Behavioral profiles: A corpus-based perspective on synonymy and antonymy*

Stefan Th. Gries, University of California, Santa Barbara
Naoki Otani, Saitama University

1 Introduction

1.1 Two empirical perspectives in the study of synonymy and antonymy

The domain of linguistics that has arguably been studied most from a corpus-linguistic perspective is lexical, or even lexicographical, semantics. Already the early work of pioneers such as Firth and Sinclair has paved the way for the study of lexical items, their distribution, and what their distribution reveals about their semantics and pragmatics / discourse function(s). A particularly fruitful area has been the study of (near) synonyms;[1] probably every corpus linguist has come across the specific example of *strong* and *powerful* – the fact that one would say *strong tea* but not *powerful tea* – as well as the general approach of studying synonyms on the basis of their distributional characteristics. However, while synonymy is probably the most frequently corpus-linguistically studied lexical relation, there is by now also quite some corpus-based work on its counterpart relation, antonymy, and we will briefly discuss examples of both below.

The study of the semantics of lexical items quite obviously presupposes two central concepts. First, one requires a notion of what it means to know a word, and for reasons that we will elaborate on more below, we follow here an approach by Miller and Charles (1991), who proposed the notion of a *contextual representation*, which is:

- "knowledge of how that word is used" (p. 4; cf. also Miller 1999);
- "some abstraction or generalisation derived from the contexts that have been encountered" (p. 5);
- "a mental representation of the contexts in which the word occurs, a representation that includes all of the syntactic, semantic, pragmatic, and stylistic information required to use the word appropriately." (p. 26).

Second, one requires some kind of scale or, more likely, multidimensional space of semantic similarity along/within which words or, more precisely, concepts or the contextual representations of words can be placed, compared, and ordered such that, on this scale / within this space, *strong* would be closer to *powerful* than it would be to *weak*. Unsurprisingly, given our assumption of a contextual representation, we consider semantic similarity a function of the contexts in which words are used (cf. Miller and Charles 1991:3), from which it somewhat naturally follows that "[t]he similarity of the contextual representation of two words contributes to the semantic similarity of those words," which Miller and Charles (1991:9) referred to as the *weak contextual hypothesis*. However, this approach only shifts the burden of difficulty from the question of how to measure the semantic similarity of, say, two words to the question of how to measure the similarity of contextual representations of two words.

The discussion of this problem in general, but also its more specific application to the issue of synonymy and antonymy has mostly contrasted two different perspectives: the *co-occurrence approach* (cf. e.g. Rubenstein and Goodenough 1965) and the *substitutability approach* (cf. e.g. Deese 1962, 1964), and we need to discuss both of them briefly.[2]

1.1.1 The co-occurrence approach

The co-occurrence approach is based on assumptions, or even axioms, that are very dear to probably nearly all corpus linguists. Corpus linguistics is inherently a distributional discipline and the study of lexical semantics with corpora is no exception: corpora do not provide meanings or functions that can be readily extracted and compared, but only the distributions of formal elements – morphosyntactic and lexical (and, depending on the corpus, sometimes phonological or orthographic) elements – so meanings and functions must be inferred from the distribution(s) of formal elements within their contexts.

The assumption underlying the co-occurrence approach is that the distributional characteristics of the use of an item reveals many of its semantic and functional properties and purposes, an assumption that has been made in various different sources: Firth (1957:11) famously stated "[y]ou shall know a word by the *company* it keeps" ; just as famously, Bolinger (1968:127) stated that "a difference in syntactic form always spells a difference in meaning"; Harris (1970:785f.), while much less quoted in this connection, asserted this even more explicitly:

> [i]f we consider words or morphemes *A* and *B* to be more different in meaning than *A* and *C*, then we will often find that the distributions of *A* and *B* are more different than the distributions of *A* and *C*. In other words, difference of meaning correlates with difference of distribution.

More recently, Cruse (1986:1) stated, "the semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts", and with a more syntactic focus, Hanks (1996:77) wrote "the semantics of a verb are determined by the totality of its complementation patterns."

This kind of logic has been applied especially fruitfully in the domain of synonymy, where contextual information of two kinds has been particularly useful and revealing:

- collocational information: what are the words that are modified by *strong* and *powerful* (Church *et al*. 1991), by *absolutely*, *completely*, and *entirely* (cf. Partington 1998: Section 3.6, also cf. Ch. 2), by *big*, *large*, and *great* (cf. Biber *et al*. 1998: Section 2.6), or by *alphabetic* and *alphabetical* and many other *-ic*/*-ical* adjective pairs (cf. Gries 2003);
- syntactic information: what are the preferred grammatical associations of *quake* and *quiver* (cf. Atkins and Levin 1995), of *little* vs. *small* or *begin* vs. *start* (cf. Biber *et al*. 1998: Sections 4.2 and 4.3 respectively), of causative *get* and *have* (cf. Gilquin 2003), of Mandarin *lian* … constructions (cf. Wang 2006), of several Finnish verbs meaning 'think' (cf. Arppe and Järvikivi 2007 and Arppe 2008), etc.

An experimental study that is often mentioned in connection with synonymy/antonymy is that of Rubenstein and Goodenough (1965), who had subjects generate sentences involving 130 target words from 65 word pairs and compared the collocational overlap of sentences created for words of the 65 pairs using an intersection coefficient, obtaining results that are compatible with a contextual approach, at least for highly synonymous words.

Especially with regard to the study of antonymy, a more specialized kind of co-occurrence approach has also been promoted. This more specialized approach, when applied to two antonymous words *x* and *y*, does not include and compare all collocates of *x* and *y* – it involves counting how often *x* and *y* themselves co-occur within the same sentence, which in turn by definition increases the amount of collocational overlap that is at the heart of the co-occurrence approach. For example, Charles and Miller (1989) reported general sentential co-occurrence counts of *big*, *large*, *little*, and *small* in the Brown corpus, which are larger than would be expected by chance. In addition, Justeson and Katz (1991) found that antonymous words also prefer to occur in substitution patterns or contrastive parallel phrases (cf. Fellbaum 1995 for similar findings and support in favor of a sentential co-occurrence approach). As a last example, in a very interesting study Jones *et al*. (2007) even used such kinds of contrastive constructions to successfully identify what has been referred to as *direct antonyms* (cf. Gross, Fischer and Miller 1989) or *canonical antonyms*, namely pairs of antonyms that are strongly associatively paired (e.g. *wet* vs. *dry*) whereas their (near) synonyms are not (e.g. *moist* vs. *dry*); the contrastive constructions used this way x *and* y *alike*, *between x and y*, *both x and* y, *either x or y*, *from x to y*, x *vs. y* and *whether x or y*.

