3

Converging Evidence II: More on the Association of Verbs and Constructions

STEFAN TH. GRIES, BEATE HAMPE, AND DORIS SCHÖNEFELD*

1 Introduction

1.1 General Introduction

Investigations within Cognitive Linguistics, Construction Grammar, and related schools of thought in linguistics are increasingly required to be 'usage-based'. As a result, many researchers from these fields have turned to samples of language produced in authentic contexts, i.e. corpora, or to experimentation, rather than basing their inquiries on constructed and isolated sentences. While this methodological shift has already resulted in a substantial increase of descriptive and explanatory reliability and validity, we believe that its full potential has not yet been utilized. More specifically, 'usage-based' approaches can in principle encompass both corpus-based and experimental perspectives, with each drawing on different kinds of data and yielding different kinds of results. Nevertheless, 'usage-based' has mainly been taken to mean 'corpus-oriented' and there are few studies in which an individual topic is tackled from more than one methodological perspective, producing what is commonly referred to as 'converging evidence' (cf. Goldberg et al. 2004; Gries 2003; Gries, Hampe, &

*The author order is arbitrary. The authors would like to thank Ken Forster for his extensive help with the reading-time study and the audience at CSDL 2004 for a variety of very helpful comments. The usual disclaimers apply.

Empirical and Experimental Methods in Cognitive/Functional Research.
Sally Rice and John Newman (eds.).

Schönefeld, 2005; Nordquist 2004).

The aim of the present paper is to advocate such a research strategy and exemplify its potential with respect to the following important methodological shift: Given the upsurge in corpus-based research within linguistics, many researchers no longer just use corpora as databases from which examples fitting one's account can be arbitrarily chosen, but rather use the frequency distributions of morphemes, words, and constructions as data such that, for example, more frequent units are considered more typical, more entrenched, more basic, etc. Though this development is generally highly welcome, what seems to have escaped notice is that reporting frequencies and percentages alone is often quite uninformative or even problematic, as studies within corpus linguistics have long shown (cf. Manning & Schütze 2000: Chapter 5).

This paper investigates and characterizes the use of a clause-level construction in English. More specifically, the converging evidence approach (i) illustrates the potential shortcomings that simple frequency data are often fraught with and (ii) demonstrates that an alternative method, *collexeme analysis* (cf. Stefanowitsch & Gries 2003), yields superior results in investigations of the associations between words and constructions.

Section 1.2 introduces the construction we investigate and Section 2 briefly summarizes an earlier case study bringing together evidence from corpora and from a sentence-completion experiment (for details see Gries, Hampe, & Schönefeld, 2005). The following two sections illustrate how we set out to strengthen our case: Section 3 discusses how we extended the corpus data, while Section 4 introduces further experimental evidence from a reading-time experiment. We conclude with Section 5.

1.2 The *as*-predicative

The syntactic construction we investigate is the *as*-predicative, a partially lexically-filled complex-transitive pattern in which the word *as* introduces an ive complement; examples from corpus data are listed in (1).

(1) a. I do not regard the Delors Report *as* [$_{NP}$ some kind of sacred text].
b. Such a force could never be described *as* [$_{P}$ purely deterrent].
c. We see the hard ECU *as* [$_{S}$ being extremely useful].

