DISCUSSION NOTE

Distinctive collexeme analysis and diachrony*

MARTIN HILPERT

1. Introduction

This discussion note argues that distinctive collexeme analysis (Gries and Stefanowitsch 2004) can be applied to analyses of diachronic corpus data, and that such an application makes it a useful tool for the study of grammaticalizing constructions. In analyses of synchronic corpus data, distinctive collexeme analysis demonstrates how several given constructions differ from each other with respect to conventionally associated lexical material. Different collocational preferences are taken to reflect semantic differences between the investigated constructions. Applied diachronically, the method can be used to compare the collocational preferences of a single construction in different periods of time. Systematic differences in the collocational preferences can be interpreted as an ongoing change in the constructional semantics.

The following three sections address the questions *How does it work?*, *Why is it useful?*, and *What does it show?*. Section 2 recapitulates the method of distinctive collexeme analysis and outlines how it can be applied to diachronic corpus data. Section 3 discusses practical considerations and identifies areas of study that can be addressed with diachronic distinctive collexeme analyses. Section 4 discusses how the results of such an analysis can be interpreted.

2. How does it work?

Distinctive collexeme analysis (Gries and Stefanowitsch 2004) contrasts two or more constructions in their respective collocational preferences. The compared constructions may be entirely unrelated, but in practice the method is particularly suited for the study of related constructions, as for example the English future constructions with *will* and *be going*

Corpus Linguistics and Linguistic Theory 2–2 (2006), 243–256 1613-7027/06/0002–0243
DOI 10.1515/CLLT.2006.012

Table 1. *Top 10 verbs with* will *and* going to *in the BNC*

WILL		GOING TO	
VERB	TOKENS	VERB	TOKENS
be	41,947	*be*	4,756
have	5,906	*do*	1,907
take	4,150	*get*	1,403
make	3,182	*have*	983
do	3,039	*take*	647
go	2,821	*say*	643
come	2,732	*make*	631
give	2,543	*go*	616
continue	2,477	*happen*	552
find	2,465	*tell*	434

to. While both constructions are used to refer to future events, several differences between the two have been pointed out (Binnick 1971, Wekker 1976, Close 1977, Haegeman 1989, Berglund 1997). Gries and Stefanowitsch (2004: 113) show that a previously overlooked difference lies in the preferred verbal collocates of each construction. Yet, this difference does not immediately fall out of the raw frequencies of the respective collocates. Table 1 lists the ten most frequent verbs in both of these constructions, based on data from the British National Corpus (Leech 1992).

In both constructions, the most frequent elements are general verbs that are either semantically light, such as *do*, or polysemous, such as *go*. Distinctive collexeme analysis affords a way to abstract away from frequent elements that are common to both investigated constructions and highlights those elements that are distinctive for each respective construction. Taking the absolute frequencies of both constructions into account, the analysis determines whether there are asymmetries in the relative frequencies of the co-occurring lexical verbs. The method identifies all verbs that occur significantly more often with one construction than with the other, and ranks these according to the degree to which they are distinctive. Table 2 illustrates the calculation with the example of *say*, which occurs significantly more often with *be going to* than with *will* (Fisher Exact, p = 5.41E-196).

The calculation in Table 2 is done for all verbs that are encountered with the respective constructions. The overall result of a distinctive collexeme analysis is a pair of lists, which rank collocating items according to their collostructional strengths. Table 3 presents the ten most distinctive collexemes for both *will* and *be going to* based on data from the

Table 2. *Input for a distinctive collexeme analysis of* say *in* will *and* going to

	say	other verbs	TOTALS
will	813	185,734	186,547
going to	643	26,294	26,937
TOTALS	1,456	212,028	213,484

BNC. The more distinct an element is, the higher its numerical value of collostructional strength (COLLSTR), which is a log-transformed probability value. Collstr values that are larger than 1.3 indicate that an element is distinct at the significance level of $p < .05$.

