

Shifts in Collocational Preferences

Welcome back to Ten Lectures on Diachronic Construction Grammar. In the lecture this morning, one of the issues that I was concerned with was the open question in Diachronic Construction Grammar of how we can make our analysis more focused on the connections between constructions, rather than maintaining a focus on the internal structure of the nodes in that network. I called that the “fat node problem”. To address that problem, I have discussed the example of the English auxiliary *may*, which has over time come to be used with a very different set of lexical collocates. You remember the journey of *may* in the system of modals that reflects its changing connections with the lexical constructions that it occurs with. In this lecture, I want to expand on the issue of shifts in collocational preferences. I want to discuss a number of case studies that allow us to assess the theoretical conclusions that we can draw from observing shifts of this kind.

One important point is that in the development of grammatical constructions, these collocational shifts are not just random events. In the lexical domain, by contrast, anything can happen. You remember from yesterday the collocate display of *gay*, and its early collocates like *gay colours* and its later collocates like *gay community*. The collocational shifts show us that some meaning change has taken place, but there aren't any broader implications of this beyond that. This is what happened to the adjective *gay*, and other lexical elements may change in completely different ways.

With grammatical constructions, however, shifts and collocates tend to reflect more systematic patterns of change, such as the ones proposed by grammaticalization theory.

Remember also the proposal by Traugott and Trousdale (2013) that grammatical constructionalization involves an increase in schematicity. We expect grammatical constructions to broaden in their collocational behavior over time. We expect them to occur with more types of lexical elements. We expect



All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: <https://doi.org/10.6084/m9.figshare.13691059>.

- Israel, Michael. 1996. *The way constructions grow*. In Adele E. Goldberg (ed.), *Conceptual structure, discourse and language*, 217-30. Stanford: CSLI Publications.

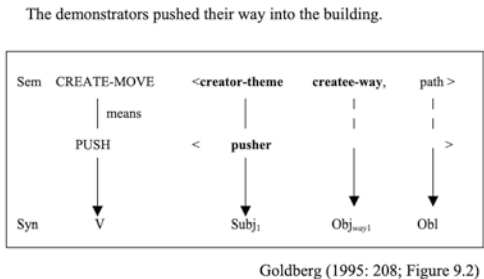


FIGURE 1

these types to have different meanings, or an increasing range of lexical meaning. That is the kind of phenomenon that I want to look at in this lecture. With all of that in mind, let's get on the way. Coincidentally, what got me on my way with regard to Diachronic Construction Grammar was a paper by Michael Israel (1996) with the curious title “The *Way* Constructions Grow”. This title very cleverly blends two ideas. On the one hand, the paper addresses by now well-known pattern that is called the *way*-construction. It presents a study of *way*-constructions and how these constructions grow. At the same time, the paper explains how this process unfolded, as it discusses the way in which constructions grow. It is terrible of me to explain the joke, you simply have to forgive me.

What you see on the slide is an example and a representation of the *way*-construction, taken from the work of Adele Goldberg (1995). The example is “*The demonstrators pushed their way into the building*”. The diagram, the box with the semantics and syntax and the arrows and the arguments, that is meant to capture that the construction can take a verb such as *push*, that has the meaning of creating something and moving something at the same time. The construction adds two arguments that are not normally called for by the verb. That is what the dotted lines mean. *Push* requires a pusher, i.e. the subject. But the two other constituents are actually added by the construction. They are not part of the ordinary argument structure of the verb. The first of the two arguments that are added by the construction is an object that encodes a created way. Goldberg calls this the createe-way argument. Then there is a

prepositional phrase, Goldberg calls it the path here, that expresses either a path or a location. In this case, it is a path: “into the building”.

How did Michael Israel analyze this construction? His starting point is an observation that Goldberg made about the construction. The construction is polysemous. It can express two basic ideas. It can express manner of motion as well as a means of motion. Manner of motion is expressed in examples such as “*The wounded man limped his way across the field*”. The movement is carried out in a limping, effortful manner. By contrast, in an example such as “*Joe cheated his way into law school*”, cheating is a means to achieve a result. There is a metaphorical movement. Joe is metaphorically moving into law school and cheating allows the subject to carry out that movement. The *way*-construction is what Goldberg calls an argument structure construction, which is to say that in synchronic usage, the construction can modify the argument structure of the verb that it takes. It imposes an unusual argument structure on the verb, in these examples the verbs *limp* and *cheat*. You notice that I can use the verbs *limp* and *cheat* intransitively. I can say “*The man was limping*” or “*John, he is always cheating*”. There are other argument structures that go with these verbs, but I can use them intransitively, and the *way*-construction allows me to use new and additional arguments with these verbs, namely a created way argument and a path or a goal argument.

Michael Israel looked at this construction by tracing its usage history in historical texts. He used a particular data resource for this, the Oxford English Dictionary, which is a dictionary that not only lists words and definitions. It also provides authentic bits of texts that illustrate how these words are used. What he found was that all the examples from the Oxford English Dictionary show that the construction established itself first with verbs that encode paths. What he found was that the early manner interpretation corresponds to examples with general motion verbs. Here is an example of this, “*The kyng took a laghtre and wente his way*”. *Went* is just the past tense of *go*, a movement verb, the king laughed and went his way. The construction thus emerges in a very specific lexical context with verbs that harmonize with the overall meaning of the construction. That is not only true of the manner interpretation. It is also true of the means interpretation. Here the construction is used early on with verbs that explicitly encode the creation of path. Here we have an example, “*Arminius paved his way*”. You pave the surface and after that you have a way to move along on.

Now, looking at these examples at this stage, you could argue that the *way*-construction has not fully constructionalized in the sense of Traugott and Trousdale, but rather what we have here is a more or less transparent

compositional use of a particular argument structure and verbs that harmonize with that argument structure. This, if you like, is the *way*-construction coming into being, constructionalizing in the sense of Traugott and Trousdale.

As time goes on, the construction goes through a type frequency increase. More and more different verb types occur in that particular argument structure pattern with the manner interpretation. Michael Israel counted all the verbs. Up to 1700, the manner interpretation is relatively sparse in the data. There are only 16 different verb types, including *go*, *pass*, *run* and a few others. Then, from the 19th century onward, there are 38 more that he finds. *Worm* here means to move like a worm. *Fumble* is a slow and clumsy movement. With regard to means, this starts earlier, so Israel finds examples as early as 1650. There we have the path-creation verbs such as *pave* or *smooth* or *cut*. You cut your way through a jungle or through some difficult terrain. Later on in the 18th century, he finds verbs that metaphorically extend this kind of path-creating idea. There is *battle*, there is *fight*. Then from the late 19th century onward, we find verbs that we find also nowadays in the *way*-construction, verbs like *elbow*, *shoot*, *spell*, and even things like *write*. You can “write your way out of a difficult situation”. You’ve offended someone and you write them an apology. You can write your way out of a hot mess that you’ve gotten yourself into. The increasing type frequency maps onto increasing degrees of schematicity and abstraction. This instantiates one of the pathways that Traugott and Trousdale proposed for grammatical constructionalization, i.e. increases in productivity, increases in schematicity, and decreases in compositionality. When we have something like “*He elbowed his way out of the subway*”, the meaning component of difficult laborious motion is carried holistically by the entire construction, rather than by any individual element on its own. When I read Michael Israel’s paper, I took away three lessons. Those lessons for me were the following.

First of all, I understood that constructionalization happens in the context of particular collocations. There are certain words and constructions that are used together. This starts the process. Second, shifts in collocational preferences reflect changes in constructional meaning. As we use a construction with more and more different verbs, our idea of what this construction can do changes along with it. Constructional change, and that is the third conclusion, shows itself in changing relative frequencies and type frequencies of lexical collocates, not necessarily in their absolute frequencies. One curious thing about the *way*-construction is that for the longest time, *make* actually has been the most frequent verb in this construction. That has been relatively constant. But there have been a lot of developments going on under the surface of the most frequent elements.

There is one further conclusion from the paper that I would like to read to you in the words of Michael Israel himself. This captures very much the way I have come to think about constructional change. Here it is.

The way-construction emerged gradually over the course of several centuries. There is no single moment we can point to and say, 'This is where the construction entered the grammar.' Rather, a long process of local analogical extensions led to a variety of idiomatic usages to gradually gain in productive strength even as they settled into a rigid syntax. As the range of predicates spread, increasingly abstract schemas could be extracted from them and this in turn drove the process of increasing productivity.

Here Michael Israel actually prefigures the main aspects of Traugott and Trousdale's concept of constructionalization, with its aspects of compositionality, schematicity and productivity. More specifically, the development of the *way*-construction clearly illustrates the phenomenon of what I talked about as host-class expansion yesterday and earlier today. I would say that Michael Israel's paper on the *way*-construction was one major inspiration for the research that I then did in my doctoral dissertation.

In my dissertation, I investigated future tense constructions across a range of languages from the Germanic family. The main overarching aim of that book was to see if we can use shifting collocational preferences to study on-going semantic change. Here is where the theoretical parts of what I have been talking about so far meet the methodological parts. I chose future constructions for my study because a lot is known about the grammaticalization of future tense markers. They tend to come from a handful of lexical sources such as movement verbs (English *be going to*), verbs of desire (English *will*, which derives from a verb meaning 'want' and 'desire'), or verbs of obligation (English *shall* falls into that category). There are verbs that mean 'turn' or 'change'. There is no English construction for that, but German has one, namely *werden*, which meant 'turn' originally. There are verbs that express intentions or at least relate to intentions. Swedish, for example, has a future construction with a verb that means 'think'. When I say *I think going to the movies*, it over time turns into an expression that tells my interlocutor I will be going to the movies later today.