Why is the *as*-predicative a *construction*, i.e. an entrenched form-meaning pairing (cf. Goldberg 1995)? For reasons of space, we cannot discuss this issue here in detail (for a more exhaustive treatment, see Gries, Hampe, & Schönefeld, 2005), so a brief discussion must suffice. First, unlike complex-transitive structures without *as*, the object complement of *as*-predicatives is uncommonly flexible syntactically since it can take a variety of different phrasal elements: NPs, AdjPs, non-finite clauses and PPs. A second and related point is that, given this structural flexibility, *as* cannot be considered a regular preposition, although it tends to be treated as such in traditional grammatical analyses (cf. Quirk et al. 1985: 1200), and is usually tagged as such in the ICE-GB (a tagged and parsed corpus of British English from the early 1990s) and the BNC-sampler. Thirdly, like other argument structure constructions, the *as*-predicative has a variety of related meanings most of which involve the notion of epistemic stance. For example, the V slot of the *as*-predicative takes a variety of mental activity verbs (e.g. *regard, view*, *recognize*, *consider*, *think of*) denoting situations where the subject of the clause conceptualizes the referent of the NP_{object} as either possessing an attribute or representing an instance of the category denoted by the $XP_{complement}$. Then, there are 'characterization'/'speech-act' verbs (e.g. *define*, *describe*, *portray, depict*) denoting activities as a result of which the referent of the NP_{object} is characterized in either of these ways. Yet another sense is instantiated by a group of 'classification' verbs (e.g., *categorize, class*, *diagnose*) closely related to both of the former groups. While these verbs usually have both a mental activity and a speech-act sense, they restrict the relation between the NP_{object} and $XP_{complement}$, in that they uniformly present the former as a member of the class denoted by the latter. On the basis of these formal and semantic characteristics, we submit that the *as*-predicative is a construction in the Construction Grammar sense of the term.

2 Collostructional Strength vs. Frequency in Sentence Completion

In a previous study (Gries, Hampe, & Schönefeld, 2005), we set out to explore which of the two approaches—absolute/relative frequency vs. collexeme analysis—yields more promising results. To that end, we compared the results of a corpus-based collexeme analysis to those of an experimental sentence-completion task.

As a first step, we extracted all occurrences of the search pattern [$_{VP}$ $V_{complex\text{-}transitive}$ [PP *as*]] from the ICE-GB. For this, we used the retrieval software that comes with the corpus (ICECUP). After the manual correction of the output, we obtained the frequencies of 107 verb types in the *as*-predicative, amounting to 687 tokens. As a second step, we conducted a collexeme analysis (cf. Stefanowitsch & Gries 2003) of the *as*-predicative to identify the verbs which are most strongly associated with the *as*-predicative and are, thus, most representative of the construction's semantics (cf. Gries, Hampe, & Schönefeld, 2005, for detailed discussion). Let us first clarify (i) what *collexeme analysis* is and (ii) how far collexeme analysis differs from traditional frequency-based analyses.

As for the former, collexeme analysis is the first of a family of methods called *collostructional analysis* (a blend of *collocational* and *constructional*). It extends established corpus linguistic methods of quantifying the collocational strength between words (i.e. collocates) to the quantification of the association between constructions (in the Construction Grammar sense of the term) and the words occurring in a given constructional slot (the so-called *collexemes*). This measure has become known as *collostructional strength*.

Regarding the difference between a frequency-based analysis and a collexeme analysis, consider the illustrative data given in Table 1 on the relation between the *as*-predicative and the verb *regard*.

	as-predicative	other constructions	row totals
regard	80	19	99
other verbs	607	137,958	138,565
column totals	687	137,977	138,664

Table 1. Input data for a collexeme analysis of *regard* and the *as*-predicative in the ICE-GB (from Gries, Hampe, & Schönefeld, 2005)

A traditional frequency analysis of *regard* in the *as*-predicative would be based either on the absolute frequency of *regard* in the *as*-predicative (i.e. the frequency 80) or on the relative frequency of *regard* in the *as*-predicative (i.e. the frequency $^{80}/_{687}$=11.6%). However, as is argued in Stefanowitsch & Gries (2003) and in more detail in Gries, Hampe, & Schönefeld (2005), this approach encounters problems since it neglects the frequency of *regard* in the whole corpus, i.e. 99 or $^{99}/_{138,664}$=0.0007%. A collexeme analysis, by contrast, takes into consideration all frequencies of Table 1 as well as their ratios and computes each verb's collostructional strength within the construction under investigation. The results are given as the negative log to the base of 10 of the Fisher-Yates exact test. The 21 most strongly attracted verbs (i.e. collexemes) resulting from this analysis are listed in Table 2.[1]

The third step was to test the predictive power of the frequency-based results against those of the collexeme analysis by means of a sentence-completion task. More specifically, the objective was to see whether the completion of sentences by native speakers could be predicted more accurately on the basis of a particular verb's frequency in the *as*-predicative—the traditional approach—or its collostructional strength to the *as*-predicative. The computation of collostructional strength requires the verbs'