The two lists in Table 3 can be used to assess the semantic differences between *will* and *going to*. In a comparison of *will* and *going to* based on data from the British component of the International Corpus of English (ICE-GB), Gries and Stefanowitsch (2004: 114) argue that among the distinctive collexemes of *will*, many are non-agentive or low in transitivity. Their results are replicated here with data from the BNC, as Table 3 lists distinctive collexemes of *will* such as *continue*, *include*, *remain*, and *depend*. Conversely, *be going to* has distinctive complements that are agentive and high in transitivity, such as *do*, *say*, *put*, or *marry*. By bringing these elements into focus, distinctive collexeme analysis accentuates semantic differences between potentially fairly similar constructions.

Distinctive collexeme analysis can not only contrast two constructions, but it can also be used to compare three or more constructions. This application is particularly apt for the analysis of grammatical domains that are instantiated by several expressions. In many languages, layers of constructions expressing similar temporal, modal, or aspectual

Table 3. *Top 10 distinctive collexemes of* will *and* going to *in the BNC*

WILL	COLLSTR	GOING TO	COLLSTR
continue	83.57	*do*	Inf
be	74.17	*get*	Inf
provide	61.39	*say*	195.36
include	56.35	*happen*	135.34
remain	44.76	*ask*	87.20
receive	42.50	*die*	78.72
become	41.15	*put*	74.96
depend	39.41	*tell*	58.85
enable	37.72	*marry*	53.99
require	36.58	*let*	42.95

meanings have grammaticalized. In Danish, constructions with the verbs *ville* 'want', *skulle* 'shall', and *komme til at* 'come to' convey future meanings (Allan *et al*. 2000). In Dutch, the posture verbs equivalent to English *sit*, *stand*, and *lie* encode progressive aspect (Lemmens 2005). English has several analytic causative constructions, such as the *make*-causative, the *let*-causative, and the *have*-causative (Stefanowitsch 2001). A multiple distinctive collexeme analysis can be used to investigate how these constructions differ with respect to their collocational preferences.

To illustrate the statistical processing that is involved in a multiple distinctive collexeme analysis, we can stay with the example of English causatives, which have been analyzed in this way by Gilquin (to appear). Gilquin compares analytical causatives with the verbs *make*, *have*, *get*, and *cause*, distinguishing further between different complementation patterns of these verbs. For instance, causative *have* can be complemented by a past participle, a gerund, or an infinitive. In order to run the analysis, Gilquin identifies and lemmatizes all co-occurring lexical verbs. This procedure yields observed frequencies for all main verb complements in the analyzed constructions. For instance, the verb *take* occurs in 8 examples of the *make*-causative, as shown below in Table 4. Based on the observed frequencies and the overall frequencies of each construction, the expected frequency of each verb can be calculated. In the *make*-causative, the verb *take* has an expected frequency of 10.9, which is more than the observed frequency.

The differences between observed and expected frequency have to be interpreted with a statistical test in order to see whether they are significant. In the version of multiple distinctive collexeme analysis implemented in *Coll.analysis* (Gries 2004), this is achieved by performing a series of binomial tests for each verb (like the Fisher-Yates Exact test, this statistic does not make the assumption of normally distributed data, and it can operate on small sample sizes). Specifically, the procedure is as follows (Stefanowitsch, p.c., cf. also the documentation of Gries 2004). Take the example of *take* in the *make*-causative. This combination occurs 8 times in Gilquin's data. As shown in Table 4, her data includes 35 cases of *take* occurring in any causative construction. Taking into

Table 4. *Observed and expected frequency and CollStr value for* take *(Gilquin to appear)*