1.1.2 The substitutability approach

The substitutability approach has been most strongly advocated by Charles and Miller in several different papers. It can be summarized as follows:

> (1) collect a set of sentences using item *A*; (2) collect a set of sentences using item *B*; then (3) delete *A* and *B*, shuffle the resulting contexts; and (4) challenge subjects to sort out which is which. The more contexts there are that will take either item, the more similar the two sets of contexts are judged to be. (Miller and Charles 1991:11)

(The sentences mentioned in (1) and (2) were either taken from a corpus or generated by subjects in a pilot study.) Charles and Miller (1989) and Miller and Charles (1991) use this measure of contextual similarity and find that it is strongly correlated with experimentally-obtained semantic similarity ratings and conclude that it appears "that a measure of contextual similarity based on substitutability gives better predictions than did Rubenstein and Goodenough's (1965) measure of contextual similarity based on co-occurrence" (p. 17; cf. Charles 2000 for a similar conclusion). In addition, they criticize co-occurrence approaches for two shortcomings. First, they dismember the contexts they are supposed to represent, and while Miller and Charles do not explain what *exactly* they mean by that, it is reasonable to assume that they refer to the fact that a simple collocational count would disregard syntactic structures, which a sorting-contexts account would preserve. Second, they claim that the fact that co-occurrence accounts yield high similarity results for antonyms is problematic because "[a]ntonyms have contrasting meanings – not just zero similarity, but negative semantic similarity, if that is possible" and thus constitute counterexamples to a co-occurrence approach.[3]

1.2 Adjectives of SIZE: Some previous findings

One lexical field that has received a lot of attention both in general and in corpus semantics is that of SIZE, presumably because it includes two pairs of canonical antonyms and many studies devote at least some space to *big*, *little*, *large*, and *small*. In this section, we very briefly review some of the findings, "very briefly," because most of them lead to very similar conclusions.

In one of the earliest empirical studies, Deese (1964) used a word association method and finds what is intuitively obvious to every native speaker of English: *big* and *large* are very similar in meaning (their size ratings are 4.3 and 4.5 respectively, while *great* scores the slightly larger size rating of 5.1), as are *little* and *small* (their ratings are 1.8 and 1.9 respectively, while *tiny* is associated with even smaller sizes: its rating is 0.8). In addition to that, Deese also found that the by far most natural pairings of antonyms are *big* vs. *little* and *large* vs. *small*. He attributed these correlations to "partial contextual equivalences" (1964:356), a formulation which does not make it all that clear whether this would translate into a co-occurrence or substitutability account.

While Charles and Miller were in general in favor of a substitutability approach based on sorting sentence contexts as outlined above, their 1989 study shows that this approach does not explain particularly well how canonical antonym relations are established. Their alternative is essentially a co-occurrence approach again. They showed that *big* and *little* as well as *large* and *small* tend to occur in one and the same sentence in the Brown corpus with a probability that is much larger than chance would predict and concluded (1989:374) that "the 'clang' association between direct antonyms […] is a consequence of frequently perceiving and using these words together in the same syntactic structures." The above-mentioned study Justeson and Katz (1991:5) also found more sentences in which antonyms co-occurred than would have been expected by chance as well as the generally accepted association of *big* and *large* to *little* and *small* respectively.[4]

Finally, Jones *et al*. (2007) focused on antonym canonicity, but also discussed canonical-antonym relations among adjectives of size. Their findings are summarized in Figure 1 (taken from Jones *et al*. 2007:148), where the numbers and percentages indicate the number of frame types and the token percentages of joint occurrences in Google searches:

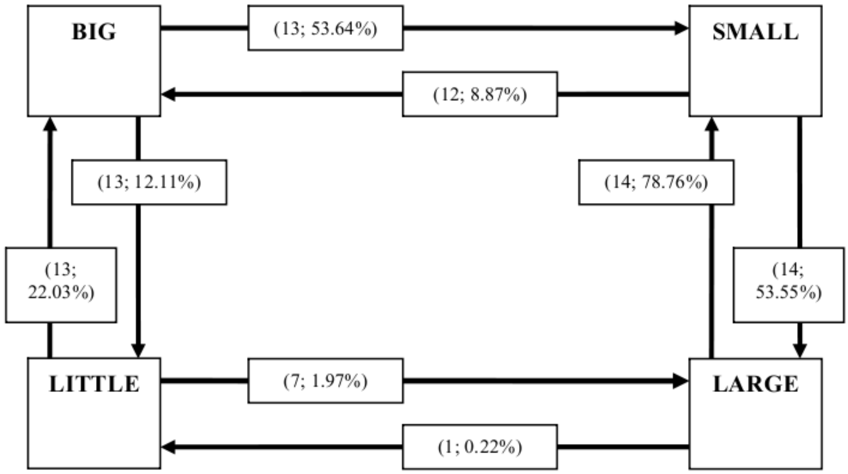

Figure 1: Summary of Jones et al. *(2007) findings regarding four adjectives of size*

With maybe one exception, the results are very encouraging: in line with nearly all previous work, *large* and *small* exhibit a strong and reciprocal attraction, and *little* prefers *big* as an antonym. The only drawback is that *big* does not prefer *little*.

1.3 The present study

The studies mentioned above have already gone very much beyond previous introspective accounts of the meanings and functions of lexical items. However, we think that some corpus-linguistic approaches still exhibit some areas of potential improvement and that the notion of contextual representation is worthy of more study.

First, there is a great degree of consensus that synonymy and antonymy both involve a large degree of semantic similarity and, thus, a high degree of common predictable distributional behavior. However, to our knowledge there is hardly any corpus-based work that explores synonyms *and* antonyms within a semantic field together; Charles and Miller (1989) and Jones *et al*. (2007) are laudable exceptions.

In addition to the above, second, it is probably fair to say that most corpus-based studies in the domains of synonymy and antonymy focus on individual pairs of synonyms and antonyms only and usually do not take larger sets of synonymous/antonymous words into consideration.

Third, many studies focus only on the base forms of the words in question as opposed to including, or differentiating between, different inflectional forms of the relevant lemmas.

Fourth, many corpus-linguistic studies of lexical relations until relatively recently focus only on collocational aspects or only on syntactic patterning, but do not combine lexical and syntactic distributional characteristics. Thus, although corpus data provide a wealth of distributional characteristics, many studies do not utilize most or even all of the available information. Some early studies that use more than just collocational or just syntactic information are cognitive-linguistic in nature. For instance, Schmid (1993) studied many lexical and syntactic characteristics of *begin* and *start* in an exemplary fashion. Also, in cognitive linguistics, Kishner and Gibbs (1996) studied collocations and syntactic patterns of *just*, and Gibbs and Matlock (2001) investigated uses of the verb *make*. Corpus-linguistic studies that are also appreciably broader in scope are Atkins (1987) study of *risk*, Hanks's (1996) study of *urge*, and Arppe and Järvikivi (2007), who all involved collocate and/or colligation analysis at an otherwise rare level of detail.

Finally, and in addition to the range of data included into an analysis, some of the studies also do not analyze their distributional data in the most revealing way but rather restrict themselves to observed frequencies of co-occurrence of a lexical item and a collocate or a syntactic pattern. Unfortunately, this is even true of some of the groundbreaking work of, for instance, Atkins (1987).

In this study, we intend to go at least in some ways beyond what some previous works have done. As for the first three points above, we do not just explore pairs of antonymous adjectives, but we also include synonymous adjectives in our data. More specifically, we explore the semantic field of SIZE by studying the distributional behavior of the adjective set {*big*, *great*, *large*} as well as the antonymous set {*little*, *small*, *tiny*}, including their comparative and superlative forms on the basis of data from the British Component of the International Corpus of English.

As for the fourth point, we do not restrict our attention to collocations or colligations only or a small set of linguistic features characterizing the use of each of the word forms included, but we use a relatively recent approach in corpus-based semantics, namely behavioral profiling. In addition, and as for the last point, our analysis is not just based on frequencies/percentages of co-occurrence, but on hierarchical agglomerative cluster analysis based on the similarity of co-occurrence vectors.

