 'bag of words' that forms the basis for its collocate frequency vector.
3. From those concordances, punctuation and 150 highly common words were removed, using a stopword list.[1] This procedure helps to reduce noise in the statistical analysis.
4. All collocates that occurred less often than 100 times combined in the retrieved concordances were removed. This step reduced the overall number of collocates entering the analysis to 5,504 different words.
5. The data was arranged in a table in which the verb types were the column labels (250) and their collocates were in the rows (5,504). The cells of that table were filled with the observed co-occurrence frequencies of all verbs with all collocates. For instance, the verb *accept* co-occurs 9 times with the word *decision* and 26 times with the word *difficult* in the corpus when a text window of 4L and 4R is used.
6. The raw co-occurrence frequencies in the table were adjusted by applying positive pointwise mutual information, or PPMI for short (Church & Hanks 1989). This is done to control for the differences in text frequency between the database items. The subsequent analysis is done on the basis of a table in which the co-occurrence frequency values are replaced by PPMI values. For the verb *accept*, the word *decision* yields a PPMI value of 1.68 and the word *difficult* a PPMI value of 2.41.
7. On the basis of the table of PPMI values, a distance matrix was computed using the cosine distance. Each verb is thus compared to each other, and cosine distances for each pair are computed on the basis of the respective pairs of PPMI value vectors.
8. Metric multidimensional scaling was used to transform the cosine distance matrix into a two-dimensional representation of the semantic vector space.

The result of this multi-step procedure is shown in Figure 3, which displays the 250 different verbal collocates of *may*. In that graph, verbs that occur with similar words at similar frequencies are placed in close mutual proximity. According to the distributional hypothesis, the graph should be interpretable in terms of meaning, so that semantically related verbs cluster together in the same area. As will

1. The list of stop words, as well as the full R code that is used in the present analysis, are available from the author upon request.

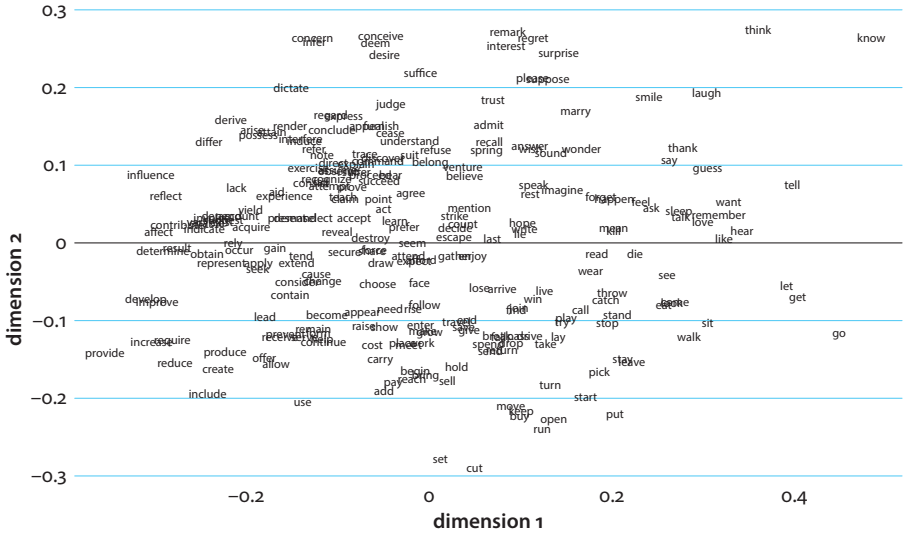

Figure 3. A semantic vector space of the 250 most frequent verbal collocates of *may*, based on data from the COHA

be discussed in more detail below, the graph indeed reflects semantic structure. However, it has to be kept in mind that lexical verbs are semantically rich and differ along many dimensions: concreteness, agentivity, transitivity, and dynamicity are only some of the semantic categories that figure prominently in the literature on verbal meaning. Given this variety, it is not surprising that the overall amount of variance that is accounted for in Figure 3 is low. The first two dimensions of the MDS solution account only for 14.3% of the entire semantic variation between the verbs that are shown. The following paragraphs will discuss which aspects of verbal meaning are captured by the graph.