TAKE	*cause* + *to*	*get* + *to*	*get* + pp	*get* + prp	*have* + inf	*have* + pp	*have* + prp	*make* + inf	*make* + *to*	*make* + pp
obs. freq.	5	8	2	1	2	9	0	8	0	0
exp. freq.	2.09	3.79	7.81	1.26	0.73	6.48	0.69	10.90	0.97	0.29
CollStr	1.26	1.51	−2.06	−0.20	0.79	0.73	−0.30	− 0.72	−0.43	−0.13

account that there are 1,156 cases of the *make*-causative and 3,713 causatives overall in the data, the expected frequency of *take* in the *make*-causative is 10.9 (1,156*35/3,713). The probability that if *take* occurs in a causative construction it will occur in the *make*-causative is thus 0.31 (10.9/35). We can now use a binomial test to calculate the probability (the binomial p-value) of 8 successes in 35 trials given an *a priori* probability of 0.31. This probability is 0.19, which indicates that the observed frequency does not deviate significantly from a chance distribution. As usual in collostructional analysis, this p-value is interpreted directly as an association measure. For expository reasons, *Coll.analysis* $\log_{10}$-transforms this value and then sets the sign to reflect the direction of association. The negative sign in the CollStr value of -0.72 shown in Table 4 indicates a negative association (repulsion), which however is not significant. Other cells in Table 4 do indicate significant degrees of attraction. The verb *take* occurs significantly more often than expected in the construction *get X to take Y* (CollStr = 1.51), and significantly less often than expected in the construction *get X taken* (CollStr = -2.06).

The output of a multiple distinctive collexeme analysis is quite similar to Table 3, except that it returns not two ranked lists, but one list for each construction that enters the analysis. In the case of Gilquin's analysis, the output consists of ten lists showing the distinctive elements of each construction. A comparison of the most distinctive elements for each causative show how the constructions map onto different types of causation events, or with what kinds of causers and causees the constructions typically occur.

The comparison of more than two alternatives is crucial to the diachronic application of distinctive collexeme analysis. Using diachronic corpora such as the PPCEME (Kroch *et al*. 2004) or the CLMET (de Smet 2005), which hold the parameter of text genre constant over several periods of time, the method can determine what types of co-occurring elements were preferred by a given construction at different historical stages. In a first application, Kemmer and Hilpert (2005) investigate the English *make*-causative and find shifting collocational preferences that can be interpreted as changes in the meaning of the construction. Here, the history of the English auxiliary *shall* will be used as an illustrating case study. Using both the PPCEME and the CLMET, all instances of *shall* with its orthographical variants and inflected alternatives are exhaustively extracted. Taken together, the two corpora cover six successive 70-year periods from 1500 to 1920, which are collapsed into three 140-year periods for the purposes of the present study. From all retrieved examples, the infinitive complements of the auxiliary are identified and brought into the form of a frequency list. Orthographical variants are identified and standardized, such that for example instances of *fynde* are

Table 5. *Top 10 verbs with* shall *over three periods of English*

1500–1640		1640–1780		1780–1920	
VERB	TOKENS	VERB	TOKENS	VERB	TOKENS
be	736	be	557	be	1,074
have	291	have	234	have	527
find	133	find	107	see	239
see	131	see	75	go	195
come	120	make	69	do	176
do	117	think	57	find	116
make	94	take	52	take	95
take	92	endeavour	52	make	89
hear	73	do	51	say	87
know	69	give	46	get	82

counted as instances of *find*. Table 5 shows for each period the ten most frequent verbs with *shall*.

The motivation for applying a distinctive collexeme analysis to present-day data is the observation that raw frequencies sometimes obscure differences that hold across sets of data. This observation carries over to raw frequencies representing historical data. As a tool to investigate the history of *shall*, Table 4 is only of limited use. As in Table 1, many of the most frequent elements are semantically light and have a high overall text frequency. The verbs *be* and *have* are the most frequent items in each period. The general verbs *do*, *see*, *find*, *make*, and *take* are among the most frequent complements of *shall* in each period. This motivates an analysis in terms of elements that are distinctive, not merely frequent. A comparison of the frequencies in Table 5 with the overall frequencies of *shall* in each period shows which items were most distinctive for the construction in a given period of time. As in the synchronic application of distinctive collexeme analysis, the collostructional method abstracts away from items that are common to each period and highlights those that are significantly more frequent than expected. Items are judged as distinctive if they occur with a higher relative frequency in one period than in the other two. In this way, differences between the three periods are accentuated and actual developments become more prominent. Table 6 illustrates the input that is needed to calculate the status of the verb *say* over the three periods.

