

# In defense of corpus-based methods: A behavioral profile analysis of polysemous *get* in English

**Andrea L. Berez and Stefan Th. Gries**

University of California, Santa Barbara

*aberez@umail.ucsb.edu stgries@linguistics.ucsb.edu*

## 1 Introduction

Linguistics is undergoing a major change these days. There is an increasing emphasis on strengthening the methodological foundations of linguistics, which manifests itself in (i) a growing body of work refining and developing both experimental and quantitative/statistical methods, and (ii) an increasing number of studies which compare data from methodologically different sources of evidence. One relatively new field in which this discussion has been particularly lively recently is usage-based cognitive linguistics in general and cognitive semantics in particular. Early studies investigating polysemous elements – morphemes, words, syntactic constructions – advocated semantic network analysis, but such networks have been criticized for their uncertain ontological status as well as for their somewhat subjective character. More recently, cognitive semanticists have taken the usage-based commitment more seriously and have put forward semantic analysis on the basis of corpus data; cf. Gries and Stefanowitsch (2006) and Tummers, Heylen, and Geeraerts (2005) for (discussion of) representative studies. However, corpus data are not immune to critique: Raukko (1999, 2003), while skeptical of intuition-based semantic analyses, criticizes corpus-based studies of polysemy because the analyst relies on his/her linguistic introspection when analyzing and classifying the instances of a word in the texts (Raukko 1999:87); he strongly argues for an alternative experimental method, seemingly to the exclusion of other methods.

In this paper, we pursue two goals. First, we briefly characterize and then challenge Raukko's assessment of corpus-linguistic methods. Second, we apply a new method of corpus-driven semantic analysis, called *behavioral profile analysis* (cf. Divjak 2006, Divjak and Gries 2006, Gries 2006a) to Raukko's object of study, the verb *to get*. The previous studies mentioned above have shown that the behavioral profile approach can be particularly useful to resolve some problems of (especially cognitive) semantic analyses, such as the number of senses to assume and the assessment of which senses are most similar to one another. However, given the recency of this method, the number of studies that investigate highly polysemous items is still limited. We therefore apply this method to the verb *to get* to illustrate that not only does it not suffer from the problems of the intersubjective approach, but it also allows for a more bottom-up/data-driven analysis of the semantics of lexical elements to determine how many senses of a word to assume and what their similarities and differences are.

In the final section of this paper we address some more broadly applied criticisms of corpus linguistics that we feel are a result of a lack of information in the field about the nature and purpose of corpus-based research.

## 2 Points of critique against corpus-linguistic approaches I

This section is concerned with points of critique of a corpus-based approach from within the framework of cognitive linguistics in general and cognitive semantics in particular. We take as representative of the corpus-critical point of view those presented in Raukko (1999, 2003), in which he discusses his 1994 study on the polysemy of the English lexeme *get*. He supplied 329 high school students with a questionnaire in which subjects are asked to provide, among other things, example sentences and introspective comments on the senses of *get*. He then used what he refers to as an *intersubjective method* to determine, among other things, the number and kinds of senses of *get* and *get*'s prototypical sense. His resulting 1999 article is a discussion of the perceived advantages of his intersubjective method over other ways of investigating polysemy and of the results of the survey. The 2003 article is an attempt to show how

the results of his experiment support the notion of polysemy as a flexible mass-like entity, rather than a more traditional network of connected sense nodes.

Surprisingly, Raukko seems to have an additional purpose beyond the investigation of the polysemy of *get*. In what we take to be an effort to promote his questionnaire-based methodology to the exclusion of all others, he marginalizes other approaches by distorting aspects of corpus-based research methods. Later in this article we discuss the results of our own corpus-based study of the polysemy of *get* and compare them to Raukko's results, but before we turn to that, let us look at just two of Raukko's fundamentally mistaken generalizations about corpus-based methods as compared to his own (cf. Berez 2007 for a full critique).