Before we proceed to the actual analysis, the latter two aspects must be clarified, so let us briefly explain the behavioral profile (BP) approach. (For reasons of space, we cannot discuss all the details of the BP approach here but need to refer to previous works. For applications of the BP approach to synonymy, cf. Divjak (2006) as well as Divjak and Gries (2006); for applications of the BP approach to polysemy, cf. Gries (2006) and Berez and Gries (2009); for experimental validation, cf. Divjak and Gries (2008); for an overview, cf. Gries and Divjak (2009).)

The application of the BP method to synonyms/antonyms involves the following steps:

i. the retrieval of (a representative random sample of) all instances of the lemmas of the synonyms/antonyms to be studied from a corpus in the form of a concordance;
ii. a (so far largely) manual analysis and annotation of many properties of each match in the concordance of the lemmas; these properties are, following Atkins (1987), referred to as ID tags and include morphological, syntactic, semantic, and collocational characteristics (cf. below for details and Table 1 for an excerpt of our data);
iii. the conversion of these data into a co-occurrence table that provides the relative frequency of co-occurrence of each lemma with each ID tag; the vector of these co-occurrence percentages for a lemma is called that lemma's behavioral profile (cf. below for details and Table 2 for an excerpt of our data);
iv. the evaluation of the table by means of exploratory and other statistical techniques, especially hierarchical agglomerative cluster analysis.

Before we begin to discuss the data, methods, and results, it is important to point out that it is quite difficult to anticipate what kinds of (meaningful?) results to expect. While it is clear that *big*, *large*, and *great* are synonyms, that *little*, *small*, and *tiny* are their synonymous antonyms, and that *big* and *little* as well as *large* and *small* are the canonical antonym pairs, it is not obvious, say, what kind of clustering to expect – a synonym-based clustering? an antonym-based clustering? a mixture of the two? Right now, we can therefore only state that clustering results which appear to order words randomly and/or do not find the canonical antonyms would undermine the BP approach, but more specific predictions are hard to come by. This is for several reasons: first, while there is already some BP-based work now on synonymous words (between and within languages) and polysemous words, there is as yet no such work on antonyms –

the corpus-based work on antonyms that exists is more restricted in terms of the number of linguistic characteristics it includes. Second, from the perspective of the degree to which distributional similarities/differences reflect functional similarities/differences, antonyms constitute an interesting case: intuitively at least, the degree of functional similarity of antonymous words should be smaller than that of synonymous words (because antonyms refer to the same semantic continuum/dichotomy, but to opposite positions/ends whereas synonyms refer to the same semantic continuum/dichotomy and to the same positions/ends), but maybe larger than that of the senses of polysemous words (because different senses of polysemous words can refer to very different semantic domains even if those may be metaphorically or otherwise related), with the caveat of the fuzziness that comes with decisions concerning distinctions between, and relatedness of, senses. On the other hand, the adjectives included in this study of course also exhibit some degree of polysemy, the evaluative sense of *great* probably being the most blatant case. It is therefore not immediately obvious how well, if at all, the BP approach can handle the distributional characteristics of antonyms, especially when, and this is a third point, there is no BP-based study in which synonyms and antonyms are combined. Given all this, while the BP approach has been validated and experimentally supported for synonymous and polysemous words, this paper is still largely exploratory and hopes to shed some first light of the joint similarity of synonyms and antonyms.

The remainder of this paper is accordingly structured as follows: in Section 2, we explain how the data were retrieved from the corpus, how they were coded for the subsequent analysis, and how they were analyzed quantitatively. In Section 3, we present the results of several different kinds of analysis of the data; the main differences will revolve around whether both morphosyntactic and semantic data are included in the statistical evaluation. In Section 4, we will very briefly summarize the main points and conclude.

2 Data and methods

2.1 Data

We first used an R script (cf. R Development Core Team 2008) to retrieve all matches of the lemmas *big*, *great*, *large*, *little*, *small*, and *tiny* (plus their comparative and superlative forms) when tagged as adjectives within their parse units as well as their and their clause's annotation from the British Component of the International Corpus of English. The data were exported into a spreadsheet software and then annotated for a variety of features:

- morphological features such as the tense, voice, and transitivity marking of the finite verb of the clause in which the adjective is used, etc.;
- syntactic features such as the syntactic frame in which the adjective is used (attributively vs. predicatively), the clause type in which the adjective is used (main vs. subordinate plus different kinds of subordinate clauses), the function of the clause in which the adjective is used, etc.;
- semantic features of the noun the adjective is modifying (count vs. non-count, concrete vs. abstract vs. human vs. organization/institution vs. quantity vs. ongoing processes vs. punctual events etc.); how SIZE is modified (literally vs. metaphorically vs. quantitatively vs. evaluatively), etc.

The resulting spreadsheet consisted of 2,073 rows (of matches) and 27 columns (of annotation, including case numbers as well as preceding and subsequent contexts).[5] It was evaluated statistically using the interactive R script BP 1.0 (Gries 2008), which has been written by, and is available from, the first author. This statistical evaluation will be explained in the following section.

2.2 Methods

The input to BP 1.0 consists of a table that contains each match of one of the adjective forms in a separate row and the annotation of each linguistic feature in a separate column. A random excerpt of this table is shown in Table 1, with the matches they belong to represented in (1):

Table 1: Excerpt of the table entered into BP 1.0

Form	Syntax	Modifiee_count	Modifiee_what	Clause function	Clause level
bigger	attributive	count	abstract	OD	depend
large	attributive	non-count	quantity	NPPO	depend
little	attributive	count	organization/institution	PU	main

(1) a. I guess size is a *bigger* problem actually than funding (S1B-076)
b. have to be transmitted in the UHF portion of the spectrum of the *large* of bandwidth required (W2B-034)
c. […] our own *little* <,,> magic circle or whatever it is […] (S1A-027)

The script converts this input table into an output table of behavioral profiles, an excerpt of which is shown in Table 2. This table, for example, shows that 87 percent of the uses of *big* were attributive while the remaining 13 percent were predicative, or that the percentages with which *big* and *bigger* are used for count and non-count modifiees are nearly identical but rather different from those of *great:* 94 percent: 6 percent vs. 95 percent : 5 percent vs. 71 percent : 29 percent; note how, in each column, the sum of the first three rows and the sum of the last two rows is 1, which means the within-ID tag percentages add up to 100 percent. These columns, i.e. the vectors of percentages, are the behavioral profiles of the word forms in the table header.

Table 2: Excerpt of the table returned by BP 1.0

ID tag	**ID tag level**	*big*	*great*	*large*	*bigger*
Syntax	adverbial	0	0.01	0	0
	attributive	0.87	0.83	0.91	0.45
	predicative	0.13	0.16	0.09	0.55
Modifiee_count	count	0.94	0.71	0.98	0.95
	non-count	0.06	0.29	0.02	0.05

While this first output of BP 1.0 is of course just a reorganization of the data, albeit a useful one, the more revealing next step is to compare the behavioral profiles of the adjective forms to each other. As in nearly all previous BP studies, we used a hierarchical agglomerative cluster analyses, but we had to decide on three parameters:

- which ID tags to include in the comparisons – all of them? only all/some morphosyntactic ones? only all/some semantic ones? some combination of both?
- how to compare the BP vectors mathematically – Euclidean distances? City-Block metric? correlational measures?
- how to amalgamate sufficiently similar (clusters of) BP vectors into clusters – average similarity, complete similarity, Ward’s method?