---

[1] All collexeme analyses were performed with Coll.analysis 3, an interactive R program (Gries 2004), which is available from the first author.

frequencies of occurrence (see above Section 1.1). We then form four groups of verbs, crossing the factors of a verb's collostructional strength to the *as*-predicative (COLLSTRENGTH: *high* vs. *low*) with that of a verb's frequency in the *as*-predicative (FREQUENCY: *high* vs. *low*). Ultimately we arrive at four groups of verbs for which different predictions can be made.[2] The frequency approach predicts that verbs with a high frequency in the construction should yield more *as*-predicative continuations irrespective of the verbs' collostructional strength while the collostructional approach predicts that verbs with a high collostructional strength should yield more *as*-predicative continuations irrespective of the verbs' frequency in the construction.[3]

Collexeme	Coll. strength	Collexeme	Coll. strength
regard	166.48	*categorize*	11.53
describe	134.87	*perceive*	8.3
see	78.79	*hail*	6.32
know	42.8	*appoint/interpret*	6.07
treat	28.22	*class*	5.92
define	23.84	*denounce*	5.38
use	21.43	*dismiss*	5.16
view	17.86	*consider*	5.08
map	12.8	*accept*	4.47
recognize	12.16	*name*	4.28

Table 2. The 21 most strongly attracted collexemes of the *as*-predicative in the ICE-GB (search pattern: [$_{VP}$ V$_{\text{complex transitive}}$ [PP *as*]])

Sixty-four native speakers of British English were given 24 sentence fragments consisting of a subject and a verb. These contained 8 experimental items and, for distraction, 16 filler items in a pseudo-randomized order, so that each subject saw each verb and each experimental condition only once. Subjects were asked to complete the sentence fragments so as to produce normal English sentences. This yielded 493 responses, of which 150 could be coded without doubt as instances of the target construction. The results showed that—while FREQUENCY did not have a significant effect—COLLSTRENGTH had a highly significant effect, and the by far strongest one (partial η^2=0.123; p<0.0001); the relevant interaction is depicted in Figure 1.[4]

---

[2] In addition, we also included a third factor, namely the voice of the sentence fragment to be completed: VOICE: *active* vs. *passive*.

[3] Of course, FREQUENCY and COLLSTRENGTH are correlated; we will return to this below.

[4] We also found that COLLSTRENGTH outperformed the verbs' subcategorization probabilities concerning the *as*-predicatives, i.e. the verbs' probabilities $^{a}/_{a+b}$ referred to as FAITH.

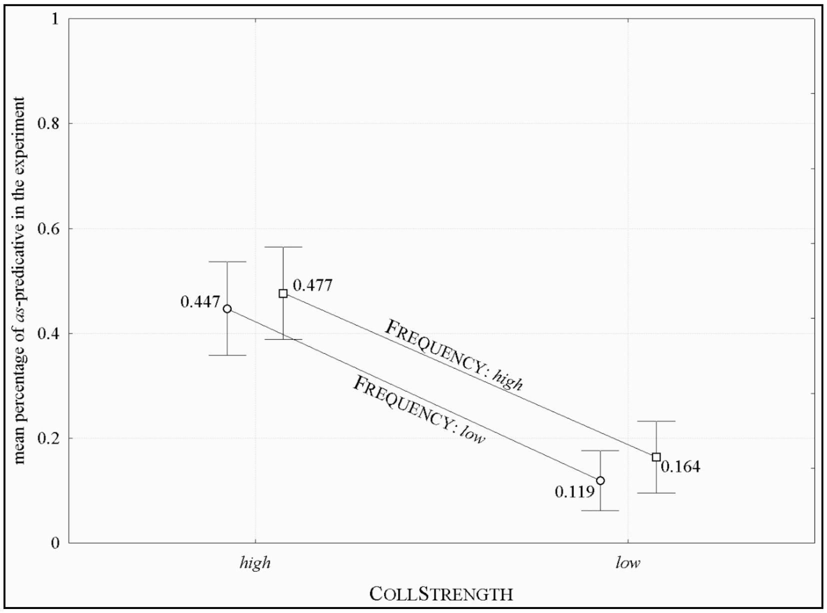

Figure 1. COLLSTRENGTH × FREQUENCY on the relative frequencies of *as*-predicatives