### 2.1 *Misconceptions about the goals of corpus linguistics*

One of Raukko's misconceptions is his characterization of the goals of corpus linguistics:

The linguist looks at a large and somewhat pre-processed selection of text material and tries to find the relevant instances (instantiations, specimens) of the item that s/he wants to study. (Raukko 2003:165)

This statement strikes us as either a straw man, a severe misunderstanding or just as severe a misrepresentation. It is a straw man in the sense that, sure, if a corpus linguist is investigating *get* in a corpus, she only looks for “relevant instances”, i.e. instances of the verb (lemma) *get*, and not for some other item, like the noun *formaldehyde*. It is a severe misunderstanding or misrepresentation to think that a corpus linguist worthy of the name would look for instances of *get* in a corpus, but only classify as relevant those that fit her theory in order to avoid dealing with problematic counterexamples. Instead, the honorable corpus linguist reports on (a hopefully representative or randomized sample of) all instances of the form under investigation. As a matter of fact, it is a particular strength of the corpus-based approach – to which everybody who has ever been surprised by naturally-occurring linguistic data can testify – that a comprehensive corpus search typically results in data that introspection alone could not have yielded, and that all of these data are taken into account.

### 2.2 *The role of introspection*

Raukko (1999:87) likewise takes issue with the fact that corpus linguists use some degree of introspection in their analysis of corpus data:

Other types of recent analyses of lexical polysemy [...] have made use of language corpora as sources of real-life data, but here also the analyst basically relies on her/his own linguistic introspection when analyzing the instances of a word in the texts and classifying them into neat semantic categories.

Just like the previous one, this statement again is either a redundant truism or a severe misunderstanding or misrepresentation. Of course, the analysis of corpus data requires classificatory decisions which are not always entirely objective – no corpus linguist would deny this fact, just as no scientist would deny that some degree of intuition plays a role in nearly *any* study. As a matter of fact, in an amazingly self-contradictory way, Raukko's method relies more on introspection than most other supposedly empirical semantic studies we are aware of: not only does Raukko have to use his own introspection in making sense of his subjects' responses, even more curiously, Raukko's method appears to simply place a large part of the analytical burden on linguistically naïve subjects since, for instance, he simply asks his subjects for what they consider the prototypical sense (cf. Raukko 1999:91).

Thus, while we openly admit that a completely objective classification of corpus data (or most other kinds of linguistic data, for that matter) is extremely unlikely, the advantages of corpus data are that

- as mentioned above, the richness and diversity of naturally-occurring data often forces the researcher to take a broader range of facts into consideration;
- the corpus output from a particular search expression together constitute an objective database of a kind that made-up sentences or judgments often do not. More pointedly, made-up sentences or

introspective judgments involve potentially non-objective (i) data gathering, (ii) classification, and (iii) interpretive processes on the part of the researcher. Corpus data, on the other hand, at least allow for an objective and replicable data-gathering process; given replicable retrieval operation, the nature, scope, and the ideas underlying the classification of examples can be made very explicit – certainly more explicit than some coding procedures in Raukko's two studies.

In the following section, we will discuss a very brief case study of Raukko's own topic – *get* – from a recently developed corpus-based approach.

### 3 A case study: *get* in English

In the face of the above misrepresentations, we felt tempted to provide an outline of a recently developed corpus-based method that is immune to many of Raukko's points of critique and that has, as we will briefly mention below, received first experimental confirmations. We conducted our own study of the polysemy of *get* using a quantitative corpus method known as *behavioral profiling* (henceforth BP; see Gries and Divjak, to appear, for a summary) described below. We show that some of our results are in fact remarkably close to Raukko's, but also provide an illustration of how BPs can combine syntactic and semantic information in a multifactorial way that is hard to come by using the kinds of production experiments Raukko discusses.

While there are now several published BP studies in the domains of (near) synonymy and antonymy (cf. Divjak 2006, Divjak and Gries 2006, to appear a, b, submitted, Arppe and Järviö 2007, and Schmid 1993 for a slightly similar early forerunner), there is so far less work on polysemy (with the exception of Gries 2006a). Thus our study also tests the discriminatory power of the BP approach for the domain of polysemy. In the next three subsections, we discuss the general principles of the BP method, the data from our small study, and our results. At the end of Section 3, we briefly compare some of our results to those of Raukko (1999, 2003).