Grammaticalization scholars have proposed very specific developmental pathways for these constructions. Movement verbs are known to turn into markers of intention, and after that into markers of future time, and after that into yet other meanings. I was curious to see whether these proposed grammaticalization pathways could be shown to be reflected in historically shifting

patterns of collocations. This turned out to be true. You can actually see these trajectories.

My study compares future constructions from five different Germanic languages. It is based on historical corpus data exploring both the synchronic meaning and the historical development of these future constructions. It applies a method that I briefly talked about this morning, collostructional analysis. I will say more about this as we go along here. The method allows us to measure how a particular future marker is connected to lexical verbs that occur with it. It allows us to determine what associations exist and how strong they are.

More importantly, we can also use it to investigate how these patterns of association change over time. Michael Israel looked at this from a qualitative point of view. He took examples from different historical periods. He noted what kind of verbs came first, what kind of verbs were added later, and how the changes reflected increasing degrees of productivity. When I say from a qualitative point of view, that is not entirely true. He did count the types, but that very much stays at the level of descriptive statistics, not inferential techniques.

In my studies, I have been fortunate to have two great teachers and mentors, Anatol Stefanowitsch and Stefan Gries, who developed a technique for the analysis of collocational relations between constructions and lexical items. This method became available just when I needed it. It fell into my lap at the exactly right time. I am of course talking about collostructional analysis. This is a method that allows you to quantify how strongly a set of lexical items is associated with a grammatical construction that has an open slot for these lexical items. Stefanowitsch and Gries developed collostructional analysis for the study of such associations in synchronic present-day usage. But one day, when I was riding my bicycle home from university, I wondered whether it would be possible to tweak the method just a little bit, so that I could use it to study change over time.

I wrote an email to Anatol later that day and asked him whether he thought it could be done. He emailed me back and told me to try it. As a dutiful student, I followed his advice. I developed the idea and publish a short paper with the title “Distinctive Collexeme Analysis and Diachrony” (Hilpert 2006), and in the same issue of the journal then Anatol published a short reply (Stefanowitsch 2006), summarizing all the points that he found problematic about it. I am getting ahead of myself here. Let me explain what collostructional analysis is all about.

The basic idea is one that I mentioned in lecture one, namely that constructional meaning is reflected in associations between syntactic patterns and lexical elements. The fact that *give* is the most frequent verb in the ditransitive

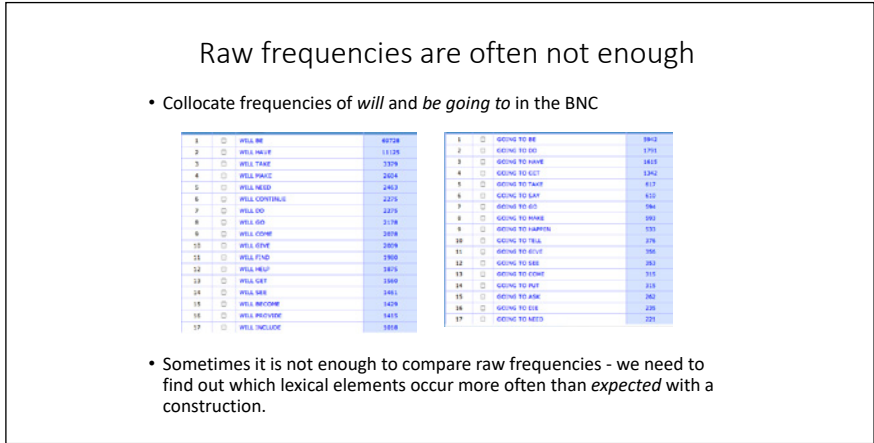


FIGURE 2

construction, and the fact that we find the verb *elbow* in the *way*-construction, that illustrates this harmony in meaning between grammatical constructions and the lexical items that occur within them. Let me just read this to you again. Stefanowitsch and Gries (2003: 236) motivate this in the following way: “*If syntactic structures served as meaningless templates waiting for the insertion of lexical material, no significant associations between these templates and specific verbs would be expected*”. But no matter where we look, pretty much any syntactic pattern that has been investigated in this way shows asymmetries that demonstrate that these distributions differ from chance in many ways. There is no random distribution of lexical elements across syntactic constructions.

This also relates to the controversy of collocation analysis versus raw frequencies. This is actually a good moment to explain what is really at the heart of this controversy. Sometimes it will turn out that when you look at the lexical elements in a grammatical construction, the raw frequencies won't tell you a great deal. Sometimes looking at raw frequencies is just not enough. As an example for that, this slide presents two lists of collocate frequencies from two near-synonymous grammatical constructions of English, namely the English *will* future and the English *be going to* future, and the most frequent lexical verbs that occur within them. If you go through these two lists, you notice that they are almost identical. We start with very frequent verbs like *will be*, *will have*, *will take*, *will make* and so on and so forth. With *going to*, we have *going to be*, *going to do*, *going to have*, *going to get*. Basically, you're just getting these long lists of semantically light, very general, very frequent verbs. You might look at these two lists and conclude that they are almost indistinguishable. These two constructions have similar functions, so they occur with similar sets

of verbs. However, if we find out which elements are not only frequent, but actually more or less frequent than expected, then we can say something more about these constructions. I mentioned collocation analysis briefly earlier this morning. Let me go back to this and explain the general logic.

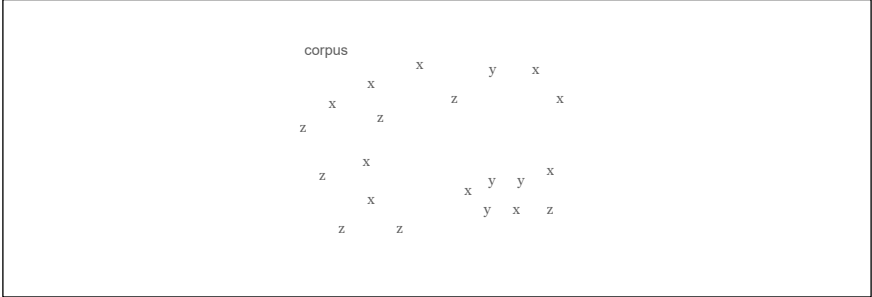


FIGURE 3

Let's say that you have a corpus that has lots and lots of different word types in it. Here I have symbolized them with letters *x*, *y* and *z*. Let's say that you extract from that corpus all the instances of a construction and count the lexical elements that you find in that construction. In this case, we have a very small sample.

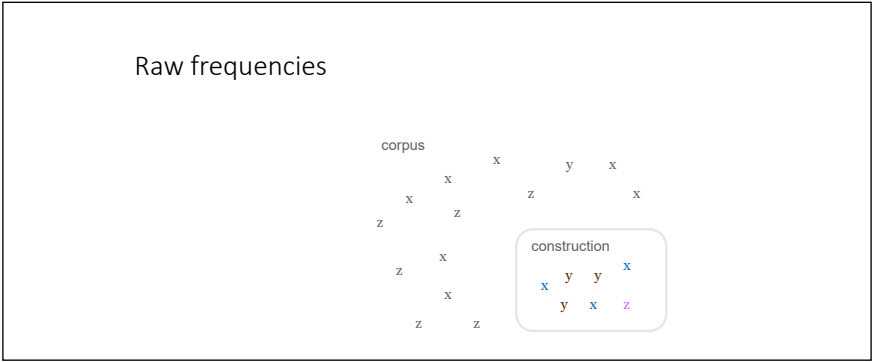


FIGURE 4

The construction occurs with seven lexical items. These seven lexical items fall into three types, *x*, *y* and *z*. Three times *x*, three times *y*, and one *z*. If we were to stay with these raw frequencies, we would say that, for this construction, *x* and *y* are of approximately equal importance.

However, as soon as we compare the frequencies of these items within the construction to the frequencies of these items outside of the construction in the corpus, it becomes apparent that *x* is an element that we find throughout

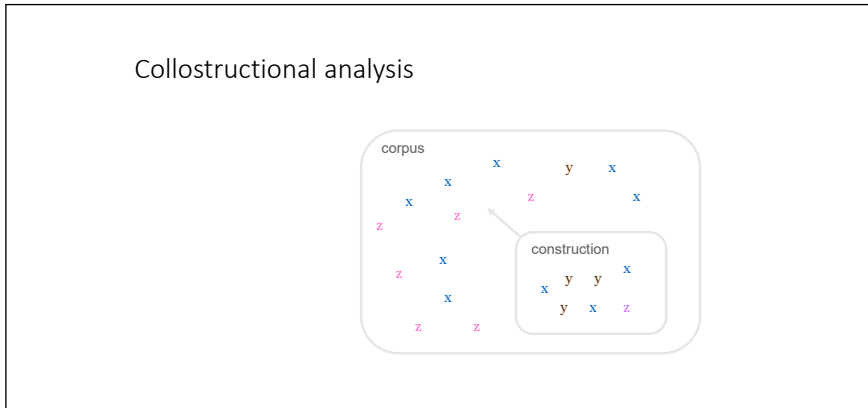


FIGURE 5

the corpus with high frequency. This is in contrast to *y*, which we find three times in the construction, but only once in other contexts. This would be similar to the case of *elbow* in the *way*-construction. *Elbow* is a very infrequent verb. To the extent that we find it in a corpus, it will tend to appear in the *way*-construction, but not anywhere else, at least not very frequently.