For each period, the table lists the frequency of *say* as compared to all other verbs that occur as complements of *shall*. For example, in the earliest period there are 48 instances of *say* out of a total of 4,472 examples. By comparing the relative frequency of *say* across the three periods,

Table 6. *Input for a diachronic distinctive collexeme analysis of* shall say

	say	other verbs	TOTAL
1500–1640	48	4,527	4,575
1640–1780	36	3,298	3,334
1780–1920	87	5,989	6,076
TOTAL	171	13,985	13,819

it can be determined accurately whether and how strong *say* is attracted to *shall* in each respective period.

The calculation illustrated in Table 6 is done for every infinitive complement of *shall* in either of the three periods, using *Coll.analysis*. The analysis yields ranked lists of the most distinctive items for each period. Table 7 shows the ten most distinctive collexemes for each of the three periods along with their actual token figures. As is apparent from the numbers, there are substantial mismatches between raw frequencies and values of collostructional strength. The applied calculation promotes the ranking of verbs that are maximally unevenly distributed. The most distinctive verb of the first period, *understand*, occurs 48 times in that period, but only once in the second period and seven times in the third period. In this way, *understand* is more typical of the first period than *come*, which occurs much more often but is also more widely distributed across the three periods.

For now, Table 7 is merely meant to demonstrate the fact that a distinctive collexeme analysis presents a different perspective on historical data than raw frequencies. The diachronic application of distinctive col-

Table 7. *Top 15 distinctive collexemes of* shall *over three periods of time*

1500–1640			1640–1780			1780–1920		
VERB	N	COLLSTR	VERB	N	COLLSTR	VERB	N	COLLSTR
understand	48	15.48	*endeavour*	52	16.36	*forget*	81	17.01
come	120	10.32	*discover*	17	7.86	*go*	194	12.91
forfeit	40	6.53	*examine*	13	6.86	*get*	81	9.46
perceive	19	6.52	*mention*	18	5.90	*try*	27	6.87
bear	30	6.49	*suppose*	14	5.67	*meet*	53	6.36
appear	37	5.65	*confine*	10	5.29	*feel*	32	5.59
serve	22	5.62	*direct*	10	5.29	*have*	527	5.07
need	28	5.48	*explain*	12	5.14	*see*	239	4.88
eat	28	5.48	*think*	57	4.70	*write*	45	4.11
bring	40	5.28	*add*	18	4.33	*return*	43	3.96

lexeme analysis thus uncovers differences between sets of data which cannot be detected through the observation of raw frequencies alone. Section 4 will return to Table 7 and offer an interpretation, arguing that the presented perspective is not only different from raw frequencies, but also more instructive. Before that, section 3 turns to practical considerations to further justify the proposed method.

3. Why is it useful?

Since distinctive collexeme analysis was not developed with the particular goal of doing historical studies, it is fair to inquire whether such an application is in fact useful. This section addresses possibilities and limits of the method.

A potential criticism follows from a general limitation of distinctive collexeme analysis. A synchronic distinctive collexeme analysis does not take into account the overall corpus frequencies of the lexical elements that occur with the compared constructions. It merely highlights differences, not characteristics that are typical of the respective constructions per se. The same limitation holds for the diachronic application, which disregards the fact that lexical elements may very well have a different relative frequency in each sub-corpus. Words may not only change in frequency, but also occasionally fall out of use entirely, such that the treatment of several periods of time as a unified corpus is clearly an idealization. This criticism would be met by the application of separate collexeme analyses (Stefanowitsch and Gries 2003) for each investigated period of time. Separate collexeme analyses take into account that each sub-corpus contains a different set of verbs with different relative frequencies, and can thus characterize the collocational preferences of a construction at a certain stage in time more faithfully. These preferences can then be compared across diachronically sequential corpora.