#### 3.1 The BP method

As a corpus-based approach, the BP approach is based on the truism that corpus data provide (nothing but) distributional frequencies. A more relevant assumption, however, is that distributional similarity reflects, or is indicative of, functional similarity; our understanding of functional similarity is rather broad, i.e., encompassing any function of a particular expression, ranging from syntactic over semantic to discourse-pragmatic. The BP method involves the following four steps:

- the retrieval of (a representative random sample of) all instances of a word's lemma from a corpus in their context (usually at least the complete utterance/sentence);
- a (so far largely) semi-manual analysis of many properties of the use of the word forms; these properties are, following Atkins (1987), referred to as ID tags and comprise
  - morphological characteristics of the usage of the word in question: tense, aspect, mood, voice, number marking, etc.;
  - syntactic characteristics of the usage of the word in question: use in main or subordinate clauses, sentence type;
  - semantic characteristics: the sense of the word, semantic roles of the word's arguments and adjuncts;
- the generation of a co-occurrence table that specifies which ID tag level is attested how often in percent with each word (of a set of near synonyms or antonyms) or sense (of a polysemous word; the columns containing the percentages for each word or sense are then referred to as the word's or sense's behavioral profile (borrowing a narrower term coined in Hanks 1996).
- the evaluation of the table by means of descriptive techniques (such as summary frequencies), correlational methods, and/or exploratory cluster analysis.

3.2 *The data*

The current study is based on a dataset consisting of 600 instances of *get* in all its inflectional forms that were randomly selected from the 3,668 total instances found in the British component of the International Corpus of English, ICE-GB. Randomization was based on proportions of each inflectional form found in the full concordance. Senses were categorized manually using WordNet 2.1 and the Oxford English Dictionary Online as a rough guide to sense distinctions. Table 1 contains the 47 senses found in the data with higher-level sense groupings and examples in the right column (The classification of senses is less uncontroversial than Table 1 may suggest since it is often difficult to decide (i) whether or not to distinguish two senses and (ii) on which level of granularity to distinguish senses; the classification below is our best guess but other classifications are certainly conceivable.)

Table 1: Senses of *get* in our study

Sense / sense group	Example
Acquire concrete metaphorical for another non-agentively contract illness/injury  hit/capture target understand	Get some jellytots or something like that I didn't expect to get that sort of reaction Uh let me get you a photograph People get the wrong injections have the wrong leg amputated And I said oh dear Harriet thinking oh you know she 's got the flu or something We got that one (hardened shelter in Iraq) [Y]ou got it (punchline of a joke)
Stable possession concrete metaphorical have plan existential	I've got a little sheet of paper somewhere and I've also got my diary I've got a sister I've got the department dinner on the Friday night [...] There was that air of expectancy about the place you get [...]
Movement in specified direction concrete metaphorical cause to move  cause to metaphorically move rise bodily rise bodily metaphorically  rise from sleep support metaphorically  act without retribution  cause to be sad board transportation  board transportation metaphorically  commence action cause to commence action cause to commence action metaphorically dispose of something dispose of something metaphorically	[...] what my emotions will be telling me when I get back home He got really into Jack Kerouac so I gave him a Jack Kerouac book [...] they were doing all they could to get their employees out of the country [...] his message is: you got them into trouble, now get them out Amy cast down her napkin upon the table-cloth [...] and got up would it not be easier for him to [...] ask him to get off his butt and do something Getting up each morning in sub-zero temperatures [...] [...] the Labour Party view would be [...] to get solidly behind a UN policy [I]t's worth whatever the government can get away with charging for it This is getting me down [...] you have to hang around HMV and then walk over or get the tube there The entire mass of birds at once got on wing and flew seawards  [I]t's just a matter of getting round to it I suppose And the steel weight ... was very difficult to get it turning On Monday morning get that brain going  all my assorted junk [...] some of it I am anxious to get rid of [...] BR's policy is to get rid of the twilight atmosphere of the old stations and trains