This, in a nutshell, brings me to the controversy between collostructional analysis and raw frequencies. The view of focusing on raw frequencies only chooses to ignore how often a lexical element appears not only in the construction, but also in the corpus as a whole.

I have a quote from Gries *et al.* (2005: 665) here, where they say, “*Arguing and theorizing on the basis of mere frequency data alone runs a considerable risk of producing results which might not only be completely due to the random distribution of words [in a corpus], but which may also be much less usage-based than the analysis purports to be.*” It may be questionable to say that words in a corpus are randomly distributed. Language does not work that way, but the point stands that we want to account for these occurrences that occur outside of the construction. Gries *et al.* (2005) have done empirical work that supports the better adequacy of collostructional results over raw frequencies. Let me review that for a minute.

When we have a verb as a cue for a construction, what is it that determines the validity or the usefulness of that cue? According to Bybee’s view, the most frequent verb that appears in a construction should be the best cue for that construction. According to Stefanowitsch and Gries (2003), it might not be the most frequent verb, it might be the verb that is most strongly attracted to the construction that provides the best cue, even if that verb is not very frequent. If all of its instances are found with the construction, as in the case of *elbow* and the *way*-construction, that would make for a very reliable, very useful cue.

Let me give you an example. The most frequent verb in the *way*-construction is the verb *make*. Now, when I say *I made*, does that make you think of the *way*-construction? Does it make you want to continue that sentence with *my way through the city*? *I made* can be continued in lots of other ways. *I made a mistake*, *I made her a sandwich*, and lots of other things. It is not the best cue for the *way*-construction.

By contrast, the verb *elbow*, if I start an utterance with *I elbowed*, there are not many ways to continue that sentence. *I elbowed my way out of the subway*, that would be a very natural continuation of that sentence fragment. This is the methodology that Gries and colleagues applied to investigate which verbs in a carrier phrase would prompt speakers to continue with a specific construction. The construction that they used as an example is a construction that they called the English *as*-predicative construction. It is instantiated by sentences such as "*The idea was perceived as too radical*". We have a verb like *perceive*, and then a prepositional phrase with *as*, and a certain predicate, so *radical* is predicated over the idea.

Here are three examples of the *as*-predicative construction: *The proposal was considered as rather provocative*; *I had never seen myself as being too thin*; *California is perceived as a place where everything is possible*. There are different verbs that appear in this construction. One of them is the verb *see*, as in sentence fragments such as *I have never seen* or *I had never seen.*, which you could continue that with the *as*-predicative. Some verbs give you a very strong cue, like the verb *hail*. When I start a sentence fragment with *The idea was hailed*, the *as*-predicative almost forces itself upon my mind. So, some verbs make you think about the *as*-predicative in very direct ways, and other verbs do not. The question now is which verbs are which. What determines cue validity? Is it frequency in the construction, or is it attraction to the construction that we measure via collostructional analysis?

	Relative Frequency	
	High	Low
Surprisingly high frequency in the AS-PREDICATIVE	define, describe, know, recognize, regard, see, use, view	acknowledge, class, conceive, denounce, depict, diagnose, hail, rate
Surprisingly low frequency in the AS-PREDICATIVE	keep, leave, refer to, show	build, choose, claim, intend, offer, present, represent, suggest

FIGURE 6

Gries and colleagues compared four different types of verb for their design. First, they took sets of verbs that are generally frequent in English. These are verbs such as *define*, *describe*, *know*, *recognize* and verbs like *keep*, *leave*, *refer to* and *show*. Some of them are surprisingly frequent in the *as*-predicative, for example *define* and *describe*. Some of them are very frequent in general but surprisingly infrequent in the *as*-predicative. This is true for *keep*, as in *He was kept as a slave*. That does not appear very often. In the right column of this table, you see verbs that are low in frequency. Some of them are surprisingly frequent in the *as*-predicative, for instance *conceive*. It is not a very frequent verb, but I can say “*This was conceived as the solution for that problem*”. *Depict* is another surprisingly frequent collocater, and *hail* is yet another one. Finally, some verbs are infrequent in general, and also surprisingly infrequent in the *as*-predicative construction. This includes *suggest*, “*This was suggested as a possible solution*”.

On the raw frequency view, high frequency verbs should pattern alike, no matter whether they are surprisingly frequent or surprisingly infrequent in the *as*-predicative. When speakers see a sentence fragment with a high frequency verb, there should be a high probability that they continue the fragment with the *as*-predicative.

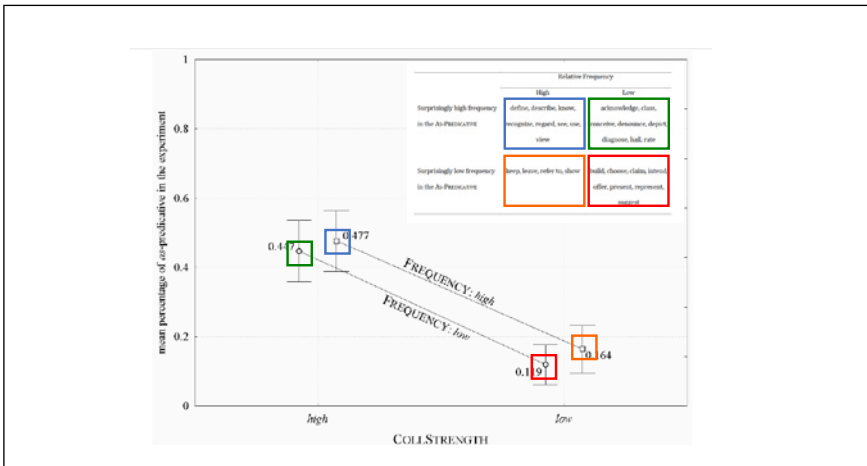


FIGURE 7

On the collostructional hypothesis, however, attraction to the *as*-predicative should matter more than just raw frequency. The verbs that are surprisingly frequent in the *as*-predicative should pattern together with low frequency verbs such as *hail* and *depict*. That means that when speakers see a sentence fragment with these verbs, they should be more likely to continue the fragment with an *as*-predicative construction.

Here’s what came out of the experiment. Here you see a chart (Gries et al. 2005: 659) that shows the rate of completion with an *as*-predicative construction that was obtained in the experiment. The *y*-axis shows you how many of the participants chose to continue a given sentence fragment with the *as*-predicative. You see that there are two types of verb that are higher up and two types of verb further down.

Higher up you see the verbs that are strongly associated with the construction. There are the frequent verbs and the infrequent verbs, but both of them are surprisingly over-represented in the construction. Down below you see the verbs that are not strongly associated with the construction. If we look at the orange box here, these verbs are frequent in the *as*-predicative. Bybee would predict that those should trigger a lot of completions with the construction, but it does not. There are not many completions with the *as*-predicative. Up here is the green box. These verbs are infrequent in the *as*-predicative, but they are strongly associated with them. This would be similar to the case of *elbow* in the *way*-construction. For the *as*-predicative, it is *hail*. It is not very frequent, neither inside nor outside the construction, but when you see *hail*, then it is a very reliable cue for the *as*-predicative.

In summary, collostructional strength, the strength of association between a construction and a lexical item, matters. Raw frequencies matter too. You see that the frequent verbs are slightly above the infrequent verbs, but that is a secondary factor.

Let me come back to how collostructional analysis works. What I have shown you so far is what is known as collexeme analysis, the basic type of collostructional analysis that compares frequency within a construction against lexical item frequency in the corpus as a whole.

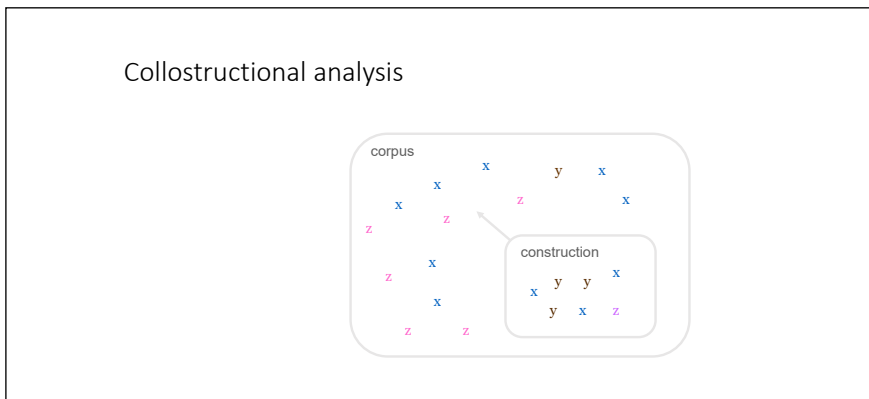


FIGURE 8

The analysis type that I briefly discussed this morning is distinctive collexeme analysis, where we compare collocate frequencies across two different constructions. We have construction A, which has a number of collocates with different frequencies. We have construction B with the same types at different frequencies. We can figure out which elements are maximally uneven in their distributions. Which elements have the greatest frequency asymmetry between construction A and construction B? That can give us a cue as to what makes these two constructions different. That kind of contrast, that kind of analysis makes sense when we are dealing with constructions that are related in some way, that have similar functions like *will* and *be going to*. Or think of different complementation patterns, *that*-clauses and *ing*-clauses, or the *get*-passive and the passive with *be*. You could also contrast broader tense patterns like the simple present and the present progressive. That would show you the different collocational preferences that these construction types have. Depending on the level of abstraction of these constructions, there are different observations you could make. So far, I have presented the synchronic way of conducting colostruational analysis. Let me get to its diachronic application.