Despite the superiority of this approach, there is still room for distinctive collexeme analysis in diachronic analyses, which is due to the nature of most available data. Collexeme analyses depend on tagged corpora to facilitate the exhaustive retrieval of the investigated construction and of co-occurring lexical items, whose frequencies in the construction and elsewhere enter the statistical calculation. Determining the frequency of a certain infinitive form is in some cases made difficult through homograph forms. For example, the form *care* instantiates a different part of speech in *take care*, *I care*, and *I don't care*. A collexeme analysis based on a large untagged corpus therefore involves much work in addition to retrieving and editing all examples of the investigated construction. Luckily, some tagged diachronic corpora such as the PPCEME are available for English, but for other languages there are few resources of this

kind. And even with its 2.5 million words representing time periods that cover 210 years, the PPCEME may neither be large enough nor temporally broad enough for certain analyses. Distinctive collexeme analysis has the advantage that it can operate on untagged data, since all necessary data points are contained in the concordances of the investigated construction. This means that tagged and untagged corpora, such as the PPCEME and the CLMET, can be combined freely to form larger databases. The method can also operate on diachronic corpora that were created from scratch, drawing on digital databases such as the Project Gutenberg. Since distinctive collexeme analysis does not process the sizes of the used corpora, it is further possible to use sources like the citations from the Oxford English Dictionary. The OED's exact size in words is not even known to its current editors (p.c.), but each citation is indexed with its year of publication, which makes it very practical for diachronic investigations. As all other corpus-based investigations that use historical data, a distinctive collexeme analysis will yield better and more reliable results with databases that are as large and as balanced as possible.

While many caveats need to be kept in mind, the method offers benefits that justify its application. The diachronic application of a distinctive collexeme analysis can offer results that bear on questions of theoretical relevance. As the method focuses on the level of grammatical constructions and documents changes in their usage, it is well-suited to address questions in the framework of grammaticalization theory, where the importance of the constructional level is increasingly often acknowledged (Traugott 2003, 2005, Noël 2006, Bergs and Diewald forthcoming).

One application of the method is to explore how a given construction changed over time, but another probably more important one is the testing of pre-existing hypotheses. Results in cross-linguistic studies of grammaticalization often take the shape of grammaticalization clines that describe a semantic change from a lexical concept to a grammatical function. Well-known examples are motion verbs that turn into future markers or body part terms that become spatial adpositions (Heine and Kuteva 2002). Distinctive collexeme analysis affords a way to follow the grammaticalization path of a given construction and to test whether the observed changes in its collocational behavior match with the proposed semantic development.

4. What does it show?

The differences between Table 5 and Table 7 show that the proposed method offers a perspective on historical data that goes beyond raw frequencies. This section discusses how these new results can be interpreted. The basic claim made here is that differences in the most distinctive collexemes represent on-going semantic change. A comparison of signifi-

cantly distinctive elements at different times therefore allows for an exploration of the semantic change that a construction has undergone. As pointed out in the previous section, pre-existing claims about semantic developments can be operationalized in such a way that they are testable against the observed shifting preferences.

In the case of *shall*, several proposals have been put forward with regard to its historical development. Traugott (1989: 35), in accordance with Bybee and Pagliuca (1987), argues that *shall* underwent a change from the meaning of obligation to the meaning of futurity. As the earliest occurrences of *shall* with future meaning considerably pre-date the PPCEME (Visser 1969: 1692, OED: *shall*, v1), data from this corpus cannot determine how exactly the meaning of obligation of *shall* gave rise to the meaning of futurity. Still, Table 7 offers an opportunity to analyze the changes that *shall* has undergone between the 16th and 20th century.

An observation that is of direct relevance to the present study is presented by Traugott (1989: 41), who argues that the semantic development of *shall* follows a trajectory towards meanings that are increasingly situated in the beliefs or attitudes of the speaker. The semantic change of subjectification is not only observed with *shall*, but with many grammaticalizing forms across different languages. With respect to the present study, Traugott's finding predicts that later uses of *shall* should show a stronger affinity to verbal complements that express subjectified, speaker-based meanings. This can actually be extrapolated from the elements in Table 7. Any interpretation needs to take into account both the lexical semantics of the observed elements, as well as the pragmatic functions with which they are used in actual examples.