As for the ID tags, we decided to run three analyses: one with all ID tags, one with all syntactic ID tags, and one with all semantic ID tags. As for the computation of the cluster analyses, we decided to use Canberra as a similarity measure and Ward’s method as an amalgamation rule. These are the settings used in *all* earlier BP studies, which makes the approach maximally comparable to previous work and also means that the present study uses the same parameters that were successfully validated in two experimental studies in Divjak and Gries (2008).

In the following section, we will discuss the results of these three cluster analyses**.**

3 Results

3.1 Results for all ID tags

The results of the cluster analysis for all ID tags are shown in Figure 2, with several relevant clusters highlighted (adjective forms not shown – *tinier* and *least* – were not found with the relevant part-of-speech tag):

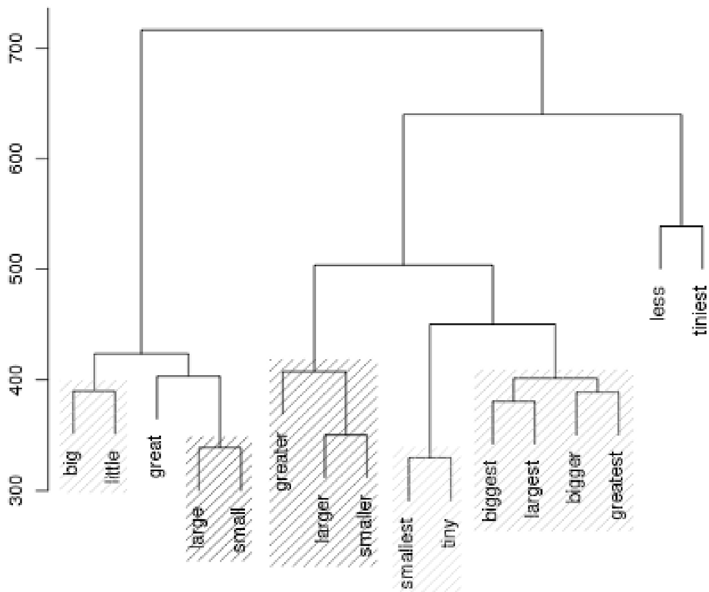

Figure 2: Dendrogram of the cluster analysis of the BP vector involving all ID tags

Given the structure of the dendrogram, these findings are surprisingly meaningful on a finer level of granularity and rather interesting. One striking finding is how different dimensions/parameters appear to have given rise to the above clustering solution. These parameters include oppositeness of meaning, sameness of meaning, and morphological parameters, i.e., all the parameters that figured in the choice of adjectives to study (more on this below).

Starting from the bottom, the first cluster to be formed is based on sameness of meaning: *tiny* of course means 'very small'. Then, focusing on the left side of the dendrogram, the analysis successfully identifies what in several other corpus-based *and* experimental studies are considered the canonical antonym pairs {*big little*} and {*large small*}. In addition, these two lowest-level clusters are clustered together with another adjective in the base form, *great*, which makes the larger leftmost cluster morphologically perfectly homogeneous as it contains only adjective base forms (but only nearly all, since *tiny* is not included). Then, there is an early lowest-level cluster that combines comparatives of at least one canonical antonym pair, {*larger smaller*}, which is joined with, again, a comparative, and again a form of the lemma GREAT, namely *greater*. Thus, this larger cluster is not only also morphologically homogeneous, but it has the same substructure as the base form cluster on the left in that the more polysemous (since evaluative) adjective *great* is somewhat less similar: {{LARGE SMALL} GREAT}.

Turning to the final cluster that is highlighted in Figure 2, it involves both a high degree of semantic sameness – all forms are from the {*big great large*} end of the size spectrum – and a slightly smaller degree of morphological coherence – three of four forms are superlatives, and the fourth form is a comparative. The last cluster in Figure 2 consists of *less* and *tiniest*, each of which occurs only once in the data so that this cluster can safely be disregarded.

Given these results, a sceptic may argue that the results are in fact not particularly interesting: after all, the three different parameters with regard to which the dendrogram can be interpreted – sameness of meaning, oppositeness of meaning, and morphological form – are the ones that synonyms and antonyms of different inflectional forms from the semantic field of SIZE would be expected to give rise to, and these parameters allow for a wide range of seemingly encouraging solutions, especially since sameness and oppositeness of meaning cover both ends of a single continuum and appear to render the approach somewhat futile. It is worth recognizing, however, that this view is mistaken, which can be easily demonstrated: if, as a sceptic may claim, any dendrogram was going to make some kind of sense – because there are so many parameters in the light of which the cluster structure could be interpreted favorably – then it should be possible to just randomly reorder the adjectives in the dendrogram and still find a lot of interesting things to say; consider the two panels of Figure 3:

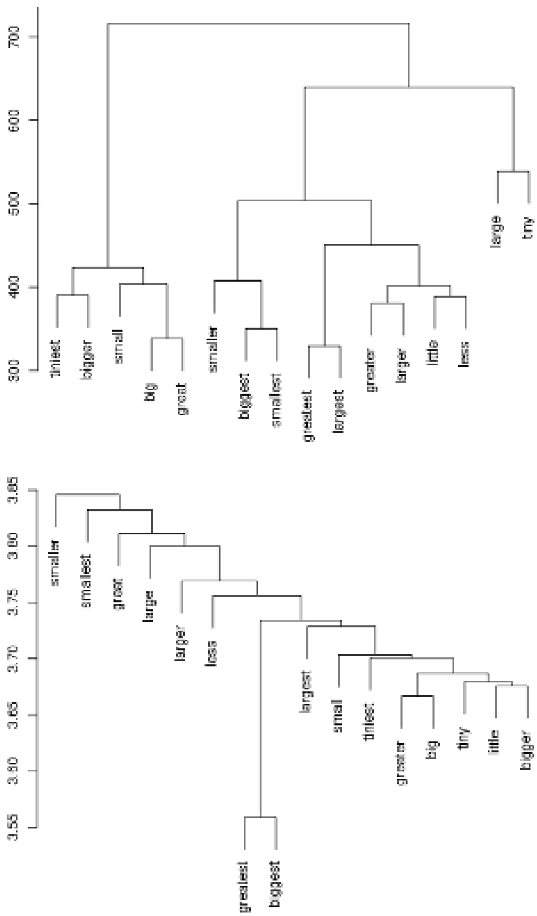

Figure 3: Dendrograms of the BP vectors with random adjective assignments

In the upper panel, we kept the existing dendrogram structure (i.e., the divisions of the branches etc.) but we reordered the adjectives in a random fashion. It is immediately obvious that there are of course still somewhat synonymous and somewhat antonymous clusters, but it is equally obvious that this random solution is much worse than the authentic one on all counts: the canonical antonyms are not identified (with any morphological forms), the amount of synonym clusters is low, and morphologically the solution is a mess.