Sense / sense group	Example
Enter state enter state  cause to enter state become acquainted become acquainted metaphorically form romantic couple rendezvous cause to rendezvous establish communication enter into established social network cause to enter into established social network be friendly  do in specified manner	You must be very careful with that cos otherwise you're going to get confused And unless we can get our transport infrastructure into [...] shape Yes yes they're extremely friendly when you get to know them It's time I got to know the sun D' you D' you really think we could ever get back together again Shall we try and get together sometime Trying to get a band together I've been trying to get in touch for months And why do they then want us to get into a political union  As your current certificate doesn't expire until June I will try to get you onto another refresher course before then I mean Father and Mother ... they still don't get on very well anyhow the highest priority for the Government is to get the economy right
Complete	[W]e have got through it haven't we
Be permitted	Criminals prefer anonymity and are less likely to get to work where there is a chance of being recognized
Cope	Europe could frankly get along without us perfectly happily
Must	if you send that to him he 's got to address the issue this time certainly
Passive Passive Cause passive	if you try to be supportive of people you so often get taken advantage of? of the English creditors who [...] were] still trying to get their money paid

Each data point was coded for 55 morphological, syntactic and semantic ID tags (cf. Table 2).

Table 2: ID tags used in our study

(Kind of) ID tag	Levels of ID tag
morphological verb form of <i>get</i> voice of <i>get</i> verb form of main verb	infinitive, pres (base or 3rd), present progressive, past tense, past participle active, passive infinitive, pres (base or 3rd), present progressive, past tense, past participle
syntactic transitivity of <i>get</i> clause function  clause type dep. clause type transitivity of clause	copula, complex trans., ditrans., intrans., monotrans., pass, prop, semi, trans. A, AJPO, CF, CJ, CS, CT, DEFUNC, ELE, FNPPPO, NOOD, NOSU, NPPO, OD, PARA, PC, PU, SU main, depend indrel, rel, zrel, sub, zsub copula, complex trans., ditrans., intrans., monotrans., trans.
abstractness of sense	abstract vs. concrete

26 senses occurring more than four times were analyzed quantitatively. The resulting co-occurrence table was entered into a hierarchical cluster analysis, the results of which are discussed in the following section.

### 3.3 Results

Even though corpus data are by definition rather noisy and the present data set is certainly not particularly large, the cluster analysis yields a dendrogram with a good deal of structure; cf. Figure 1. Moreover, several of these clusters are fairly straightforward to interpret. For example, there is

- a cluster with most possession senses (and one other): 'possess', 'possess/acquire', 'possess metaphorical' plus 'contract illness', which can be metaphorically understood as involving possession;
- a cluster with most acquisition senses: 'acquire for another', 'acquire non-agency', 'acquire', 'acquire metaphorical';
- a cluster with most movement senses (and others): 'rise bodily', 'rise from sleep', 'move in specified direction', 'move in specified direction (metaphorical)';
- a cluster containing the grammaticalized senses *must* and *passive*.