The first period from 1500 to 1640 is characterized by several distinctive collexemes that encode the closely related concepts of perception and appearance, as illustrated in (1).

(1) a. Furthermore, ye shal vnderstand that the brayne is a member colde and of moyste complexion.
    b. But bycause my Trueth and his Falsehood shall the better appear unto you, I will declare his Inconstancy in vttering this his Euidence.
    c. Ye the tyme shall come, that whosoever killeth you, will thinke that he doth God service. (a–c: PPCEME)

With the most distinctive collexeme *understand*, 42 of the observed 48 examples are in the second person and allow an imperative reading. This tendency reflects that obligation, which is no longer expressed by *shall* in most modern varieties of British English, is still a strong semantic

component of the construction in the 16th and 17th century. The affinity towards the meaning of obligation does not mean that the future meaning has not yet grammaticalized; example (1d) shows that future meanings with abstract subject referents are fully possible with *shall* at this stage of English. The distinctive collexeme *come* appears less often in the concordance as a motion verb than as a verb that indicates abstract changes.

In the second period, a shift away from the meaning of obligation towards the meaning of intention can be observed. This is evidenced by several meta-linguistic verbs that appear as distinctive collexemes. The verbs *mention*, *suppose*, and *explain* explicitly denote a communicative act, while the verbs *endeavour*, *discover*, *examine*, *confine*, and *add* are not semantically meta-linguistic, but can be used as such, as illustrated in the examples in (2).

(2) a. I shall only recall on this occasion one of these arguments, which I shall endeavour to render still more conclusive.
(a–b: CLMET)

b. But I shall now no longer confine my remarks to single errours, but observe that there is one general defect, by which the whole bill is made absurd.

c. To the foregoing Experiments, whose success is wont to be uniform, I shall adde the Recital of a surprising Phaenomenon.
(PPCEME)

It has been observed that *shall* in modern British English is largely confined to first person uses and to conservative written genres (Wekker 1976, Close 1977, *inter alia*). These tendencies begin to conventionalize in the 17th and 18th century, when *shall* is increasingly used by writers to express their intentions of structuring a text with sequential parts. The use of a future construction is motivated in this context, because there is a temporal sequence in the production as well as in the reception of a text. The semantic shift from obligation to intention can be interpreted as subjectification. Both concepts involve a force which is meant to compel an agent to act, but this force is external to the subject in the former, and subject-internal in the latter.

In the third period, the construction continues to be attracted towards meta-linguistic verbs such as *write*, as well as towards the verbs *see* and *leave*, which occur in examples with meta-linguistic meaning. Emerging meanings that continue the trend of subjectification are the functions of expressive and commissive speech acts. The distinctive collexemes *forget* and *feel* occur in examples in which the speakers disclose their personal emotions, and thus make expressive speech acts. Examples with the distinctive collexemes *go*, *get*, and *try* convey commissive speech acts, in

which the speakers make a promise or signal their willingness to comply with a request. Illustrating examples are given in (3).

(3) a. You see, Charlotte, your kindness − I shall never forget it!
b. We shall feel the loss of these two most agreeable young men exceedingly.
c. 'You will really do that?' 'Yes. I shall go and look for rooms to-morrow.'
d. I shall try to be as little disturbance to you as possible.
(a−d: CLMET)

Expressive and commissive speech acts convey even further subjectified meanings than declarations of intentions. In the examples in (2), speakers announce their intentions about events that will be objectively observable, such as writing up an argument. By contrast, the events denoted in examples (3a) and (3b) are not objectively verifiable, as they are internal to the speaking subjects. Similarly, in (3c) and (3d), the felicity of the given promises depend on mental states that are internal to the speaking subjects.