The idea that I had on my bicycle, going home from university, was that we have a diachronic corpus with data from different historical periods, for example starting in the 1600s, then the 1700s and then 1800s and so on and so forth. We look for the same construction across different time slices of the same corpus. That enables us to identify asymmetries across these time slices.

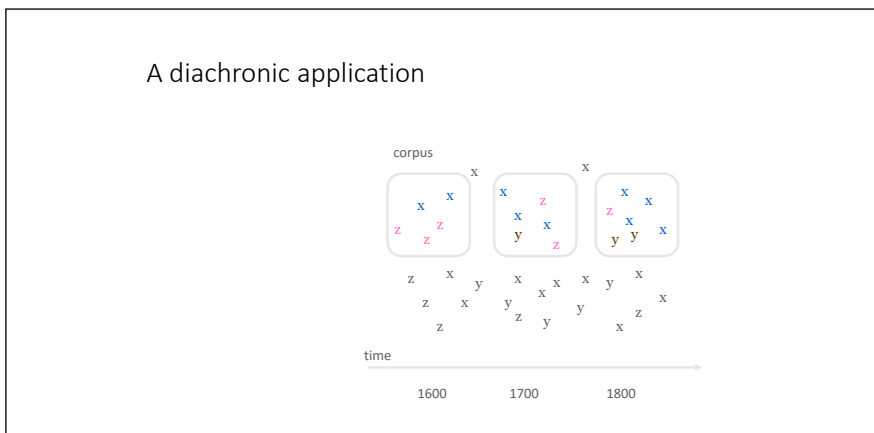


FIGURE 9

We can examine the collocates that a construction has in the 1600s, and we can ask whether they are the same as the ones that we find in the 1700s and in the 1800s. We can identify elements that have become more or less frequent

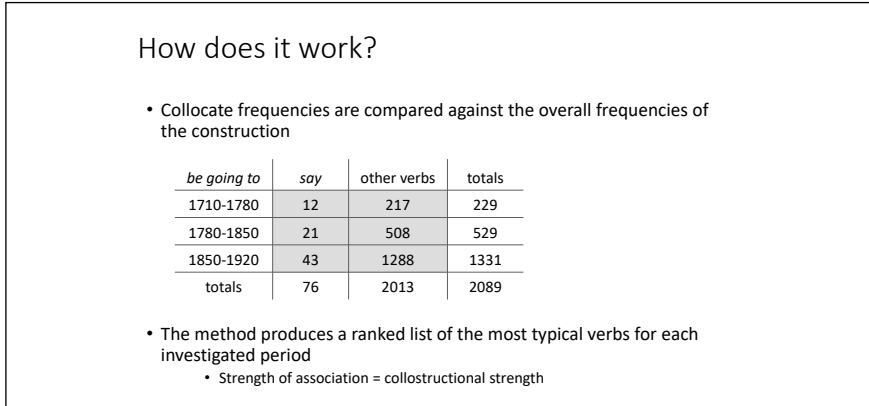


FIGURE 10

over time. We can identify elements that are typical for a particular section of the corpus, and we can draw conclusions from that.

The way it works in practice is that the collocate frequencies of a construction are compared against the overall frequencies of that same construction for each collocate that you observe. You see an example here of the *be going to* construction that is used with the verb *say* across three periods of time.

We observe the raw frequencies of *say*. They happen to increase. From 12 to 21 to 43, but then you see that also the construction as a whole increases in frequency as well. It goes from about 230 examples to 530 examples to more than 1300 examples. So *say* goes up in frequency, but also the construction itself goes up in frequency. Does that mean that *say* becomes more or less attracted to the construction? Does that mean that the attraction stays at the same level? This is something that we can figure out on the basis of a statistical analysis that takes all the verbs and their frequencies into account. The method produces a ranked list of the most typical verbs for each investigated period. I have used the term collostructional strength without properly defining it. Collostructional strength would be the strength of association between a construction and a lexical item that occurs within it, as measured by a collocation statistic.

What does this type of analysis show? It determines the elements that are most typical for a construction in modern usage, when you analyze a synchronic corpus. Applied diachronically, we can also determine the most typical elements for a construction in any given historical period. The collocational preferences can be used to describe the modern semantics and how it came to be that way. That was important for me. One other thing that attracted me to the method was that if we find that there are systematic changes, that there are attracted sets of semantically related collocates, not just individual items,

then that would suggest that there is a broader trajectory of semantic change going on that reaches beyond just individual words, individual histories of lexical items.

Why would this be useful? The method allows explorative analyses of constructions and their diachronic variation, how they have developed. More usefully perhaps, it allows us to test proposed semantic developments. I mentioned grammaticalization paths that have been proposed for the development of future constructions. My idea was to go into the data and see if these proposals that have been made could be falsified or substantiated. This is especially useful if we want to distinguish between competing hypotheses, which you can often find in the grammaticalization literature. One account might claim that this future construction developed in this way, and another account might propose a very different semantic pathway. How do we decide between the two? This method actually can get to the bottom of what kind of meaning came first, what kind of meaning it developed later, and how it all ended up.

Semantic classes of distinctive verbs would indirectly reflect stages of grammaticalization paths that have been proposed for the development of future markers, for example, in the work by Joan Bybee and colleagues (1991, 1994), who have argued for a path from the meaning of intention to the meaning of future and from thereon to epistemic or speaker-oriented modality.

To give you a taste of how all of this can be applied to concrete case studies, let's look at an example. So far, I have exclusively shown you data from English, so I think it is high time to broaden the outlook a bit. You know that English has a future with the modal auxiliary *will*. What you perhaps do not know is that a small language close by, Danish, has a *vil* future as well. The word looks almost exactly the same, etymologically it is the same.

There are further parallels in that the construction in its synchronic usage shows a preference for certain types of verbs. It is highly productive. It is highly general, but there are preferences, there are asymmetries. The construction prefers abstract atelic verbal complements such as verbs like *require*. There is nothing dynamic or agentive about *require*. It is a state. A Danish verb that means 'to be' is also among the most attracted items.

My questions for this particular case study were the following: Can the semantic developments be described in terms of shifting collocational preferences? Were there certain semantic verb classes that were central to the development? Can the collocates address the hypothesis that future constructions of this kind develop out of markers of intention? That would be the default hypothesis. If we have a verb that means "want", then it is a marker of intention that may eventually shade off into a use that encodes just future time reference, so that also inanimate subjects can be used with it. I can say things like, *It will rain*, which does not mean that *It wants to rain*, but rather that it will happen in the future.

Data

- raw-text historical corpora

Corpora	Century	Size	Search	Ex.
ACOD	12-14	36 k	<i>vil, will,</i>	99
DSST 1	15-16	400 k	<i>wil, will</i>	642
DSST 2	17-18	240 k		408
Gutenberg	19-20	700 k		999
Total		1.4 M		2,148

FIGURE 11

Here’s an overview of the kind of data I used. A collocation analysis does not require you to have millions and millions of words. In this case, I had four different historical periods ranging from the 12th to the 20th century. It is a long and thin corpus really, with all in all about 1.4 million words. That, by today’s standards, is considered very small. I looked for different orthographic variants of the auxiliary. I had a total of some 2000 examples to work with.

Absolute frequencies

ACOD (12-14)			DSST (15-16)			DSST (17-18)			Gutenberg (19-20)		
Verb	Gloss	N	Verb	Gloss	N	Verb	Gloss	N	Verb	Gloss	N
give	<i>give</i>	9	være	<i>be</i>	44	have	<i>have</i>	27	være	<i>be</i>	126
have	<i>have</i>	9	gøre	<i>do</i>	42	være	<i>be</i>	27	sige	<i>say</i>	84
tage	<i>take</i>	8	give	<i>give</i>	38	gøre	<i>do</i>	20	have	<i>have</i>	82
vide	<i>know</i>	5	have	<i>have</i>	31	sige	<i>say</i>	20	blive	<i>become</i>	41
lade	<i>let</i>	4	sige	<i>say</i>	26	give	<i>give</i>	19	se	<i>see</i>	29
søge	<i>seek</i>	4	lade	<i>let</i>	19	gå	<i>go</i>	11	gå	<i>go</i>	27
fære	<i>travel</i>	4	blive	<i>become</i>	16	bevise	<i>prove</i>	10	gøre	<i>do</i>	21
gøre	<i>do</i>	3	tale	<i>talk</i>	15	lade	<i>let</i>	10	fi	<i>get</i>	19
svare	<i>answer</i>	3	bevise	<i>prove</i>	10	tro	<i>believe</i>	10	tage	<i>take</i>	19
mæle	<i>speak</i>	2	ride	<i>ride</i>	10	forlade	<i>forgive</i>	9	komme	<i>come</i>	14
bytte	<i>swap</i>	2	skrive	<i>write</i>	10	komme	<i>come</i>	9	finde	<i>find</i>	13

FIGURE 12

When we examine the absolute frequencies of the most frequent collocates across the four periods, there are already a few observations that we can make.