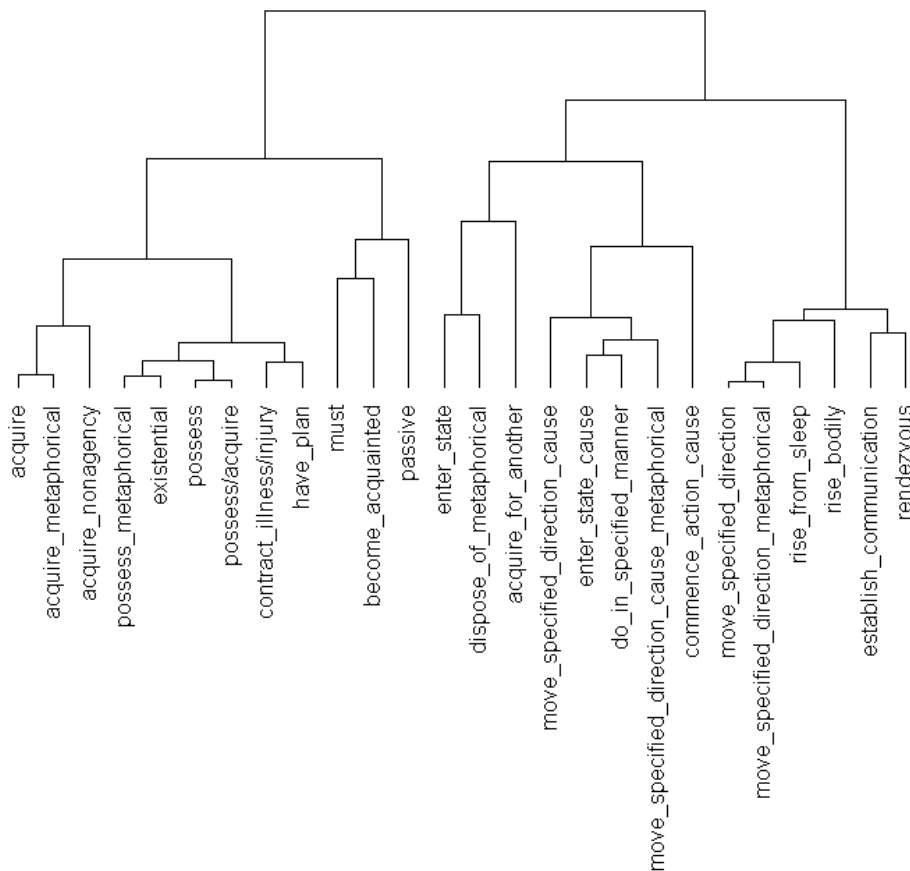


Figure 1: Result of a HAC on the BP of *to get*

In addition to this somewhat intuitive interpretation, we also ran a hierarchical agglomerative cluster analysis on the data and calculate  $p$ -values based on multiscale bootstrap resampling (cf. Shimodaira 2004, Suzuki and Shimodaira 2006). Again, in spite of the small sample size, we find that, as represented in Figure 2.

- the 'possess' cluster, the 'acquire' cluster, and the cluster with the grammaticalized senses reach marginal significance ( $p \approx 0.07$ ,  $p \approx 0.1$  and  $p \approx 0.08$ );

- the non-causative 'move' cluster reaches significance ( $p \approx 0.03$  \*);
- a cluster that contains all causative senses (but also two other senses;  $p \approx 0.21$  ns).

It is especially interesting that these clusters emerge at least relatively clearly because it seems what underlies the clusters are semantic aspects, but the number of semantic criteria – i.e., ID tags – that were coded is in fact negligible.