In summary, the data in Table 7 allow the identification of a semantic change that follows the trajectory of subjectification. The application of a distinctive collexeme analysis to historical data can thus inform analyses that address theoretical issues in frameworks such as grammaticalization theory, and it can supply a perspective that brings developments into focus which would remain elusive on inspection of mere raw frequencies. It can therefore be concluded that the proposed application is a useful tool for the study of grammaticalizing constructions. As pointed out by Gries and Stefanowitsch (to appear), collostructional analyses do not render the interpretative work of a human analyst obsolete. This note hopes to reaffirm the view that it can in fact make it more fruitful, even in diachronic analyses.

Note

* I would like to thank Suzanne Kemmer for our many discussions on this topic. Thanks are also due to Anatol Stefanowitsch and Stefan Th. Gries for patiently answering the numerous questions that I have had about collostructional analysis. Yet, it seems like there will be more questions in the future.

References

Allan, Robin, Philip Holmes, and Tom Lundskaer-Nielsen
2000 *Danish. A comprehensive Grammar.* London: Routledge.
Berglund, Ylva
1997 Future in Present-Day English: Corpus-Based Evidence on the Rivalry of Expressions. *ICAME Journal* 21, 7−19.

Bergs, Alexander and Gabriele Diewald (eds.)
to appear *Constructions and Language Change*. Berlin: Mouton de Gruyter.
Binnick, Robert I.
1971 Will and Be Going To. *Chicago Linguistic Society* 7, 40−51.
Close, Reginald A.
1977 Some Observations on the Meaning and Function of Verb Phrases Having Future Time Reference. In W. D. Bald and R. Ilson (eds.), *Studies in English Usage: The Resources of a Present-Day English Corpus for Linguistic Analysis*. Frankfurt am Main: Peter Lang, 125−56.
De Smet, Hendrik.
2005 A corpus of Late Modern English Texts. *ICAME Journal* 29, 69−82.
Gries, Stefan Th.
2004 *Coll.analysis 3. A program for R for Windows 2.x.*
Gries, Stefan Th. and Anatol Stefanowitsch.
2004 Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9/1, 97−129.
to appear Cluster analysis and the identification of collexeme classes. In J. Newman and S. Rice (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford, CA: CSLI.
Gilquin, Gaëtanelle
to appear The verb slot in causative constructions. Finding the best fit. *Constructions*.
Haegeman, Liliane M.V.
1989 Be going to and will: A Pragmatic Account. *Journal of Linguistics* 25, 291−317.
Heine, Bernd and Tania Kuteva
2002 *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Kemmer, Suzanne and Martin Hilpert
2005 Constructional grammaticalization in the make-causative. Paper presented at the 17th ICHL, Madison, WI.
Kroch, Anthony, Beatrice Santorini, and Lauren Delfs
2004 *The Penn-Helsinki Parsed Corpus of Early Modern English*. URL <http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-1/>
Leech, Geoffrey N.
1992 *100 million words of English: the British National Corpus*. Language Research 28/1, 1−13.
Lemmens, Maarten
2005 Aspectual posture verb constructions in Dutch. *Journal of Germanic Linguistics* 17/3, 183−217.
Noël, Dirk
2006 Diachronic construction grammar vs. grammaticalization theory. Leuven: Katholieke Universiteit Leuven. Preprint Nr. 255.
Stefanowitsch, Anatol
2001 *Constructing causation: A construction-grammar approach toanalytic causatives*. Doctoral dissertation, Rice University
Stefanowitsch, Anatol and Stefan Th. Gries
2003 Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8/2, 209−43.
Traugott, Elizabeth
1989 On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 57/1, 33−65.

2003 Constructions in grammaticalization. In Brian Joseph and Richard Janda (eds.), *The Handbook of Historical Linguistics*. Oxford: Blackwell, 624–47.

2005 On the relation between constructions and grammaticalization. Paper presented at the 17th ICHL, Madison WI.

Visser, Frederikus Th.

1963 *An historical syntax of the English grammar*. Leiden: Brill.

Wekker, Hans Christian

1976 *The Expression of Future Time in Contemporary British English*. Amsterdam: North Holland.