For example, there is a verb meaning “give”, which is the most frequent one in the first period, after which it decreases. Eventually it disappears from the list of the most frequent elements.

In the opposite direction, there is a verb meaning “say” that does the exact opposite. First, it is not among the most frequent verbs, but then it gradually works its way up the list. In the last period it is on position two. But overall,

Decrease of *give*

ACOD (12-14)			DSST (15-16)			DSST (17-18)			Gutenberg (19-20)		
Verb	Gloss	N	Verb	Gloss	N	Verb	Gloss	N	Verb	Gloss	N
<i>give</i>	<i>give</i>	9	være	<i>be</i>	44	have	<i>have</i>	27	være	<i>be</i>	126
have	<i>have</i>	9	gøre	<i>do</i>	42	være	<i>be</i>	27	sige	<i>say</i>	84
tage	<i>take</i>	8	<i>give</i>	<i>give</i>	38	gøre	<i>do</i>	20	have	<i>have</i>	82
vide	<i>know</i>	5	have	<i>have</i>	31	sige	<i>say</i>	20	blive	<i>become</i>	41
lade	<i>let</i>	4	sige	<i>say</i>	26	<i>give</i>	<i>give</i>	19	se	<i>see</i>	29
søge	<i>seek</i>	4	lade	<i>let</i>	19	gå	<i>go</i>	11	gå	<i>go</i>	27
fare	<i>travel</i>	4	blive	<i>become</i>	16	bevise	<i>prove</i>	10	gøre	<i>do</i>	21
gøre	<i>do</i>	3	tale	<i>talk</i>	15	lade	<i>let</i>	10	fi	<i>get</i>	19
svare	<i>answer</i>	3	bevise	<i>prove</i>	10	tro	<i>believe</i>	10	tage	<i>take</i>	19
mæle	<i>speak</i>	2	ride	<i>ride</i>	10	forlade	<i>forgive</i>	9	komme	<i>come</i>	14
bytte	<i>swap</i>	2	skrive	<i>write</i>	10	komme	<i>come</i>	9	finde	<i>find</i>	13

FIGURE 13

Increase of *sige*

- (*det vil sige* – that is to say, that means)

ACOD (12-14)			DSST (15-16)			DSST (17-18)			Gutenberg (19-20)		
Verb	Gloss	N	Verb	Gloss	N	Verb	Gloss	N	Verb	Gloss	N
give	<i>give</i>	9	være	<i>be</i>	44	have	<i>have</i>	27	være	<i>be</i>	126
have	<i>have</i>	9	gøre	<i>do</i>	42	være	<i>be</i>	27	<i>sige</i>	<i>say</i>	84
tage	<i>take</i>	8	give	<i>give</i>	38	gøre	<i>do</i>	20	have	<i>have</i>	82
vide	<i>know</i>	5	have	<i>have</i>	31	<i>sige</i>	<i>say</i>	20	blive	<i>become</i>	41
lade	<i>let</i>	4	<i>sige</i>	<i>say</i>	26	give	<i>give</i>	19	se	<i>see</i>	29
søge	<i>seek</i>	4	lade	<i>let</i>	19	gå	<i>go</i>	11	gå	<i>go</i>	27
fare	<i>travel</i>	4	blive	<i>become</i>	16	bevise	<i>prove</i>	10	gøre	<i>do</i>	21
gøre	<i>do</i>	3	tale	<i>talk</i>	15	lade	<i>let</i>	10	fi	<i>get</i>	19
svare	<i>answer</i>	3	bevise	<i>prove</i>	10	tro	<i>believe</i>	10	tage	<i>take</i>	19
mæle	<i>speak</i>	2	ride	<i>ride</i>	10	forlade	<i>forgive</i>	9	komme	<i>come</i>	14
bytte	<i>swap</i>	2	skrive	<i>write</i>	10	komme	<i>come</i>	9	finde	<i>find</i>	13

FIGURE 14

when we examine the raw frequencies, most of what we see are highly frequent verbs, like *have*, like *be*, *do*, *give*, *take*, and *go*. These verbs, for the most part, do not have a whole lot of semantic substance, and thus they do not allow us to say much about how the construction develops semantically.

The absolute frequencies revealed some tendencies, but not tangible developments. There is a constant of light verbs with high absolute frequencies, which could be taken to suggest that the changes that happened were either non-substantial or unsystematic. These could be seen as chance fluctuations in the lexical domain. If I had been working with a raw-frequency approach, I would have given up at this point with the conclusion that nothing has happened. However, I was curious to see if the collocation approach would yield a different outcome, which it actually does.

Distinctive collexemes

ACOD			DSST 15-16			DSST 17-18			Gutenberg 19-20		
Verb	Gloss	CS	Verb	Gloss	CS	Verb	Gloss	CS	Verb	Gloss	CS
tage	take	3,5	ride	ride	5,2	bekende	confess	2,9	se	see	6,7
fare	travel	3,4	skrive	write	4,3	bevise	prove	2,8	være	be	6,2
søge	seek	2,8	gøre	do	4,3	tro	believe	2,8	sige	say	4,3
bytte	swap	2,7	give	give	3,8	forlade	forgive	2,5	spille	play	4,0
bøtte	pay	2,7	antegne	note	3,7	føre	lead	2,3	blive	become	3,5
fange	catch	2,7	forklare	explain	3,7	æde	eat	2,3	flå	get	3,4
løse	solve	2,7	råde	advise	3,7	nægte	deny	2,3	rejse	travel	3,0
mæle	talk	2,7	slås	fight	3,7	skikke	send	2,2	finde	find	2,8
rebe	tie	2,7	love	promise	2,1	vis	show	2,2	høre	hear	2,8
vide	know	2,6	tale	talk	1,9	bo	live	2,1	bringe	bring	2,4

FIGURE 15

Here's a table with the distinctive collexemes for each of the four periods. For each period you see the elements that are maximally over-represented in that particular corpus period. Let me briefly go through the periods individually and point out some developments along the way.

Period 1: ACOD

Verb	Gloss	CS
tage	take	3,5
fare	travel	3,4
søge	seek	2,8
bytte	swap	2,7
bøtte	pay	2,7
fange	catch	2,7
løse	solve	2,7
mæle	talk	2,7
rebe	tie	2,7
vide	know	2,6

- Distinctive verbs require animate, intentional subjects.
- Examples with distinctive verbs express intention, not future.

tha a bonden siæl uæ doom um at standæ hwilkæ bõtær han wil takæ
 'Then the farmer can decide on his own what fees he wants to take.'

Wil man witæ of manz konæ gør hoor læggæ thænnæ steen undær hænnæ houæth.
 'If you want to know whether your wife is cheating, put this stone under her pillow.'

FIGURE 16

In the first period, this is Old Danish, as it was written between the 12th and 14th century. We find exclusively verbs that require animate intentional subjects, things like *take*, *travel*, *seek*, *pay* and *catch*. Human beings can do that, inanimate beings cannot. If we are examining the examples with distinctive verbs, we find that these examples express intention, not future. We have things like *The farmer can decide what fees he will take*, that is, what fees he wants to take.

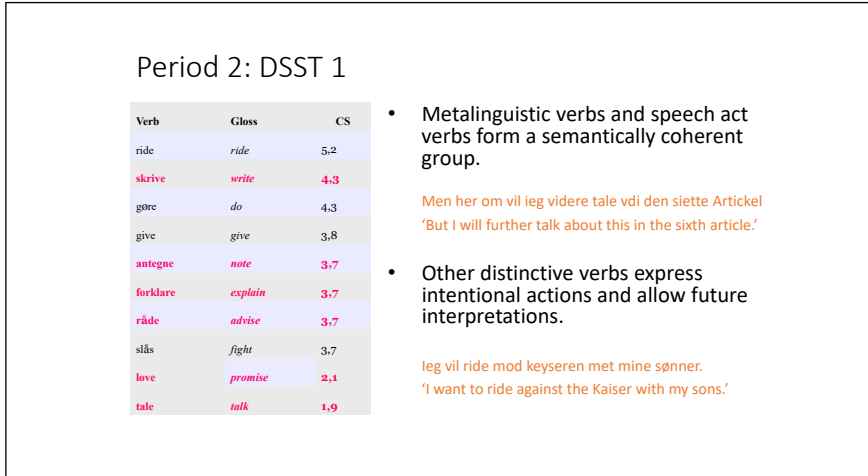


FIGURE 17

That will happen in the future, but the main idea is intention or volition. I love this example here: *If you want to know whether your wife is cheating, put this stone under her pillow. If you want to know, you will know eventually. I am not exactly sure what's supposed to happen with that stone, though.*

Moving on to Period 2, we find that the profile of verbs changes in that we see an over-representation of verbs that are metalinguistic, that encode speech acts. There are quite a few of them here. We have *write*, *note*, *explain*, *advise*, *promise* and *talk*, among the most attracted ones. There are a few other distinctive verbs that express intentional actions and that allow a future interpretation. The top collocate here is the word that means "ride", as in *ride a horse*. The relevant example that you have at the bottom of the slide here translates as "*I want to ride against the Kaiser with my sons*". Someone wants to do something, but it is also clear that this will happen in the future. This you can see as the future interpretation making inroads and establishing itself more strongly across the semantic spectrum of the construction.