While these results clearly show that the collostructional approach is superior to the frequency approach (cf. Gries, Hampe, & Schönefeld, 2005 for more detailed discussion), we subsequently discovered a problem with the data from the ICE-GB. According to the corpus compilers, the annotation of the corpus has been "fully checked" (http:// www.ucl.ac.uk/english-usage/ice-gb/). However, in a few cases we noted that the actual number of *as*-predicatives seemed somewhat higher than that obtained on the basis of our search parameters. We therefore decided to broaden the basis of our retrieval by searching for *as* alone (tagged as a preposition) and doing the remaining coding manually. Additionally, we further extended the data set and carried out a similar search in a different, but theoretically comparable corpus, namely the sampler of the British National Corpus (BNC) containing two million words of running text balanced for spoken and written registers.

3 Extending the Corpus Data

The extraction of all instances of *as* tagged as a preposition from the ICE-GB and the BNC sampler and subsequent manual identification of all occurrences of the *as*-predicative yielded the following results. In contrast to the results of our first search, this time we obtained 1,131 instances of the *as*-predicative (i.e. 65% more than in the first attempt on the basis of the corpus parse) comprising 261 verb types. From the BNC sampler, we obtained 1,251 tokens of the *as*-predicative, comprising 309 verb types. At this point the question arose whether it would be feasible and legitimate to actually merge the two corpora in order to provide a more solid foundation for our corpus analysis and thus create a more robust verb classification on which to base further experimental studies.

Even though the corpora have been designed with a similar general focus on British English from the 1990s, it is by no means obvious that the corpus results on the *as*-predicative from the two corpora are similar enough, i.e., reasonably comparable in order to justify such a merging procedure. While there are some studies on how to compare different corpora, most of these are based on word frequencies, which is of course a feasible approach (cf. Johanson & Hofland 1989; Kilgarriff & Rose 1998; Kilgarriff 2001; Rayson & Garside 2001). As our main focus is on the relation between verbs and their occurrence in the *as*-predicative, we did collexeme analyses for all corpora in isolation as well as one collexeme analysis on the basis of the merged corpus data. We then investigated the similarity of the two corpora and their collexeme lists in three different ways. Table 3 summarizes the results of the collexeme analysis on the basis of both corpora.

Collexeme	Coll. strength	Collexeme	Coll. strength
regard	inf	*refer to*	29.21
describe	209.39	*recognize*	28.23
see	137.15	*class*	27.27
use	124	*interpret*	21.35
treat	97.38	*perceive*	19.55
know	73.1	*hail*	19.47
think of	52.31	*classify*	17.04
define	37.99	*present*	16.91
consider	30.32	*map*	16.48
view	29.75	*categorize*	14.75

Table 3. The 20 most strongly attracted collexemes of the *as*-predicative in the corpus resulting from merging the ICE-GB and the BNC sampler

A quick comparison of Tables 2 and 3 already indicates a high degree of similarity. To legitimize the merging procedure, we first tested whether there are significant correlations between the observed frequencies and the collostructional strengths of shared verbs in the *as*-predicative across the two corpora.[5] As it turned out, the correlations for both are high and highly significant ($r=0.94$; $t=32.06$; $p<0.001$ *** for the observed frequencies of shared verbs and $r=0.94$; $t=32.07$; $p<.001$ *** for the collostructional strength of shared verbs, respectively).

The degree of compatibility was further determined by testing whether, in the data set obtained from the merged corpora, the verbs originally chosen for the four experimental groups would still fall into the same classes on the basis of their values for frequency (*high*/*low*) and collostructional strength (*high*/*low*). A simple cross-tabulation revealed that their value combinations were as in the ICE-GB before. For example, the verb group for COLLSTRENGTH: *low* and FREQUENCY: *high* in the merged data still exhibited means representing the very same tendency when compared to all experimental items.