For a first orientation in the semantic vector space of the verbal collocates of *may*, it is instructive to inspect the graph for pairs of verbs that are placed in close proximity and that are semantically similar. For example, there are verb pairs such as *throw* and *catch* (x = 0.2, y = −0.1), *offer* and *allow* (x = −0.2, y = −0.15), or *smile* and *laugh* (x = 0.3, y = 0.2), which appear close to each other and which hence indicate that the analysis is picking up a statistical signal that reflects human intuitions. Beyond pairs like *smile* and *laugh*, it is possible to see more general semantic distinctions in the graph. What is striking is that the lower right quadrant is mainly occupied by verbs that denote concrete physical actions such as *run*, *open*, *put*, *pick*, *leave*, *sit*, *walk*, and so on. By contrast, the upper left quadrant features verbs such as *concern*, *differ*, *influence*, *reflect*, *derive*, *render*, and others, which describe processes that are less concrete and that are not necessarily tied to a volitional agent that carries out an action. In the upper right, there are the verbs *think* and

know as well as a number of speech act verbs such as *tell*, *say*, and *thank*, which are volitional and concrete, but do not describe physical actions. In the lower left, there are verbs such as *provide*, *create*, *increase*, *reduce*, or *produce*, which can denote physical actions that are however much more abstract and schematic than *put* or *walk*. On the whole then, the semantic landscape of *may* that is reflected in this graph suggests that two dimensions of meaning play important roles. On the x-axis, there is a continuum from relatively abstract meanings on the left to relatively concrete meanings on the right. This is in line with the observation that the left side of the graph features a lot more 'long words' such as *determine* or *represent* than the right-hand side. On the y-axis, we see a continuum from volitional and potentially physical actions at the bottom (*use*, *cut*, *put*, etc.) to mental and partly involuntary processes at the top (*concern*, *desire*, *interest*, *know*, etc.). While individual verbs on the graph, as for example *act* (x = –0.05, y = 0.05) may show a less than optimal fit with this characterization (*act* would have been expected to be in the vicinity of volitional and physical verbs such as *put*), it has to be kept in mind that frequent verbs are usually polysemous, so that not all occurrences of *act* describe concrete volitional actions. The above characterization is therefore adopted as a working hypothesis of how the semantic landscape of *may* is organized.

The crucial question to be addressed in this study is how *may* has changed diachronically with regard to its relation to this semantic landscape. In order to explore this question in more detail, the semantic vector space shown in Figure 3 needs to be enriched with diachronic information. Here we come back to the COHA frequencies of *may* plus infinitive that were mentioned in the beginning of this section. For each of the 250 verbs in Figure 3, the COHA provides diachronic frequency information — we can easily find out whether strings such as *may use*, *may need*, or *may determine* increased or decreased in frequency. It is thus possible to investigate which of the verbs in Figure 3 have strengthened their connections to *may* over time. Figure 4 visualizes those shifting connections by means of three density maps.

The maps indicate which areas of the semantic landscape have been particularly densely populated during three sequential periods of time. The contour lines in the graphs represent the relative density of token frequencies. Frequent verbs lead to higher density in a given area, which is then represented by increasingly darker lines that mark frequency 'peaks' in the landscape. Several such peaks can be seen in the graphs. Some of these peaks are caused by single high-frequency verbs, as for instance the verb *go* during the first period (1800s-1860s), which is situated in the lower right corner (x = 0.45, y = –0.15). Other peaks are in fact semantic 'plateaus' that involve several verbs. In the first period, there is a heavily overplotted peak area in the upper left (x = –0.2, y = 0) that includes the verbs *depend*, *exist*, *involve*, *enable*, and *indicate*. The combined representation of the