In the last period, this trend of speech act verbs being strongly represented continues. Speech act verbs are still the largest coherent group of distinctive collexemes. The verbs *confess*, *deny*, or *forgive* express actions that you do linguistically, but the interesting thing that we see in Period 3 is the first occurrence of inanimate subject referents: *If your sins are about to lead you into desperation*. Your sins do not want anything. Your sins are actions, not intentional beings.

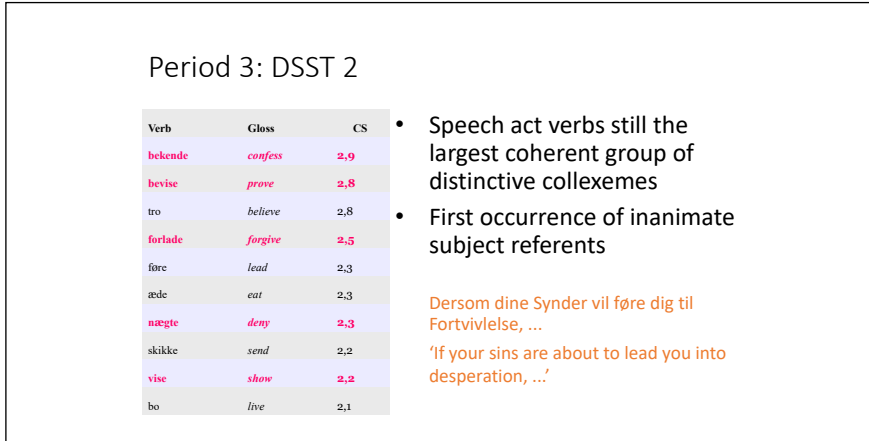


FIGURE 18

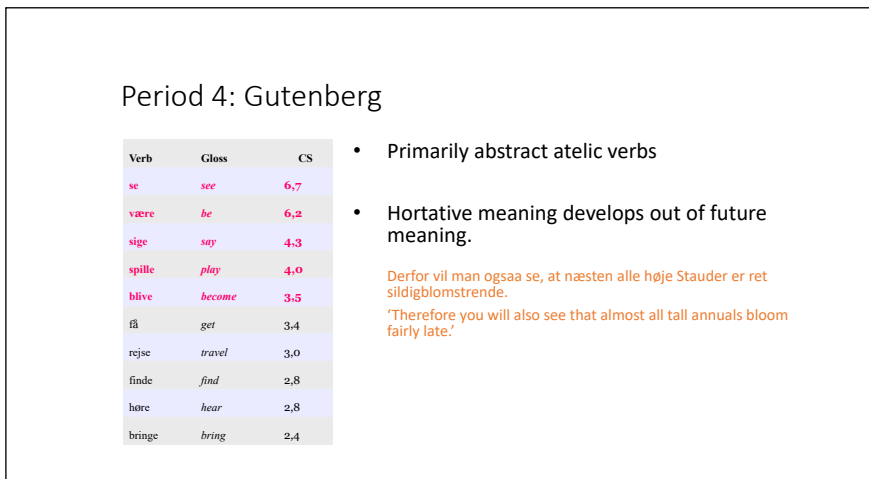


FIGURE 19

Finally, in the fourth period, which represents Present-day Danish, we see the profile that we are used to seeing in modern usage of the construction. The top collexemes are abstract atelic verbs, like *see*, *be* and *say*. There are a couple of others, and there are even extensions out of future meaning. There is a certain type of meaning that we can call hortative, which encourages someone else to do something. This occurs primarily in the context of *see, if you will see. You will see that almost all tall annuals bloom fairly late*. This is someone who is giving a piece of advice and who's inviting someone else to take this particular point

of view. That is something that does not necessarily encode future time, but it is an extension out of that temporal meaning.

What can we conclude from this? These data are corroborating the idea that future meaning with this construction developed out of intentional meaning. We see that very firmly ingrained in the construction's profile early on. You could say this is something that has been predicted all along, and we have lots of typological evidence for this particular pathway. What is the big deal? I would say the big deal is that the corpus-based material gives us a lot more detail than secondary resources that we can glean from descriptive grammars or even from native speaker's intuitions. One thing that definitely goes beyond the generic story of future tense development that has been proposed in grammaticalization studies is that we have this group of speech act verbs, which seem to be central to the development of this construction here.

There is a second example that I want to mention. For that, we will go from *will* futures to *be going to* futures. There is the English *be going to* future of course, but there is another small language close by, Dutch, which has its own *be going to* future, even though that form is a little different. It is not a progressive *going* type form, but it is a basic verbal form of a verb (*gaan*) that means 'go'.

These two constructions, Dutch *gaan* and English *be going to*, are often seen as more or less equivalent. They translate into each other. I can say "*It is going to rain tomorrow*" and there is an equivalent sentence in Dutch. I can say "*This is what we are going to do*". Again, there is a way to say that in Dutch. If I say "*I think that is going to happen*" with an inanimate subject, also this I can translate into Dutch almost word for word. The constructions share characteristics such as an orientation to the present, a preference for intentional or premeditated actions, and there is a preference for dynamic events as opposed to the stative and atelic predicates that *will* in English prefers.

That was my starting point. I thought this would be a good opportunity to develop a parallel case study, but then I looked at corpus data, and mutual translations of these two constructions. I examined corpus data from European Parliament Proceedings, which features speeches that are held in different languages and which are translated into several other European languages. The same text is thus available in different languages.

I looked at a data set of more than 7500 examples of *be going to*. Out of those 7500 there are only about 1000 that are translated into Dutch with *gaan*. That is 15%, which is not much. Had it been half, I would have suspected a stylistic reason, so that translators want to avoid colloquialisms and therefore choose the more conservative construction with *will*. A rate of 15% however suggests

to me that the two constructions are functionally different. It also made me wonder how much overlap I should expect in the first place.

I looked at two other future constructions that are not etymologically related and that would not be considered as translational equivalents. I looked at the *will* future and the Dutch *zullen* future, which is etymologically related to English *shall*. I took a smaller sample of 500 examples, and I found 215 translations with *zullen*, so 43%. The overlap between these unrelated construction is significantly larger than the overlap that we find between *be going to* and *gaan*.

That sparked my interest. What semantic characteristics of *be going to* and *gaan* work against a mutual translation in present-day usage? Why are they not translated into each other more than they are? Second, did the two constructions drift apart only recently, or did they grammaticalize in different ways? That would be something of a surprise, because movement-based futures are thought to develop along similar lines.

My analysis was based on synchronic and diachronic corpus resources from English and Dutch. For modern English, I used the British National Corpus (BNC), a 100 million words corpus. For the historical part, I use the CLMET, which is a 9.5 million words corpus. I also used the Oxford English Dictionary, the source that Michael Israel used for the *way*-construction. The corpora for Dutch are smaller. The modern data comprise about 8 million words and the historical data from the Project Gutenberg hold about 4 million words. I exhaustively retrieved all examples with *going to* and *gonna* plus infinitive, and then the Dutch forms of the verb *gaan*, for which there are several morphologically inflected variants. I analyzed the data both in a synchronic way, with a collexeme analysis, and then diachronically with a diachronic distinctive collexeme analysis.

Comparing the collexemes

<i>be going to</i>		<i>gaan</i>		
Verb	CollStr	Verb	Gloss	CollStr
do	Inf	regenen	<i>rain</i>	73.56
get	Inf	praten	<i>talk</i>	32.80
happen	Inf	gebeuren	<i>happen</i>	32.02
say	168.39	kosten	<i>cost</i>	29.42
die	125.70	waaien	<i>storm</i>	26.26
cost	93.45	werken	<i>work</i>	25.56
put	91.57	samenwerken	<i>collaborate</i>	21.81
ask	59.91	zitten	<i>sit</i>	18.83
go	58.13	onderzoeken	<i>analyze</i>	18.52
marry	52.95	schijnen	<i>shine</i>	17.99

FIGURE 20

Overlap

<i>be going to</i>		<i>gaan</i>		
Verb	CollStr	Verb	Gloss	CollStr
do	Inf	regenen	<i>rain</i>	73.56
get	Inf	praten	<i>talk</i>	32.80
happen	Inf	gebeuren	<i>happen</i>	32.02
say	168.39	kosten	<i>cost</i>	29.42
die	125.70	waaien	<i>storm</i>	26.26
cost	93.45	werken	<i>work</i>	25.56
put	91.57	samenwerken	<i>collaborate</i>	21.81
ask	59.91	zitten	<i>sit</i>	18.83
go	58.13	onderzoeken	<i>analyze</i>	18.52
marry	52.95	schijnen	<i>shine</i>	17.99

FIGURE 21

Let me show you the synchronic results first. On this slide, you see two lists with the synchronic colllexemes of *be going to* and *gaan*. The lists show the most attracted lexical verbs for both *be going to* and *gaan*.