In the lower panel, we illustrate that a cluster analysis can also return dendrograms with very different kinds of structure, which are then also much less revealing to interpret. Instead of a tree with several fairly delimited substructures as obtained in the authentic data, we here get a long-chained substructure which, with the exception of *greatest* and *biggest*, exhibits hardly any structure, given the long and rather indiscriminate chaining of adjectives. Thus, the structure and patterns of the authentic data in Figure 2 are not just a to-be-expected artifact of a powerful clustering algorithm – the range of possible clusters solutions is huge but the one that was actually obtained allows for meaningful interpretation on many different levels. We therefore consider these encouraging and interesting results.

3.2 Results for all morphosyntactic ID tags

In this section, we will very briefly discuss the results of the cluster solution obtained on the basis of only the morphosyntactic ID tags. However, since the morphosyntactic ID tags constitute the vast majority of all ID tag levels (520 out of 539), the result of this analysis is nearly identical to that of Figure 2; we therefore do not show the dendrogram here. The only difference to Figure 2 is that *biggest* moves from *largest* to become part of the cluster {*smallest tiny*}. In this case, the explanatory parameters *do* motivate both the first and the second solution: in Figure 2, *biggest* was grouped together with another superlative form from its synonym set (*largest*) whereas in the syntax-only tree *biggest* is grouped together with a strong synonym cluster that contains its canonical antonym in the superlative form *(smallest*). However, since this is the only change in an otherwise identical solution and since here even both solutions make sense – different kinds of sense, though – the BP approach at least does not suddenly yield a very counterintuitive result.

3.3 Results for all semantic ID tags

Let us now finally turn to the last cluster analysis, which is based on our semantic annotation only. Recall that this analysis is based on only 19 ID tag levels and that these are the ID tag levels which were more problematic to code. Consider Figure 4 for the resulting dendrogram:

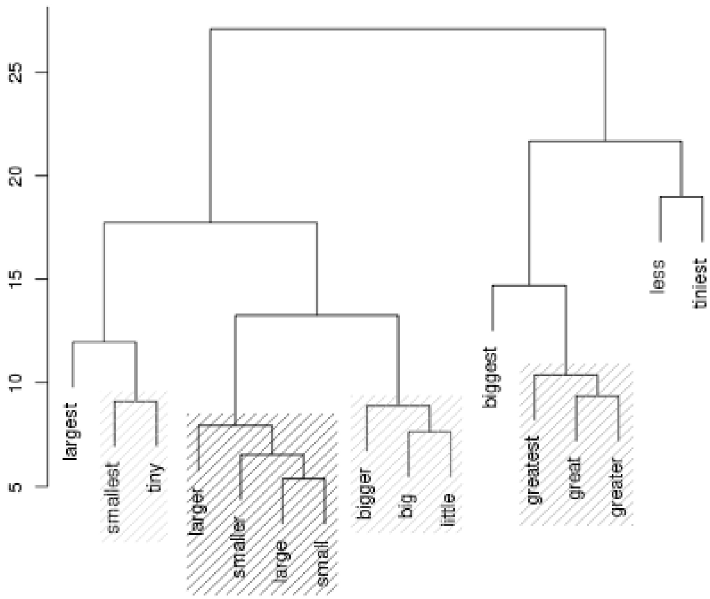

Figure 4: Dendrogram of the cluster analysis of the BP vector involving semantic ID tags

The dendrogram is a bit different than the one based on all the data or the syntactic data, but the results are still far from erratic. In fact, the dendrogram structure is maybe even a bit more revealing in terms of canonical antonymy (but maybe less revealing with regard to other things): supporting Jones *et al*. (2007), who find that among the size adjectives, *large* and *small* are most canonical, the pair *large*/*small* is here instantiated in the earliest cluster, which contains these adjectives' base *and* comparative forms, and the pair *big*/*little* is instantiated by another similar cluster from which only *less* is missing. We still obtain the cluster {*smallest tiny*}, which is now, just like the strong correlation of the canonical antonyms, attested in all three solutions. One interesting change compared to the previous dendrogram involves the lemma GREAT, whose three inflectional forms now constitute one cluster. Another change is the location of *biggest*, which does not appear to reveal much anymore – the only positive thing that can be said about it is that it is closest to a cluster involving an adjective from its synonym set that includes a superlative, but we have to admit that this seems somewhat far-fetched. Also, *largest* is not positioned ideally, but at least part of the cluster that also contains *smallest*, the superlative form of its canonical antonym. Finally, *less* and *tiniest* can again be disregarded given their rarity.

Again, the results are rather encouraging for the BP approach. Even with a much smaller number of ID tag levels to work with and a coding that may be a bit less reliable than the syntactic annotation that comes with the corpus, the results strongly reflect distributional similarities in the form of clusters that reflect previous corpus-based findings and experimental results regarding distributional similarities and canonical antonymy.

3.4 Post hoc results for all ID tags

In this section, we would like to briefly explore one useful added benefit of the BP approach. In addition to the multivariate analysis of different words, it is also very easy to determine how words differ from each other in a pairwise fashion, especially words that have been clustered together – i.e. are highly similar – but are of course still different. Divjak and Gries (2006) use *z*-scores to this end, but a conceptually much more straightforward approach is discussed in Divjak and Gries (2009). This approach is based on the recognition what BP vectors are; recall from Table 2 that BP vectors are simply aligned observed percentages of co-occurrence. For example, Table 2 above already revealed that, with regard to the countability of the modifiee, *big* is more similar to *large* than to *great*. Thus, if one computes the pairwise differences between percentages of ID tag levels and sorts the vectors according to the size of the difference, then one can see where two words exhibit their most marked differences. Consider Figure 5 for this kind of comparison of *little* and *big* (using all ID tags):