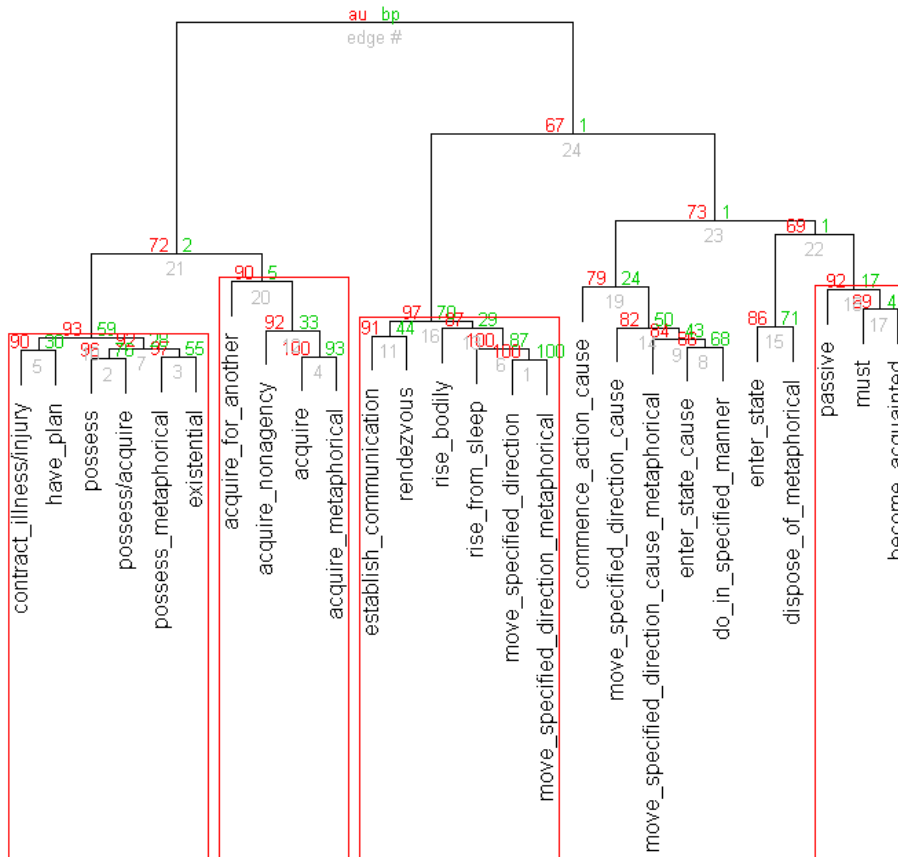


Figure 2: Result of a HAC on the BP of *to get* with multiscale bootstrap resampling

### 3.4 How our results compare to Raukko

So how do our results compare to those found by Raukko in his survey? He doesn't provide much in terms of descriptive statistics, but he does provide some frequency percentages for major sense categories. We can compare these frequencies with our own to make some initial observations about the relationship between introspective evidence and corpus evidence.

Table 3 shows the relative frequencies of three groups of senses. Our categorization of the senses of *get* were different from those Raukko found, but nonetheless we were able to create similar groupings.

Table 3: Comparison of occurrence frequencies

Results from Raukko (1999)		Results from corpus	
'Obtaining' + 'receiving' + 'stable possession'	43.27%	'Acquire' (all types) + 'possession' (all types)	42.00%
'Change of state'	21.15%	'Enter state' (all types)	9.83%
'Motion'	18.37%	'Move in specified direction' (all types)	16.83%

Note that two of the major categories, the 'obtain/acquire/possess' category and the 'motion/movement' category, have remarkably similar frequencies. The 'change of state' category, however, was quite different between the two studies: in Raukko, this is the third most frequently given meaning type, but it only accounted for less than ten percent of our dataset. Not included in the chart is the sense 'understand'. Raukko's informants produced examples of this sense 103 times (nearly 5% of all examples), but this sense showed up in our data only once (<1%). While we cannot make firm statements about how introspections about sense usage are borne out in actual usage based on these results (we need to consider dialect and age differences between the two populations), the data here suggest that the results of a BP approach, while coming close to the results of an experimental approach, are not the same as the latter.

#### 4 Points of critique against corpus-linguistic approaches II

Apart from Raukko's points, there are some other concerns all too often raised with *individual* corpus-based studies, both outside of and within cognitive linguistics. These can be summarized in two frequent reactions to corpus-linguistic presentations (cf. Gries and Divjak, submitted, for more discussion):

- comments aimed at the corpus as a whole: “but isn't all this true in your corpus only?” or “I bet you would find something entirely different if you looked at a different corpus!” and “but the two corpora you are comparing are not sufficiently similar, your results are invalid!”;
- comments aimed at subpart of the corpus: “I bet you would find something different if you looked at different registers!” or “I'm sure you would find something different if you looked at word forms/lemmas instead of lemmas/word forms.”

In spite of their frequency, these comments are weak on two counts. First, they are procedurally problematic: The 'asker' hypothesizes a deviation from the null hypothesis (that there is no effect of or distributional difference between corpora), i.e., an alternative hypothesis, yet places the burden of proof on the 'askee'. If the asker thinks the distributional data obtained and reported on would be different in another corpus, the asker should test this alternative hypothesis instead of stipulating a difference for which (so far) no evidence exists.