Among the top colllexemes, there are exactly two verbs that match. With *be going to*, we have *happen* and *cost*. With *gaan*, we have two Dutch verbs that also mean “happen” and “cost”, but that is where the similarities end.

Weather phenomena

<i>be going to</i>		<i>gaan</i>		
Verb	CollStr	Verb	Gloss	CollStr
do	Inf	regenen	<i>rain</i>	73.56
get	Inf	praten	<i>talk</i>	32.80
happen	Inf	gebeuren	<i>happen</i>	32.02
say	168.39	kosten	<i>cost</i>	29.42
die	125.70	waaien	<i>storm</i>	26.26
cost	93.45	werken	<i>work</i>	25.56
put	91.57	samenwerken	<i>collaborate</i>	21.81
ask	59.91	zitten	<i>sit</i>	18.83
go	58.13	onderzoeken	<i>analyze</i>	18.52
marry	52.95	schijnen	<i>shine</i>	17.99

FIGURE 22

It is instructive to look at the different verb types that we find across the two constructions. In Dutch, there are quite a few weather phenomena, such as *rain*, *storm*, and *shine*. That is not the typical kind of verb that you expect to see with a movement-based future. Rain is not an intentional agent, it is a natural phenomenon.

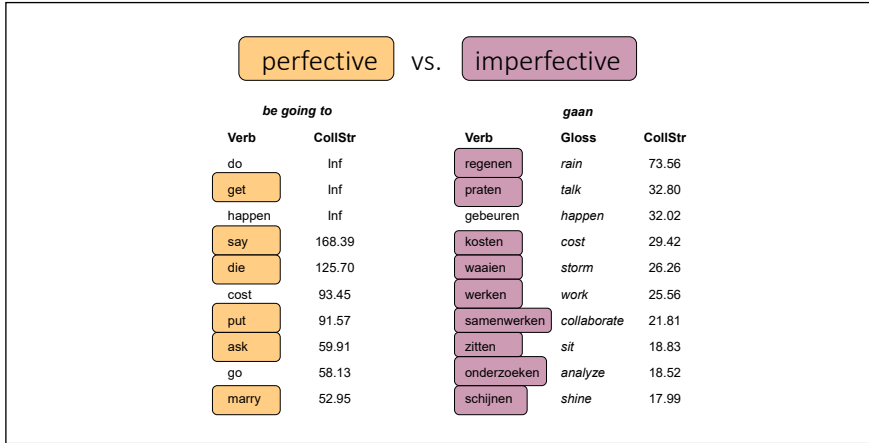


FIGURE 23

What struck me more though is that there is a great asymmetry with regard to the lexical aspect of verbs that we find in either construction. *Be going to* has a strong preference for verbs that I describe here as perfective. They encode a particular start point or end point, or they are even just happening at one singular point in time. Consider a verb such as *get*, which describes a punctual event. One moment you do not have it, then you get it, and then you have it. The same is true for the verb *say*. The verb *die* is perhaps the most drastic of them all. *The verb marry* also encodes a clear, punctual division of before and after.

With *gaan*, the most strongly attracted elements are imperfective. *Rain* can go on for a long time. Verbs such as *talk*, *storm*, *work*, *collaborate*, *sit*, or *analyze* encode activities that I can carry on and continue for an unlimited amount of time, or for as long as I choose to do so. That is one considerable difference.

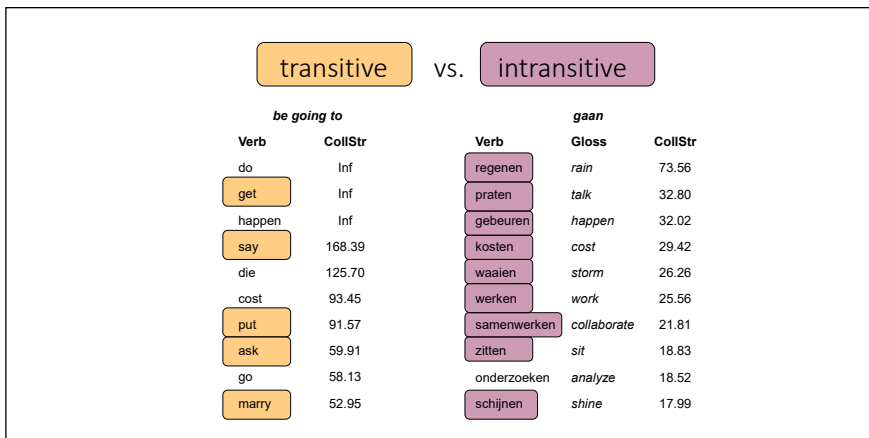


FIGURE 24

There is another difference in terms of transitivity. *Be going to* has a strong preference for transitive predicates. *I get something, I say something, I put something somewhere, I ask someone, I marry someone.* These are transitive verbs. The most attractive elements in the Dutch construction are intransitive, i.e. *rain, talk, happen, cost, storm, work* and so on and so forth.

agentive		vs.	non-agentive		
<i>be going to</i>			<i>gaan</i>		
Verb	CollStr		Verb	Gloss	CollStr
do	Inf		regenen	rain	73.56
get	Inf		praten	talk	32.80
happen	Inf		gebeuren	happen	32.02
say	168.39		kosten	cost	29.42
die	125.70		waaien	storm	26.26
cost	93.45		werken	work	25.56
put	91.57		samenwerken	collaborate	21.81
ask	59.91		zitten	sit	18.83
go	58.13		onderzoeken	analyze	18.52
marry	52.95		schijnen	shine	17.99

FIGURE 25

Another parameter that is strikingly asymmetric is agentivity: *I do something actively, I get something, I say something, I put something somewhere.* Things like *rain* or *happen* or *sit* are not agentive in the same way. There is no patient argument that would be affected by these activities.

To summarize, when we compare the colllexemes, the main differences in synchronic usage are concerned with lexical aspect, transitivity and agentivity. These are aspects of transitivity that have been described famously in a paper by Hopper and Thompson (1980). I concluded that these grammatical differences really motivate the low rate of mutual translations. Grammatically, the two constructions both refer to future time events, but that is where the similarity ends. With regard to what kinds of events are encoded, the two constructions are almost diametrically opposed to each other, which brings me to the historical development. How did all this come to be? Let's look at the distinctive colllexemes historically of English *be going to*.

This table shows you the overall development across the three periods with lists of the most attracted verbs for each historical time slice. Let me go through each period on its own.

We start with data in the 1700s, so 1710 to 1780. Here the most attracted colllexemes encode intentional activities, and movement is often still a possible interpretation. When I am saying something like "*You're going to fight for your country*", you're not going to do it in your living room. You're going somewhere

English *be going to*

1710-1780		1780-1850		1850-1920	
Verb	CollStr	Verb	CollStr	Verb	CollStr
fight	3.48	hunt	2.31	be	10.66
publish	2.32	speak	2.21	do	4.81
answer	1.94	commence	1.79	get	4.18
observe	1.94	expose	1.79	have	3.19
embrace	1.92	part	1.79	try	2.94
ravish	1.92	strike	1.79	die	2.30
relate	1.69			stay	1.76
begin	1.56			happen	1.74
visit	1.48			run	1.57
				talk	1.37

FIGURE 26

English *be going to*

1710-1780	
Verb	CollStr
fight	3.48
publish	2.32
answer	1.94
observe	1.94
embrace	1.92
ravish	1.92
relate	1.69
begin	1.56
visit	1.48

- Intentional activities, movement still a possible interpretation
 - And now my boy, I cried, you are going to fight for your country.
 - I am going to visit the Marquis, and will talk further with thee at my return.
- Metalinguistic verbs
 - By the circumstances of the story which I am going to relate, you will be convinced of my candour.
 - As he was going to begin his narrative, Rasselas was called to a concert.

FIGURE 27

else to do the fighting. In “*You’re going to visit someone*”, it is implied that this visiting will take place somewhere else. Again, there are verbs that encode speech acts. There are a number of metalinguistic verbs that figure in this early period. Metalinguistic verbs feature in examples such as *the story which I am going to relate* or *as he was going to begin his narrative*. These verbs are attracted to *be going to* during this early period.

In the second period, all distinctive collexemes are compatible with intentional actions. Speech act verbs still continue to be represented. But we find the first events that are independent of human actions, as for instance the distinctive collexeme *strike*. While *strike* could be viewed as an agentive, intentional verbs that requires a human agent who is carrying out an action, the example on this slide shows a different meaning. In the example “*When ten*

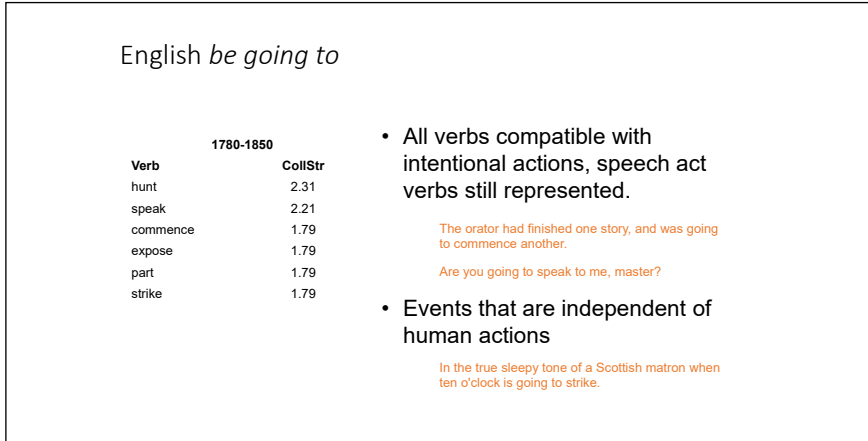


FIGURE 28

o'clock is going to strike”, it is a clock doing the striking, not a human being. This opens up a pathway to other inanimate entities accomplishing actions, and eventually the construction broadens semantically.