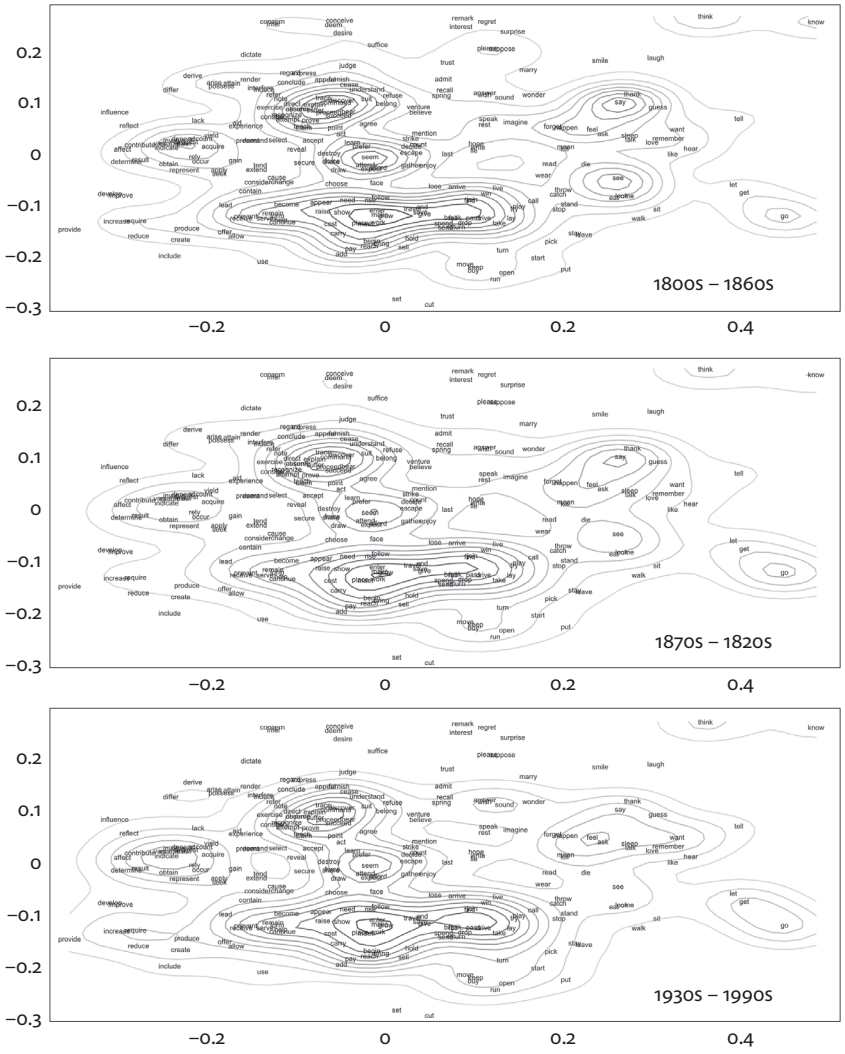

Figure 4. Density changes in the semantic vector space of *may*

different verb types that occur with *may* and the density plots of their aggregated token frequencies allow for a thorough inspection of the way in which *may* is used in the COHA data. It is apparent that verbs that may be semantically central (as for instance *act*) are not necessarily very frequent in their co-occurrence with *may*, and vice versa (cf. *think* or *go*). It can also be gleaned that the semantic landscape of *may* is pluricentric. Instead of a single group of verbs acting as a solitary attractor, there are several peaks that reflect the varied and heterogeneous meaning potential of *may*.

Through a comparison of the three panels in Figure 4, continuities and changes in the collocational preferences of *may* can be identified. As could perhaps be expected, the overall frequency profile of the semantic landscape remains relatively constant. The three panels do not look radically different, indicating relative temporal stability in the collocational profile of *may*. Some changes are nonetheless worthy of discussion. The right side of the graphs shows the gradual demise of two fairly frequent verbs, namely *say* and *see*. The collocations *may say* and *may see* form clear peaks in the first period, but those flatten out as time goes on. Earlier in this section, the right-hand side of the graph was characterized in terms of relatively greater concreteness. The demise of concrete verbs is in line with the observed trajectory of *may* into more informational and less involved texts that was discussed in Section 3. However, not all areas of receding density are found on the right side of the graphs. In the upper left quadrant, the abstract and non-volitional verbs *concern*, *infer*, *conceive*, *deem*, and *desire* form a low semantic plateau in the first period. Over the subsequent decades of the COHA, that plateau disappears, and also the first density line, found just below the verbs *dictate* and *suffice* in the first period, recedes continually downward. This development does not find a satisfactory explanation in anything that has been discussed above or in the analysis of Hilpert (2013b). Some collocational shifts, it appears, happen in relative independence from more general trends.

Turning towards developments of increasing density, the following remarks can be made. In the upper left quadrant of the graphs, the peak area that includes *depend*, *exist*, *involve*, *enable*, and *indicate* ($x = -0.2$, $y = 0$) rises and extends over time. Abstract verbs such as *affect* and *determine* ($x = -0.3$, $y = 0$), or *apply* and *seek* ($x = -0.2$, $y = -0.05$), which in the first period remain at the very margins of that plateau, are eventually incorporated into it. The expansion of this semantic plateau is an expression of the shift of *may* into more informational textual genres. Another development in the upper left quadrant concerns a disappearing 'trough' between two peaks that can be seen in the first COHA period. In the first period, the peak of *seem* ($x = 0$, $y = 0$) and the small plateau that includes *explain*, *command*, and *discover* ($x = -0.07$, $y = 0.1$) are divided by nine contour lines. In the third period, the path from *seem* to *explain* only crosses five contour lines. The collocation *may seem*, which is a solitary peak in the first period, decreases in relative prominence over time, as other verbs in the semantic neighborhood of *seem* gain in their collocation frequency with *may*. This development can be interpreted in terms of the strengthening of epistemic meanings in the meaning pole of *may* plus infinitive.