Second, assertions like these are empirically problematic: The kinds of difference often hypothesized by askers is usually far from 'a given'. There is now increasing evidence that simple generalizations of what does and what does not remain constant across corpora, registers, word forms etc. are often inaccurate or exaggerated. Some of this evidence is based on BP approaches, other evidence is based on data regarding the distribution of occurrences of syntactic variables or the distribution of co-occurrences of lexico-syntactic variables.

##### 4.1 Comments about a corpus (as a whole)

As for the comments aimed at the use of a particular corpus, for example, the results obtained by Schmid (1993), who worked with the LOB corpus, are – while less comprehensive in terms of annotation and more comprehensive in terms of sense differentiation – to a considerable degree compatible with Divjak and Gries's (to appear) results. This is noteworthy because the composition of the two corpora are of such a different nature that might compel many an audience to doubt the corpus comparability: Schmid's (1993) LOB consists exclusively of written and published texts representative for British English of the 1960s, whereas approximately 60% of the ICE-GB corpus used in Divjak and Gries (to appear) consists of spoken language and even the 40% of written language in the ICE-GB contains a sizable amount of unpublished material.



#### 4.2 *Corpus parts, registers, genres, etc.*

Similar findings have been reported for the cherished distinction between spoken and written data. Stefanowitsch and Gries (to appear) and Gries (to appear) show that distinguishing between spoken and written data has no substantial effect in analyses of lexico-syntactic preferences of active vs. passive voice, the two word orders of verb-particle constructions, and the *will* vs. *going-to* future. Gries (to appear) shows that the same holds true for the ditransitive vs. prepositional dative alternation and that the 'real' division of the corpus – 'real' in the sense of explaining the maximally meaningful amount of variance in the corpus data as obtained by a principal component analysis – cuts across both spoken vs. written and all register distinctions present in the corpus. More specifically, the four corpus parts that are most homogeneous internally and most different from each other are based neither only on spoken vs. written nor only on subregisters; instead, they are mixed groups based on both these levels of granularity (cf. Gries, to appear, for details). Gries (to appear) also finds that looking at word forms does not necessarily yield results different from a lemma-based analysis.

More generally, Gries (2006b) demonstrates on the basis of three very different case studies – the frequencies of the present perfect, the predictability of particle placement, and lexicosyntactic associations of the ditransitive constructions – that the usual suspects of mode, register and even subregister account for much less variability than the above-mentioned after-presentation comments suggest. In each of the above cases, different samples from even a single corpus may yield very different results; the size of within-corpus differences is often similar in size to between-corpus differences so there is little reason to *a priori* assume that other corpora will automatically yield different results. Bottom line: the issue of corpus homogeneity and comparability can only be determined (i) empirically and (ii) individually for each phenomenon, each corpus, and each level of corpus division(s) – it cannot be determined or objected *a priori* as one sees fit.

## 5 Conclusion

As the field of linguistics increasingly turns to usage-based and quantitative methods, corpora can supply supporting evidence for questions answered with other methods and go beyond them in terms of both description and explanation. Here we have shown how a rejection of corpus-based investigations of polysemy is premature: our BP approach to *get* not only avoids the pitfalls Raukko mistakenly claims to be inherent in corpus research, it also provides results that are surprisingly similar to his own questionnaire-based results, and Divjak and Gries (to appear b) show how predictions following from a BP study are strongly supported in two different psycholinguistic experiments. It is therefore our hope that in addressing the widely held misunderstandings that Raukko but serves to exemplify, we can encourage the use of corpus-based methods in linguistics more widely.