As a final test, we used a Monte Carlo-like simulation technique. This simulation procedure tests how much overlap one may expect between any two corpora by chance and compares this expected degree of overlap to that actually obtained for *as*-predicatives. Initially, from the number of verb types determined by the collostructional analysis of the data obtained from the ICE-GB and BNC sampler, respectively (i.e. 261 and 309), simplified lists (i.e. 231 and 286) were produced for use in the simulation procedure.[6]

The observed number of verb types shared by the construction in both corpora is 136. Most crucially, we had to determine whether an overlap of this size is due to chance or whether it represents a significant deviation from what would have to be expected. To this end, we determined the frequencies of all verb lemmas in both corpora by extracting all forms tagged as verbs and then identifying the relevant items using a mixture of semi-automatic and traditional manual coding.[7] To ensure a maximum degree of reliability, we generated 50,000 random samples of 286 verb lemmas from the BNC sampler.[8] In each of these 50,000 samples we counted how many

[5] The procedures to check the legitimacy of the merger were developed and carried out exclusively by the first author.

[6] This is due to the fact that, in a small number of cases, the analysis had to be simplified by reducing phrasal verbs to their verbal component only, e.g., *build up* was subsumed under the verb type *build*. Given that the relevant types constitute only a very small portion of the corpus overlap, we are confident that this had no influence on the accuracy of the results.

[7] The procedure required some decisions as to verb identity. For example, we decided to code spelling variants of the *-ise*/*-ize* and *-yse*/*-yze* types as belonging to a single lemma; similarly, cases of contractions (e.g. *fussin'* and *'m*) were coded as belonging to their non-contracted lemmas. Finally, we discarded within-verb hyphenation and coded the few Middle English forms we found as belonging to their ModE equivalents.

[8] In this connection the sampling was weighted by word frequency, i.e. the probability of a verb lemma from one corpus to be sampled was proportional to its overall frequency in that corpus. We adopted this strategy because if every verb lemma had had the same chance of

of the verb lemmas thus obtained were actually verb lemmas also occurring in the *as*-predicative in the ICE-GB. In addition, we also performed the same simulation the other way round, generating 50,000 random samples of 231 verb lemmas from the ICE-GB. Again, we counted how many of these verb lemmas also occur in the *as*-predicative in the BNC sampler. As a result, we obtained one normally distributed sample for each corpus. Given the similarity of these samples, we collapsed them into one for the final evaluation. Firstly, we checked how often the expected overlap of randomly sampled verb lemmas reached the observed number of 136. This did not happen a single time. In other words, p (i.e., the probability to obtain the observed overlap just by chance), is smaller than 0.00001. Secondly, the perfectly normal sampling distribution allows for computing the average expected overlap and its confidence interval. As it turns out, the average expected overlap with a 99.9% confidence interval is 76.3±0.05 verb lemmas. Both of these results show that the overlap of 136 verb lemmas found in our data is virtually impossible to obtain on the basis of chance alone. That is to say, the similarity between the verb lemma lists resulting from the collostructional analysis of the two corpora is enormous.

While this discussion of the similarities between the two corpora may seem overly thorough and technical, it was necessary to ensure that the merging of the two corpora can in fact be justified. What is more, it offers interesting strategies for future studies where a successive accumulation of data appears desirable and should therefore be interesting to corpus-minded linguists of all persuasions. Be that as it may, the data show that the two corpora are very similar with respect to how the *as*-predicative is used. That is why we merged the results from the two corpora to increase the reliability and precision of the collostructional analysis and, thus, to obtain a more refined set of stimulus items for the reading-time experiment.

4 Collostructional Strength vs. Frequency in a Reading-time Study

4.1 Data

On the basis of the merged corpus data, we compiled sets of verbs that systematically crossed the three factors: COLLSTRENGTH, FREQUENCY in the *as*-predicative, and VOICE. From these verbs, we chose those given in Table 4 to test in a reading-time study.

being sampled, we would have neglected the fact that high frequency verbs such as *do* or *go* already result in relatively high chances of overlap—our approach integrates this possibility and is, therefore, both more realistic and more conservative.