To sum up these observations, the graphs in Figure 4 offer a perspective on the shifting collocates of *may* that is complementing other approaches that offer either a 'bird's eye view' (e.g. the development of *may* in Figure 3 that averages over all

collocates) or what could be called a 'tip of the iceberg view' (e.g. the collostructional approach that singles out those collocates that exhibit the most substantial diachronic asymmetries). It goes without saying that each approach has its own benefits and shortcomings, and that they are most fruitful when used in combination. In fact, their combined use may help to distinguish actual linguistic changes from mere chance fluctuations in the corpus data. The benefits of methodological pluralism and converging evidence (Arppe & Järvikivi 2007; Arppe et al. 2010) apply here as well as elsewhere. An encouraging conclusion is that in the case of *may* and its changes, the three approaches used in this paper yield results that are compatible and mutually enriching.

5. Concluding remarks

This paper has argued for a theoretical understanding of grammatical constructions that explicitly makes reference to their associations with a large range of lexical items. The idea that grammar is tied to lexis is in fact a core assumption of usage-based linguistics (Diessel 2011: 834), so that it could be asked whether presenting an extended argument for that idea is perhaps a moot point. The question is certainly justified, but it has to be conceded that even among those members of the Construction Grammar community that hold this view, there is no established consensus about its practical consequences. The most popular definitions of constructions remain those of Goldberg (1995) and Goldberg (2006), which invite an understanding of grammatical constructions as 'schemas with slots', rather than 'networks of connections', even though the latter is of course very much in line with Goldberg's ideas about how linguistic knowledge is organized in speakers' minds. The present study has therefore aimed to illustrate, with the example of a modal auxiliary construction, how a more network-oriented analysis of grammatical constructions could proceed. Modal auxiliary constructions, such as *may* plus infinitive, are a particularly fitting example for this purpose, because they fail several criteria of constructionhood that are typically used as reference points: With the core modal auxiliaries, it is hard to argue for construction status on the basis of non-compositional meanings, idiosyncratic formal characteristics, or unpredictable constraints. Yet, it is possible to show that modal auxiliaries such as *may* entertain a complex network of associations with lexical verbs, which motivates a constructional view of modal auxiliaries. The empirical sections of this paper have presented a number of different approaches that can be used for the analysis of associations between constructions and lexical items, with a particular focus on diachronic change in these associations. With regard to *may*, corpus-based analyses of shifting collocational preferences are in line with the interpretation that *may* has

been undergoing continuous semantic change, away from deontic modal meanings and 'involved' textual uses towards epistemic meanings and a higher degree of informativeness. None of these developments implicate any tangible change in morpho-syntactic form, and with regard to the change at the semantic pole of the construction it is actually a matter of debate whether any of the 'new' meanings are genuinely new, or just meanings that were already present and have merely gained in relative importance. Nevertheless, rearrangements of this sort do constitute changes in speakers' knowledge of language, and by virtue of that instantiate constructional change (Hilpert 2013a: 16). Constructional analyses of modal auxiliaries can therefore benefit from a perspective in which collocational networks are in the focus of attention.

References

Arppe, A., & Järvikivi, J. (2007). Every method counts - Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3(2), 131–159.

Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., & Zeschel, A. (2010). Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. *Corpora*, 5(2), 1–27. doi: 10.3366/cor.2010.0001

Bergs, A. (2010). Expression of futurity in contemporary English: A construction grammar perspective. *English Language and Linguistics*, 14(2), 217–238. doi: 10.1017/S1360674310000067

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511621024

Boogaart, R. (2009). Semantics and pragmatics in construction grammar: The case of modal verbs. In A. Bergs & G. Diewald (Eds.), *Contexts and constructions* (pp. 213–241). Amsterdam: John Benjamins. doi: 10.1075/cal.9.09boo

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173. doi: 10.1075/ijcl.20.2.01bre

Bybee, J. (2010). *Language, usage, and cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526