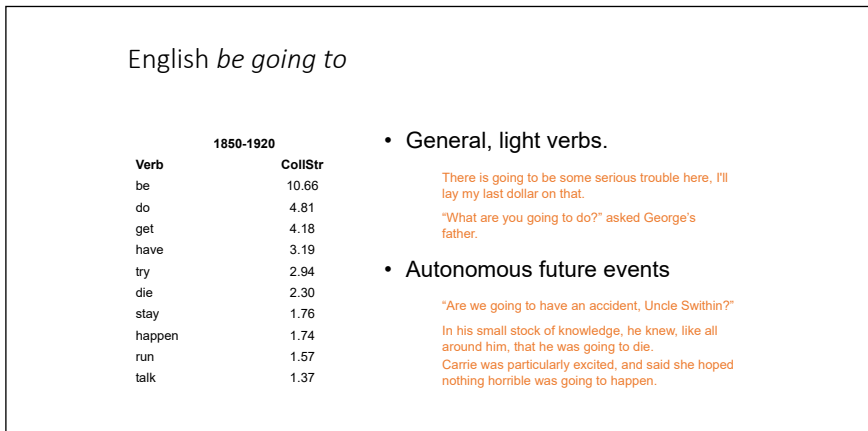


FIGURE 29

Moving on to the third period, here we find again the present-day profile of what *be going to* is like. There are general light verbs such as *be*, *do*, *get* and *have*, and these encode at least to some extent autonomous future events that haven't been planned and that haven't been executed by evolutionary agents. Examples like "*Are we going to have an accident*" appear in the form of a question, "*He knew that he was going to die*" or "*She hoped nothing horrible was going to happen*" encode spontaneous or hypothetical events. Examples of this kind are over-represented in this last period. In the shifting patterns of collocations, we can thus see a development towards more abstract meanings.

Also with *be going to*, the intentional source of future meaning seems very solid as a hypothesis. Speech act verbs as prototypical intentional verbs are central. The development towards future meaning is accompanied by a growing preference for general light verbs. All in all, this corroborates existing accounts of the grammaticalization of *be going to*.

Let's look at the diachrony of Dutch *gaan*. This slide shows an overview, but we are going to look at each period in turn.

Dutch *gaan*

16th – 17th century			18th – 19th century			20th century		
Verb	Gloss	CS	Verb	Gloss	CS	Verb	Gloss	CS
lopen	walk	3.52	opzoeken	find	2.67	beminnen	love	3.44
strijken	run off	3.30	onderzoeken	analyze	2.35	denken	think	3.44
stellen	put	3.25	vertellen	tell	2.25	gebeuren	happen	2.45
reizen	travel	2.86	doorbrengen	spend	2.01	bevrijden	liberate	1.96
liggen	lie	2.46	geven	give	1.68	studeren	study	1.96
preken	preach	2.00	varen	travel	1.68	voelen	feel	1.96
treden	step	2.00	verkoopen	sell	1.68	eten	eat	1.66
verhuizen	move	2.00	leggen	put	1.50	werken	work	1.49
leiden	lead	1.48	sterven	die	1.49	krijgen	get	1.47
rechten	straighten	1.48	brenge	bring	1.49	twijfelen	doubt	1.47
spreken	speak	1.41	halen	get	1.41	kijken	look	1.32
drinken	drink	1.36	roepen	call	1.39			
bezigen	use	1.33	nemen	take	1.37			

FIGURE 30

Gaan unsurprisingly starts with predicates that are movement verbs such as *walk*, *run off*, *travel*, *step* and *move*, which are among the most strongly attracted elements in the first period. There are caused posture verbs like *put* or *straighten*. Most verbs have a imperfective aspectual profile, as for example *walk* or *travel*, which encode actions that you can continue for a long time. These verbs further encode intentional human actions.

Dutch *gaan*

16th – 17th century		
Verb	Gloss	CS
lopen	walk	3.52
strijken	run off	3.30
stellen	put	3.25
reizen	travel	2.86
liggen	lie	2.46
preken	preach	2.00
treden	step	2.00
verhuizen	move	2.00
leiden	lead	1.48
rechten	straighten	1.48
spreken	speak	1.41
drinken	drink	1.36
bezigen	use	1.33

- Movement verbs
 - Nu willic gaan loopen al in mijn huus.
now want.I go walk all in my house
'Now I want to go home.'
 - Daer gaet hij strijcken!
there goes he run.off
'There he's running off!'
- Caused posture verbs
- Most verbs are imperfective
- Intentional human actions

FIGURE 31

Dutch *gaan*

18th – 19th century		
Verb	Gloss	CS
opzoeken	go to	2.67
onderzoeken	analyze	2.35
vertellen	tell	2.25
doorbrengen	spend	2.01
geven	give	1.68
varen	travel	1.68
verkoopen	sell	1.68
leggen	put	1.50
sterven	die	1.49
brengen	bring	1.49
halen	get	1.41
roepen	call	1.39
nemen	take	1.37

- Movement verbs
 - We zullen haar eens gaan opzoeken.
we shall her immediately go go.to
'We're going to go to her immediately.'
 - Hij kan als scheepsjongen gaan varen.
he can as cabin.boy go travel
'He can sail the sea as a cabin boy!'
- Transfer of objects
 - Zij gaan die verkopen te Volosca.
they go them sell to Volosca
'They're going to sell them to Volosca.'

FIGURE 32

Dutch *gaan*

20th century		
Verb	Gloss	CS
beminnen	love	3.44
denken	think	3.44
gebeuren	happen	2.45
bevrijden	liberate	1.96
studeren	study	1.96
voelen	feel	1.96
eten	eat	1.66
werken	work	1.49
krijgen	get	1.47
twijfelen	doubt	1.47
kijken	look	1.32

- Cognitive and emotive verbs
 - Wat zou men wel gaan denken?
what should people well go think
'What is everybody going to think?'
- *gebeuren* 'happen'
 - Wat gaat er dan gebeuren, Sander?
what goes there then happen Sander
'What is going to happen then, Sander?'
- mostly imperfective activities
- often no intentionality

FIGURE 33

In the second period, movement verbs are still strongly represented. There is a verb that means “go to” and there is a verb that means “travel”. Other distinctive collexemes encode transfers, as for example *spend*, or *give*, or *sell*, or *bring*. Suddenly these show up as a relatively homogeneous class. They are different in terms of their aspectual profile, as they have an endpoint. If you spend your money, then it is gone.

In the third period, the distinctive collexemes include cognitive and emotive verbs. A verb that means *happen* is another strongly attracted element. Most of the distinctive elements for this period are imperfective activities, typically with no intentionality at all. A verb like *love* encodes a human activity, but I can't intentionally decide to love someone, and when I love someone that is a state that has a temporal extension. That is very different from verbs such as

break or *cut*. The list of verbs on this slide includes *love*, *think*, *feel*, and *doubt*, which form the profile of this movement-based future construction.

In summary, the early usages of *gaan* commonly involve literal, intentional motion. Later, movement verbs are joined by verbs of transfer. All along the constructional meaning broadens, accommodating verbs without the meaning of intentionality and then in present-day usage, *gaan* preferentially occurs with atelic predicates, and intention is no longer a part of the constructional semantics. In synchronic usage, we find weather verbs as attracted collexemes, and in the third historical period, these are cognitive response verbs such as *doubt*.

What can we conclude from all of this? English *be going to* and Dutch *gaan* both follow the general path of movement-based future constructions, which start with the idea of motion, then merge into the idea of intention, and finally settle into future time reference. Even though they are moving along the same path that is typologically well-attested, this does not mean that they function similarly in language use. They have converse preferences for perfectivity, transitivity, and agentivity. If we are looking at the collocational patterns in their shifts, we see substantial developmental differences. *Be going to* has a preference for speech act verbs in the second period. *Gaan* starts out with movement verbs that are imperfective like *travel*.

I have talked enough about future constructions for one day. To sum up, I hope you see what I find attractive about the concept of looking at shifting patterns of associations. If you come to it for the first time you might think that this gives you a very diffuse idea of how language changes. Wouldn't it be much clearer to look at morphosyntactic changes or at first instances of a construction with a tangible meaning difference? I think that the study of collocational shifts can yield important insights.

I would submit that shifts in collocational preferences constitute one central type of change in the network of constructions. I have mentioned host-class expansion as one important concept in this context. A construction increases its number of links either to lexical collocates or to syntactic carrier phrases, constructions. I have argued that shifting collocational preferences can actually reflect systematic patterns of semantic change with a sufficient amount of accuracy, so that we can test hypotheses or test claims that have been proposed about semantic change elsewhere. Collocational preferences change as a construction develops along a grammaticalization path. Different paths then are embodied or realized by collocational changes that differ with respect to each other as well. With this, I would like to come to a close for today. Thank you for your attention.