We then composed the test sentences around these verbs. In fact, the sentences were based on authentic sentences from the BNC (World Edition) containing these verbs, but were altered so that they were comparable both in the length and the complexity of their constituent phrases. Additionally, sentence elements that might have caused comprehension problems were replaced by less context-dependent or better-known expressions (for example, proper names, such as *De Klerk* were replaced by *Smith*). Finally, 95 sentences were generated as filler items to distract the subjects' attention from the construction under investigation.

	FREQUENCY: *high*	FREQUENCY: *low*
COLLSTRENGTH: *high*	*see, regard, treat, think_of, view, class, hail, portray*	*preselect, enunciate, depose, ridicule, personify, elicit, cultivate, evoke*
COLLSTRENGTH: *low*	*find, leave, pay, suggest, grow, form, discuss, obtain*	*want, provide, understand, apply, collect, take on, appreciate, handle*

Table 4. Verbs used in the reading-time study

The experimental stimuli were then tested with 33 subjects, who were told that the purpose of the experiment was to find out how easily people can understand sentences. Each subject saw 24 sentences, of which 8 contained verbs selected to represent each of the experimental conditions (with no repetition of verb types), while the remaining 16 items served as fillers. The experiment was designed as a non-cumulative segment-by-segment self-paced reading experiment. That is, the subjects were presented with the sentences in a word-by-word fashion such that they pressed a button to request the next word as soon as they thought they understood the sense of the part of the sentence already encountered. For the dependent variable, we measured the time from the presentation of a word until the subjects pressed the button to get the next word. After each sentence, a comprehension question was asked to mask the purpose of the experiment. In order to respond, subjects were required to select the correct answer from two alternatives by pressing another key. Five practice trials were presented initially to familiarize the subjects with the experimental procedure.

As a result, we obtained 254 reading times. In order to assess the effect of COLLSTRENGTH and FREQUENCY on reading time, we first considered using the reading times obtained for *as*, since here the subjects get an explicit clue that they are encountering an *as*-predicative. However, some reading-time studies have shown potential effects to appear one word after the facilitatory element (cf. Just, Carpenter, & Woolley 1982; Hare, McRae,

& Elman 2003: 292-4).[9] For our experiment, we consider relying on the reading times of the word immediately following *as* the most useful strategy, since this word is often needed to determine whether the stimulus presented constitutes an *as*-predicative rather than a clause of a completely different kind (e.g., *He saw the problem as well as the solution*). Furthermore, reading time may also be influenced by general word token frequency. Thus, we also determined the token frequencies of the words at position *as* + 1 in the BNC (World Edition), entering their base-10 logarithms into the analysis.[10] The effects were then investigated by means of an ANCOVA with three independent factors (COLLSTRENGTH: high vs. low; FREQUENCY (in the *as*-predicative): high vs. low; VOICE: active vs. passive) and one covariate (TOKENFREQUENCY). The results are summarized in Table 5. Figure 2 represents the interaction COLLSTRENGTH × FREQUENCY.

Factor / interaction	F	p	effect size
TOKENFREQUENCY	0.257	0.612	0.001
VOICE	0.180	0.672	0.001
COLLSTRENGTH	3.438	0.065	0.014
FREQUENCY	1.111	0.293	0.005
VOICE × COLLSTRENGTH	0.021	0.886	0.000
VOICE × FREQUENCY	0.053	0.819	0.000
COLLSTRENGTH × FREQUENCY	0.609	0.436	0.002
VOICE × COLLSTRENGTH × FREQUENCY	0.622	0.431	0.003

Table 5. Results of the ANCOVA

4.2 Results and Discussion

The results in Table 5 show that no single factor or interaction reaches standard levels of significance. Although this does not appear to be a very encouraging result, it is probably due to the relatively small number of subjects. 254 reading-time observations proved to be too few for eight experimental conditions, an assumption also supported by the fact that the observed power for this effect is only 0.432 rather than the usually recommended value of 0.8. On closer inspection, it turns out that the results do in fact exhibit a strong tendency conforming to the expectations we had derived both from the theoretical approach and from the first experimental

[9] We thank Harald Baayen for pointing this out to us.