Bybee, J., & Fleischman, S. (1995). *Modality in grammar and discourse*. Amsterdam: John Benjamins. doi: 10.1075/tsl.32

Bybee, J., & Pagliuca, W. (1987). The development of future meaning. In A. Giacalone Ramat, O. Carruba, & G. Bernini (Eds.), *Papers from the 7th International Conference on Historical Linguistics* (pp. 109–122). Amsterdam: John Benjamins. doi: 10.1075/cilt.48.09byb

Bybee, J., Pagliuca, W., & Perkins, R. D. (1991). Back to the future. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization*, Vol. II (pp. 17–58). Amsterdam: John Benjamins. doi: 10.1075/tsl.19.2.04byb

Church, K., & Hanks, P. (1989). Word association norms, mutual information, and lexicography. Proceedings of the *27th Annual Conference of the Association of Computational Linguistics*, Vancouver, British Columbia (pp. 76–83).

Coates, J. (1983). *The semantics of the modal auxiliaries*. London: Croom Helm.
Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 400+ million words*. Available online at http://corpus.byu.edu/coca.
Davies, M. (2010). *The Corpus of Historical American English (COHA): 400+ million words, 1810-2009*. Available online at http://corpus.byu.edu/coha.
Diessel, H. (2011). Review article of 'Language, usage and cognition' by Joan Bybee. *Language*, 87(4), 830–844. doi:10.1353/lan.2011.0082
Fillmore, C., Kay, P., & O'Connor, M.C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64, 501–38. doi:10.2307/414531
Fischer, O. (2007). *Morphosyntactic change: Functional and formal perspectives*. Oxford: Oxford University Press.
Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Godlberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Goldberg, A. E., & van der Auwera, J. (2012). This is to count as a construction. *Folia Linguistica*, 46(1), 109–132. doi:10.1515/flin.2012.4
Gries, S. Th., & Stefanowitsch, A. (2004a). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9(1), 97–129. doi:10.1075/ijcl.9.1.06gri
Gries, S. Th., & Stefanowitsch, A. (2004b). Co-varying collexemes in the *into*-causative. In M. Achard & S. Kemmer (Eds.), *Language, culture, and mind* (pp. 225–236). Stanford: CSLI.
Hilpert, M. (2008). *Germanic future constructions. A usage-based approach to language change*. Amsterdam: John Benjamins. doi:10.1075/cal.7
Hilpert, M. (2011). Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics*, 16(4), 435–461. doi:10.1075/ijcl.16.4.01hil
Hilpert, M. (2013a). *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139004206
Hilpert, M. (2013b). Die englischen Modalverben im Daumenkino: Zur dynamischen Visualisierung von Phänomenen des Sprachwandels. *Zeitschrift für Literaturwissenschaft und Linguistik* 42, 67–82.
Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The *What's X Doing Y?* construction. *Language*, 75(1), 1–33. doi:10.2307/417472
Langacker, R. W. (1987). *Foundations of cognitive grammar*, Vol. 1. Stanford: Stanford University Press.
Millar, N. (2009). Modal verbs in TIME. Frequency changes 1923-2006. *International Journal of Corpus Linguistics*, 14(2), 191–220. doi:10.1075/ijcl.14.2.03mil
Plank, F. (1984). The modals story retold. *Studies in Language*, 8(3), 305–364. doi:10.1075/sl.8.3.02pla
Sagi, E., Kaufmann, S., & Clark, B. (2011). Tracing semantic change with latent semantic analysis. In J. Robynson & K. Allan (Eds.), *Current methods in historical semantics* (pp. 161–183). Berlin/New York: Mouton de Gruyter.

Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–43. doi: 10.1075/ijcl.8.2.03ste

Taylor, J. R. (2012). *The Mental corpus. How language is represented in the mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199290802.001.0001

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard: Harvard University Press.

Traugott, E. C. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, 57(1), 33–65.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188.

Visser, F. Th. (1963). *An historical syntax of the English Language*. Leiden: Brill.

Wheeler, E. S. (2005). Multidimensional scaling for linguistics. In R. Koehler, G. Altmann, & R.G. Piotrowski (Eds.), *Quantitative linguistics. An international handbook* (pp. 548–553). Berlin: Walter de Gruyter.

Author's address

Université de Neuchâtel
Institut de langue et littérature anglaises
Espace Louis-Agassiz 1
CH-2000 Neuchâtel

mhilpert@gmail.com
martin.hilpert@unine.ch