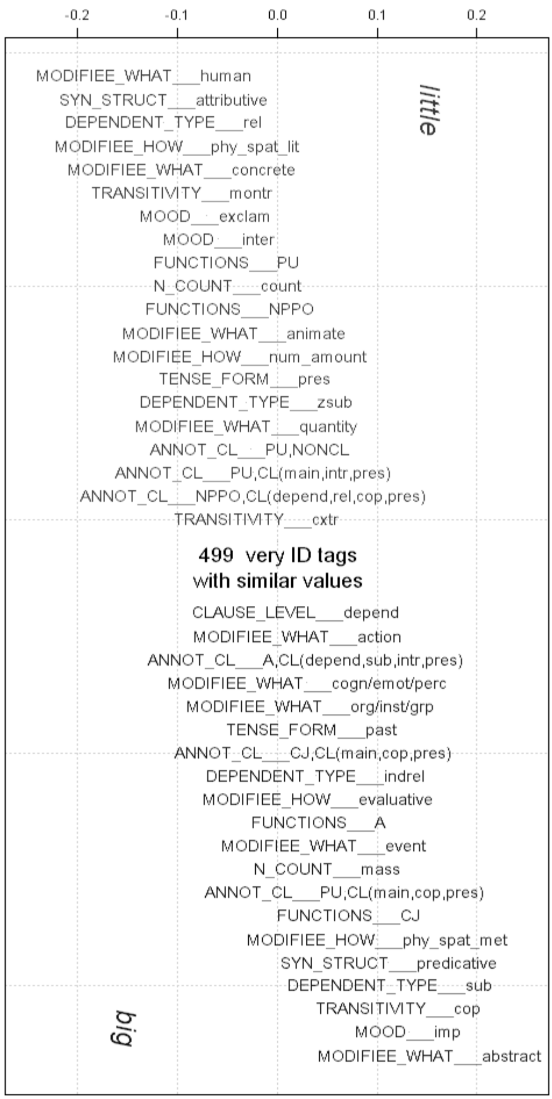

Figure 5: Snakeplot of differences between BP vector values: % big - *%* little

On the *y*-axis, we plotted the differences between the observed percentages between the two vectors for *little* and *big*. Thus, the further away from 0 (the center of) an ID tag level is plotted, the more distinctive it is for *little* or *big*. For example, 29 percent of the modifiees of *big* are abstract, but only 13 percent of the modifiees of *little* are, too, so *Modifiee_what: abstract* is plotted (centered) at 0.29-0.13-0.16. It is therefore immediately obvious that, while *big* and *little* are very similar to each other, they differ most strongly with regard to the semantic characteristics of their modifiees (and this kind of pronounced difference is why the semantic clustering could yield a very good result with far fewer ID tags): for instance, *big* is preferred with nouns that refer to abstract things, events, or organizations/institutions and metaphorically in predicative constructions. In comparison with *big*, *little* is preferably used for the actual physical size of (referents of) nouns referring to concrete things, in particular humans and other animate entities, and in attributive constructions. Similar comparisons are of course available for all desired pairwise comparisons but, given space constraints, one additional example shall suffice. In Figure 5, we compared *big* to its canonical antonym *little* so as to illustrate how snakeplots reveal patterns it makes sense to now also compare *big* to its closest synonym in the present data, *large*; consider Figure 6:

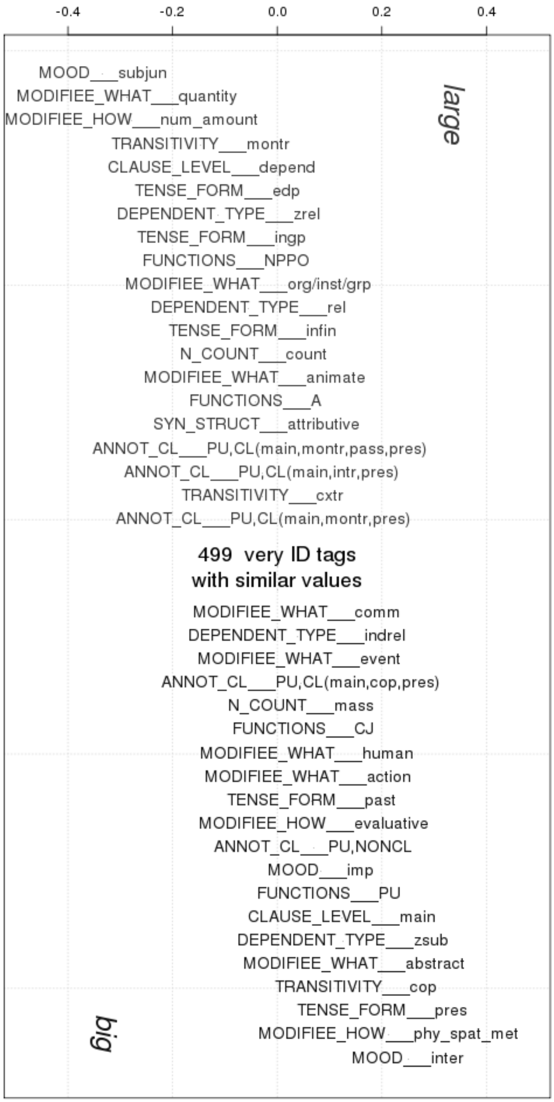

Figure 6: Snakeplot of differences between BP vector values: % big *- %* large

It is immediately clear why *big* was clustered with *little* and not with *large*: the deviations from 0 are much larger in Figure 6 than in Figure 5, since some differences become as small as -0.3. But how do the two words differ? The major difference is that *large* is preferably used to modify count nouns that refer to quantities, but also organizations/institutions and animate entities (not humans) – *big*, by contrast, modifies rather non-count nouns and co-occurs with abstract nouns, but also humans and actions. Interestingly, *big* disprefers attributive use compared to both *little* and *large*, but, compared to *large*, *big* has no similar preference for predicative use.[6]

4 Discussion

4.1 Interim summary and initial evaluation

In terms of the structure of the lexico-semantic space of the adjective studied here, the BP approach alone yields many results that were previously obtained in very many (methodologically) different studies, but nearly all of which make a lot of sense. In a completely data-driven fashion but without any subject ratings etc.

- the dendrogram finds that *tiny* 'means' *smallest* (cf. Deese's 1964 ratings);
- the dendrogram returns the canonical antonym pairs *big*/*little* and *large*/*small*; (cf. most studies on canonical antonyms, but recall that e.g. Jones *et al*.'s (2007) approach was less inclined to associate *big* with *little*;
- the morphologically clean clusters reflect subjects' preference to respond to a stimulus words with one of the same inflectional form (cf. Ervin-Tripp 1970).

And, to a considerable degree at least, these findings are obtained both from a large number of syntactic ID tags, from a small number of semantic ID tags, and the combination of both. In addition and as a by-product, the pairwise differences of ID tags allow us to immediately identify which distributional differences between the near synonyms are most pronounced (and would maybe even be most useful to learners). While similarly good results have already been obtained for polysemous words and sets of only synonymous words before (and in the case of the latter these results could even be experimentally validated), the present study is the first to have applied BPs to a potentially more noisy/chaotic data set in which synonyms and both direct and indirect antonyms were included – but the results still turned out to be very reasonable.

4.2 Theoretical implications

This successful application is of course good news for behavioral profilers in particular, but also for corpus linguists in general since it strongly supports the working assumption underlying most corpus-based work – formal differences reflect functional differences. However, we feel that the BP approach takes this approach more seriously and, thus, provides a better link to recent theoretical and empirical developments than simple collocational or colligational studies can offer. The main difference between the BP approach (or its earlier but similar precursors) and other simpler methods is the range of contextual/distributional characteristics that are taken into consideration. Rather than looking at one position (e.g. R1 for adjectives), at a window around a search word (4L to 4R), one syntactically-defined slot (e.g. the noun slot in an NP, the direct object slot of a verb, or one vacant slot in a contrastive construction), the BP approach includes dozens of linguistic features from different linguistic levels. This is not only desirable because it is *per se* more comprehensive. It is also desirable because this property of the BP approach and, more generally, this kind of perspective makes it possible to

- relate Miller and Charles's (1991) notion of a contextual representation to more recent development in corpus linguistics;
- connect this approach to recent cognitive and psycholinguistic trends, in which corpus linguists should be very interested;
- at least in some sense, re-evaluate the dichotomy between co-occurrence and substitutability.

Recall Miller and Charles's (1991:26) definition of a contextual representation: "a mental representation of the contexts in which the word occurs, a representation that includes all of the syntactic, semantic, pragmatic, and stylistic information required to use the word appropriately." It may not be immediately obvious, but this quote is in fact remarkable in the way it connects cognitive or psycholinguistic approaches to language to corpus-linguistic approaches – a possibility that many older-school corpus linguists are reluctant to entertain. One particularly clear connection of this notion of contextual representation is to Hoey's theory of lexical priming. Consider this recent quote:

> The notion of priming as here outlined assumes that the mind has a mental concordance of every word it has encountered, a concordance that has been richly glossed for social, physical, discoursal, generic and interpersonal context. This mental concordance is accessible and can be processed in much the same way that a computer concordance

> is, so that all kinds of patterns, including collocational patterns, are available for use. (Hoey 2005:11)

While Hoey does not refer to Miller and Charles's work, we still find the overlap between this metaphor of a comprehensive mental concordance and contextual representations rather striking, and it is exactly this richly glossed mental concordance, this contextual representation, that the BP approach tries to approximate. Given this close connection between contextual representations and Hoey's new and exciting new framework, exploring these notions is therefore still a very relevant issue.