[10] We took the logarithm of the frequency rather than the frequency as such since the logarithm of the frequency allows for identifying linear correlations with other interval variables. In addition, log (token frequency) is known to correlate with psycholinguistic processes such as word recognition (cf. Howes & Solomon 1951).

results. COLLSTRENGTH has a marginally significant effect, while FREQUENCY clearly has not. Similarly, the effect size of COLLSTRENGTH is nearly twice as high as that of FREQUENCY. Lastly, as the investigation of the means for both variables shows, the results are also in the expected direction. Both COLLSTRENGTH: *high* and FREQUENCY: *high* result in shorter reading times than the corresponding low counterparts, and only the gain resulting from COLLSTRENGTH is marginally significant. However, we regard this as evidence in favor of collostructional strength as the more powerful predictor of the subjects' performance—the results are simply impossible to account for if one assumes that FREQUENCY is the decisive factor.

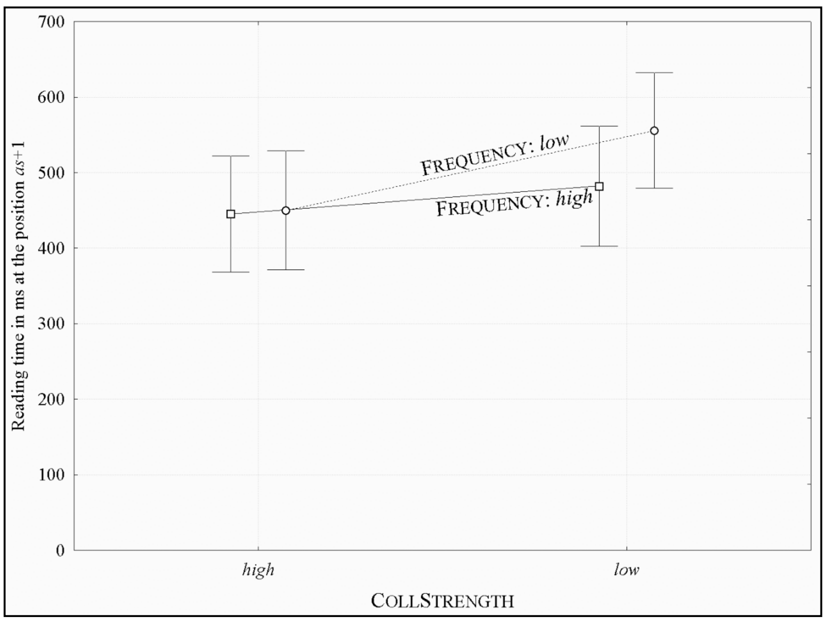

Figure 2. The effects of COLLSTRENGTH × FREQUENCY on the reading times of *as*-predicatives

5 Conclusion

This is our second study investigating which kinds of corpus data are most useful within usage-based approaches to the analysis of the association between verbs and constructions. In our first study (Gries, Hampe, & Schönefeld, 2005), we contrasted a more traditional and widespread method based on absolute or relative frequencies with the more recent method of collexeme analysis. In order to assess the predictive power of the verb rankings produced by the respective methods, we carried out a sentence-completion experiment, obtaining a very strong effect in favor of collostructional strength. In the present study, we firstly improved the database in two ways: (i) by employing a less constraining search strategy based on the word-class tags, and (ii) by merging the results from two different, but comparable, corpora. In this regard, we outlined a method for testing whether data from different corpora are homogeneous enough to be merged.

Secondly, we found additional experimental evidence in favor of collostructional strength on the basis of a reading-time study. Though the results from the reading-time study were not as clear as those obtained in our sentence-completion experiment, they point in the same direction and can also be taken to argue against a mere frequency-based explanation. On the basis of these results, we feel that we have made a strong case against one of the most prominent corpus-based methods in contemporary cognitive/functional linguistics, namely the use of frequency data. In other words, while cognitive linguists regularly regard frequency data as directly reflecting the degree of routinization or entrenchment, we have shown that (i) frequency alone runs the risk of severely misrepresenting speakers' behavioral patterns and that (ii) collostructional strength outperforms frequency as a predictor of speakers' behavior in both production and comprehension tasks.