## References

- Arppe, Antti and Juhani Järviö 2007. Take empiricism seriously: combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3.2:131-59.
- Atkins, Beryl T. Sue. 1987. Semantic ID tags: Corpus evidence for dictionary senses. *Proceedings of the Third Annual Conference of the UW Centre for the New Oxford English Dictionary*, p. 17-36.
- Berez, Andrea L. 2007. The polysemy of English *get*: Past and current approaches. Ms., University of California, Santa Barbara.
- Divjak, Dagmar S. 2006. Ways of intending: Delineating and structuring near synonyms. In Gries and Stefanowitsch (2006), p. 19-56.
- Divjak, Dagmar S. and Stefan Th. Gries. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2.1:23-60.
- Divjak, Dagmar S. and Stefan Th. Gries. to appear a. Corpus-based cognitive semantics: A contrastive study of phasal verbs in English and Russian. In: Dziwirek, Katarzyna and Barbara Lewandowska-Tomaszczyk (eds.). *Studies in cognitive corpus linguistics*. Frankfurt am Main: Peter Lang.
- Divjak, Dagmar S. and Stefan Th. Gries. to appear b. Clusters in the mind? Converging evidence from near-synonymy in Russian. *The Mental Lexicon*.

- Gries, Stefan Th. 2006a. Corpus-based methods and cognitive semantics: the many meanings of *to run*. In Gries and Stefanowitsch (2006), p. 57-99.
- Gries, Stefan Th. 2006b. Exploring variability within and between corpora: some methodological considerations. *Corpora* 1.2:109-51.
- Gries, Stefan Th. to appear. Corpus data in usage-based linguistics: What's the right degree of granularity for the analysis of argument structure constructions?. In: Brda, Mario, Milena Žic Fuchs, and Stefan Th. Gries (eds.). *Expanding cognitive linguistic horizons*. Amsterdam, Philadelphia: John Benjamins.
- Gries, Stefan Th. and Dagmar S. Divjak. to appear. Behavioral profiles: a corpus-based approach to cognitive semantic analysis. In: Evans, Vyvyan and Stephanie S. Pourcel (eds.). *New directions in cognitive linguistics*. Amsterdam, Philadelphia: John Benjamins.
- Gries, Stefan Th. and Dagmar S. Divjak. submitted. Quantitative approaches in usage-based cognitive semantics: myths, erroneous assumptions, and a proposal.
- Gries, Stefan Th. and Anatol Stefanowitsch (eds.). 2006. *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin, New York: Mouton de Gruyter.
- Hanks, Patrick. 1996. Contextual dependency and lexical sets. *International Journal of Corpus Linguistics* 1.1:75-98.
- Oxford English Dictionary Online. 2006. *Get*, *v*. Retrieved 21 December 2006 from <http://dictionary.oed.com>.
- Raukko, Jarno. 1999. An "intersubjective" method for cognitive-semantic research on polysemy: the case of *get*. In: Hiraga, Masako K., Chris Sinha, and Sherman Wilcox (eds.). *Cultural, psychological and typological issues in cognitive linguistics. selected papers of the bi-annual ICLA meeting in Albuquerque, July 1995*. Amsterdam, Philadelphia: John Benjamins, p. 87-105.
- Raukko, Jarno. 2003. Polysemy as flexible meaning: experiments with English *get* and Finnish *pitää*. In: Nerlich, Brigitte, Zazie Todd, Vimala Herman, and David D. Clarke (eds.). *Polysemy: flexible patterns of meaning in mind and language*. Berlin, New York: Mouton de Gruyter, p. 161-93.
- Schmid, Hans-Jörg. 1993. *Cottage and co., idea, start vs. begin*. Tübingen: Max Niemeyer.
- Shimodeira, Hidetoshi. 2004. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Annals of Statistics* 32:2616-2641.
- Stefanowitsch, Anatol and Stefan Th. Gries. to appear. Register and constructional semantics: A constructional case study. In: Kristiansen, Gitte and René Dirven (eds.). *Cognitive sociolinguistics*. Berlin, New York: Mouton de Gruyter.
- Suzuki, Ryota and Hidetoshi Shimodeira. 2006. pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22.12.
- Tummers, Jose, Kris Heylen, and Dirk Geeraerts. 2005. Usage-based approaches in Cognitive Linguistics: a technical state of the art. *Corpus Linguistics and Linguistic Theory* 1.2:225-61.
- WordNet, n.d. WordNet Serach 2.1: "get". Retrieved 23 December 2006 from <http://wordnet.princeton.edu>.