One way of exploration is concerned with the question of how the contextual representations that give rise to subjects' behavior in experiments and the patterns in the corpus data come about in the first place, a question that the previous literature referred to so far does not discuss in much detail. Consider Charles (2000:507):

> Similarly, the contextual representation of a word is not an actual linguistic context but an abstraction of information in the set of natural linguistic contexts in which a word occurs. […] *Although the process that is used to derive a contextual representation from multiple encounters is difficult to describe*, many derivatives have been characterized. (our emphasis)

While the quote from Charles states that the process resulting in a contextual representation is still "difficult to describe," recent developments in, among other areas, cognitive linguistics, first language acquisition, and phonology, have made substantial progress in this regard (both theoretically and empirically). Our own take on this is that an exemplar-based framework is probably the best way to conceive of the generation of a contextual representation or a mental concordance.

Consider Dabrowska (2009) as one recent cognitive-linguistic example that is exemplar- and usage-based. Dabrowska investigates the meanings of rare verbs of walking or running such as *stagger*, *hobble*, *plod*, or *saunter* and shows that verbs are reliably associated with semantic and collocational preferences of the main arguments and complements of the verbs. Learners acquire the meanings of words on the basis of contextual and distributional cues provided in usage events by (i) storing lexically-specific knowledge of semantic and collocational preferences and (ii) forming more phonologically and semantically abstract generalizations or schemas on the basis of recurrent exposure to particular components of meaning.

It is immediately apparent how well this characterization can be mapped onto the adjectives of size studied here. Following this logic and earlier BP-based exploration of synonyms by Divjak and Gries (2008), we here also argue for a view

> [...] in which acquisition depends on exemplar learning and retention, out of which permanent abstract schemas gradually emerge and are immanent across the summed similarity of exemplar collections. These schemas are graded in strength depending on the number of exemplars and the degree to which semantic similarity is reinforced by phonological, lexical, and distributional similarity. (Abbot-Smith and Tomasello 2006:275)

(Cf. also, for example, Bybee 2000; Pierrehumbert 2001; Goldberg 2006: part II). This hybrid view implies that acquiring contextual representations of the size adjectives involves memorizing a 'cloud' of exemplars – in multidimensional syntactic-semantic space. Whenever a speaker encounters yet another instance of these adjectives, the memory representation of these adjectives and their actual uses – the mental concordance – is updated with the information contained in the most recent usage event. But recall that Charles as well as Abbot-Smith and Tomasello also speak of abstraction: of course not all actual instances are remembered. Memory traces may decay over time, and while particular salient usage events may remain accessible, what remains for the most part may well be generalizations based on many similar but now forgotten usage events.

These generalizations are assumed to involve probabilistic knowledge of distributional patterns (in this case, for example, the combination of semantic properties of modifiees with grammatical co-occurrences or colligations), and in the BP approach these straightforwardly correspond to the distributions of ID tag levels. On this view, the results of both experimental techniques and authentic language production as captured in corpora result from speakers accessing traces of memory representations for the use of the adjectives. More specifically, in contexts-sorting tasks as advocated within the substitutability approach, the contextual clues provided in the contexts facilitate access of a particular sub-region of the syntactic-semantic space containing a cloud of traces for adjectives that were used in a similar way. The likelihood that subjects produce the same adjective thus increases strongly, and the same sub-region is accessed in the case of language production when a speaker's processing system seeks for the right lexical and syntactic material to express a combination of concepts.

From this perspective, the *conceptual* distinction between co-occurrence and substitutability approaches pretty much collapses: the collocation data from co-occurrence studies and the sorting data from substitutability approaches are ultimately due to the speakers processing the same regions in syntactico-semantic space. A bit polemically, all that remains is the *methodological* distinction between 'noisy' co-occurrence data and 'nicely controlled' experimental data. Crucially, however, the BP approach with its multidimensional annotation of every single corpus example captures what both co-occurrence and substitutability approaches tap into, the richly glossed mental concordance with distributional data of past usage events.

Given the consistent empirical utility of the BP approach for several lexical relations, its ability to provide many results previously obtained separately, its experimental validation for synonyms, its compatibility with the most recent psycholinguistic developments from language acquisition – after all, the observed patterns must be learnable somehow – and the new perspective it offers, we hope that this study has contributed to BPs being used to explore an even larger variety of linguistic phenomena.

Notes

* We thank Carita Paradis for her valuable comments and feedback. The usual disclaimers apply.

1. To avoid unnecessary prolixity, we will from now on just use the term *synonyms* to refer to words that are semantically similar and may on at least one occasion be used functionally interchangeably.
2. Our characterization of "two empirical perspectives" is of course not comprehensive and heavily biased towards the methods employed in studies of synonymy and antonymy as well as, for reasons that will be discussed below, towards studies that involve Miller and Charles's notion of a contextual representation. Naturally, many methods other than the two we focus on here have been employed, such as word association tests, judgment tests, forced-selection tests, gap-filling tests, and no doubt many others.
3. They note that "[t]his situation had been noted before, of course, but was dismissed with a rationalisation that antonyms may not be as different in meaning as they seem" (1991:26) but unfortunately, they do not really provide any argument why this rationalisation, which is after all not that uncommon, may not be correct.
4. Jones (2001) again replicates this by then already comprehensively documented finding, but there are also some potential problems with his

approach. First, he argues that his findings are more trustworthy than Justeson and Katz's:

> For example, Justeson & Katz calculate an Observed/Expected ratio of 19.2 for *happy*/*sad*, based on individual frequencies of 89 and 32 respectively and observed co-occurrence in just one sentence. My corpus yields an Observed/Expected ratio of 6.8 for *happy*/*sad*, based on individual frequencies of 28,217 and 9,420 respectively and observed co-occurrence in 140 sentences. It seems fair to conclude that statistics derived from the latter corpus will be more trustworthy.

However, we consider this far from "fair to conclude" since even if Jones's corpus is larger, it is also a much more restricted convenience sample. Jones (2001) uses a 280m words corpus, whereas Justeson and Katz (1991) use the 1m word Brown corpus, but Jones's corpus contains journalese from a single newspaper only, whereas the Brown corpus has been sampled much more widely and carefully.

Second, he criticizes previous studies' antonym lists as too intuition-based and outdated, but his own solution is to just make up a list on his own, which, even if that list is praised as "customised to meet the demands of this research and relevant to a 21st Century investigation of antonymy" (p. 299) is certainly not as objective as, say, Deese's more empirical approach.