As the title of the paper suggests, we therefore strongly recommend the employment of evidence from different methodologies. As summarized above, this paper demonstrates how generalizations drawn from corpus results can be validated in what can be called an empirical cycle. Not only did we use corpus results to validate corpus results, we also put our hypotheses—made on the basis of an analysis of the usage data thus corroborated—to a number of experimental tests (cf. Schönefeld 1999: 165f. and 2001: 110-3). As for the use of frequency data, our procedure has made explicit that frequencies must be normalized and checked against chance levels before they can be interpreted.

Apart from this general conclusion, which cannot be emphasized enough, there are a few other methodological implications that merit mentioning. For example, we have demonstrated that relying on automatic corpus tagging can threaten the reliability of the data considerably. In the present study, the number of hits was increased by nearly two thirds once we performed a more exhaustive and almost fully manual retagging. While it is admittedly time-consuming and did not alter our initial results, computational shortcuts as they are frequently found in computational and corpus linguistics may seriously undermine the reliability of the data. Manual coding is the only way to guarantee maximum precision and recall. In addition, we have shown that merging data from different corpora can be quite useful and we have exemplified a few strategies to test the legitimacy of such an approach. In this respect, we have very briefly discussed a few issues concerning the comparability of corpora to be merged. All in all, we are confident that the methodological issues and empirical findings discussed here will contribute to raising methodological awareness within usage-based theorizing and help to inspire a lively debate on these issues.

References

Goldberg, A.E. 1995. *Constructions: A Construction-Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A.E., D.M. Casenhiser, & N. Sethuraman. 2004. Learning Argument Structure Generalizations. *Cognitive Linguistics* 15: 286-316.

Gries, St.Th. 2003. Towards a Corpus-based Identification of Prototypical Instances of Constructions. *Annual Review of Cognitive Linguistics* 1: 1-27.

Gries, St.Th. 2004. *Coll.analysis 3*. A program for R for Windows.

Gries, St.Th., B. Hampe, & D. Schönefeld. 2005. Converging Evidence: Bringing Together Experimental and Corpus Data on the Association of Verbs and Constructions. *Cognitive Linguistics* 16: 635-676.

Hare, M.L., K. McRae, & J.L. Elman. 2003. Sense and Structure: Meaning as a Determinant of Verb Subcategorization Preferences. *Journal of Memory and Language* 48: 281-303.

Howes, D.H., & R.L. Solomon. 1951. Visual Duration Threshold as a Function of Word Probability. *Journal of Experimental Psychology* 41: 401-410.

Johanson, S. & K. Hofland. 1989. *Frequency Analysis of English Vocabulary and Grammar, Based on the LOB Corpus*. Oxford: Clarendon.

Just, M., P. Carpenter, & J. Woolley. 1982. Paradigms and Processes in Reading Comprehension. *Journal of Experimental Psychology: General* 111: 228-238.

Kilgarriff, A. 2001. Comparing Corpora. *International Journal of Corpus Linguistics* 6: 1-37.

Kilgarriff, A. & T. Rose. 1998. Measures for Corpus Similarity and Homogeneity. *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, 46-52.

Manning, C.D. & H. Schütze. 2000. *Foundations of Statistical Natural Language Processing*. 4th printing with corrections. Cambridge, MA: MIT Press.

Nordquist, D. 2004. Comparing Elicited Data and Corpora. *Language, Culture, and Mind*, eds. M. Achard & S. Kemmer, 211-224. Stanford: CSLI Publications.

Quirk, R., S. Greenbaum, G. Leech, & J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rayson, P. & R. Garside. 2001. Comparing Corpora Using Frequency Profiling. *Proceedings of the Workshop on Comparing Corpora*, held in conjunction with *The 38th Annual Meeting of the Association for Computational Linguistics*, 1-6.

Schönefeld, D. 1999. Corpus Linguistics and Cognitivism. *International Journal of Corpus Linguistics* 4: 137-171.

Schönefeld, D. 2001. *Where Lexicon and Syntax Meet*. Berlin: Mouton de Gruyter.

Stefanowitsch, A., & St.Th. Gries. 2003. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics* 8: 209-243.