5. The annotation of the data was mostly unproblematic, since we could rely on the remarkable coding of the ICE-GB corpus compilers. The only tricky part of the annotation was the annotation of the semantic characteristics of the modifiee. While disagreements between the authors were resolved in discussion, there are of course ambiguous cases which were hard to decide. While we tried to code those to the best of our abilities and did quite some checking for consistency, it is unlikely that all of our decisions are uncontroversial. However, given the size of the multivariate data set, it is inconceivably difficult to bias the data in one particular direction on purpose, and in the analyses discussed below, we compare the results of an analysis based on the many uncontroversial morphosyntactic characteristics to those of the few more controversial semantic characteristics as a control.
6. In our data, the preferences for attributive and predicative use of *little* and *small* are different from the findings reported by Biber *et al*. (1998). Our BP vectors indicate that *little* prefers attributive use and *small* disprefers predicative use, whereas Biber *et al*. (1998:93f.) find that *small* prefers predica-

tive use more strongly than *little* (esp. in conversation). This kind of difference is not particularly encouraging, since their conversational data are from British English from the 1990s just like ours, but they may still most likely be due to the different corpus compositions and different parsing schemes. However, most relevant to the antonymy perspective of the present study is the fact that Biber *et al*.'s (1998) findings also result in the association of *big* to *little* and *large* to *small* that we and many other studies found.

References

Abbot-Smith, Kirsten and Michael Tomasello. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23: 275–290.

Arppe, Antti. 2008. *Univariate, bivariate, and multivariate methods in corpus-based lexicography: A study of synonymy*. Helsinki: University of Helsinki.

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

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*, 17–36.

Atkins, Beryl T. Sue and Beth Levin. 1995. Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography* 8 (2): 85–114.

Berez, Andrea L. and Stefan Th. Gries. 2009. In defense of corpus-based methods: A behavioral profile analysis of polysemous *get* in English. *Proceedings of The 24th Northwest Linguistics Conference* (University of Washington Working Papers in Linguistics 27), 157–166. Seattle, WA: Department of Linguistics.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Bolinger, Dwight L. 1968. Entailment and the meaning of structures. *Glossa* 2: 119–127.

Bybee, Joan. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow and S. Kemmer (eds.). *Usage-based models of language*, 65–85. Stanford: CSLI Publications.

Charles, Walter G. 1988. The categorization of sentential contexts. *Journal of Psycholinguistic Research* 17 (5): 403–411.

Charles, Walter G. 2000. Contextual correlates of meaning. *Applied Psycholinguistics* 21 (4): 505–524.

Charles, Walter G. and George A. Miller. 1989. Contexts of antonymous adjectives. *Applied Psycholinguistics* 10 (3): 357–375.

Church, Kenneth Ward, William Gale, Patrick Hanks and Donald Hindle. 1991. Using statistics in lexical analysis. In U. Zernik (ed.). *Lexical acquisition: Exploiting on-line resources to build a lexicon*, 115–164. Hillsdale, NJ: Lawrence Erlbaum.

Church, Kenneth Ward, William Gale, Patrick Hanks, Donald Hindle and Rosamund Moon. 1994. Lexical substitutability. In B.T.S. Atkins and A. Zampolli (eds.). *Computational approaches to the lexicon*, 153–177. Oxford and New York: Oxford University Press.

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Dabrowska, Ewa. 2009. Words as constructions. In V. Evans and S.S. Pourcel (eds.). *New directions in cognitive linguistics*, 201–223. Amsterdam and Philadelphia: John Benjamins. [The typesetting software used did not reproduce the second letter in the author's last name as it appears in the source.]

Deese, James. 1962. Form class and the determinants of association. *Journal of Verbal Learning and Verbal Behavior* 1 (2): 79–84.

Deese, James. 1964. The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal Behavior* 3 (5): 347–357.

Divjak, Dagmar S. 2006. Ways of intending. In St. Th. Gries and A. Stefanowitsch (eds.). *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, 19–56. Berlin and New York: Mouton de Gruyter.

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. 2008. Clusters in the mind? Converging evidence from near-synonymy in Russian. *The Mental Lexicon* 3 (2): 188–213.

Divjak, Dagmar S. and Stefan Th. Gries. 2009. Corpus-based cognitive semantics: A contrastive study of phasal verbs in English and Russian. In K. Dziwirek and B. Lewandowska-Tomaszczyk (eds.). *Studies in cognitive corpus linguistics*, 273–296. Frankfurt a.M.: Peter Lang.

Ervin-Tripp, Susan. 1970. Substitution, context, and association. In L. Postman and G. Keppel (eds.). *Norms of word association*, 383–467. New York: Academic Press.

Fellbaum, Christiane. 1995. Co-occurrence and antonymy. *International Journal of Lexicography* 8 (4): 281–303.

Firth, John R. 1957. *Papers in linguistics*. Oxford: Oxford University Press.

Gibbs, Raymond W. Jr. and Teenie Matlock. 2001. Psycholinguistic perspectives on polysemy. In H. Cuyckens and B. Zawada (eds.). *Polysemy in cognitive linguistics*, 213–239. Amsterdam and Philadelphia: John Benjamins.

Gilquin, Gaëtanelle. 2003. Causative *get* and *have*. *Journal of English Linguistics* 31 (2): 125–148.

Goldberg, Adele E. 2006. *Constructions at work*. Oxford: Oxford University Press.

Gries, Stefan Th. 2003. Testing the sub-test: A collocational-overlap analysis of English *-ic* and *-ical* adjectives. *International Journal of Corpus Linguistics* 8 (1): 31–61.

Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: The many meanings of *to run*. In St. Th. Gries and A. Stefanowitsch (eds.). *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, 57–99. Berlin and New York: Mouton de Gruyter.

Gries, Stefan Th. 2008. *BP 1.0*. A program for R 2.7.1 and higher.

Gries, Stefan Th. and Dagmar S. Divjak. 2009. Behavioral profiles: A corpus-based approach towards cognitive semantic analysis. In V. Evans and S.S. Pourcel (eds.). *New directions in cognitive linguistics*, 57–75. Amsterdam and Philadelphia: John Benjamins.

Gross, Derek, Ute Fischer and George A. Miller. 1989. The organization of adjectival meanings. *Journal of Memory and Language* 28 (1): 92–106.

Hanks, Patrick. 1996. Contextual dependency and lexical sets. *International Journal of Corpus Linguistics* 1 (1): 75–98.

Harris, Zellig S. 1970. *Papers in structural and transformational linguistics*. Dordrecht: Reidel.

Hoey, Michael. 2005. *Lexical priming: a new theory of words and language*. London and New York: Routledge.

Jones, Steven. 2001. Corpus approaches to antonymy. In P. Rayson, A. Wilson, A. McEnery, A. Hardie and S. Khoja (eds.). *Proceedings of the Corpus Linguistics 2001 conference*, 297–310. Lancaster University: University Centre for Computer Corpus Research on Language.

Jones, Steven, Carita Paradis, M. Lynne Murphy and Caroline Willners. 2007. Googling for ‘opposites’: A web-based study of antonym canonicity. *Corpora* 2 (2): 129–154.

Justeson John S. and Slava M. Katz. 1991. Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics* 1 (1): 1–19.

Kishner, Jeffrey M. and Raymond W. Gibbs Jr. 1996. How *just* gets its meanings: polysemy and context in psychological semantics. *Language and Speech* 39 (1): 19–36.

Miller, George A. 1999. On knowing a word. *Annual Review of Psychology* 50: 1–19.

Miller, George A. and Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6 (1): 1–28.

Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam and Philadelphia: John Benjamins.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.). *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam and Philadelphia: John Benjamins.

Rubenstein, Herbert and John B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM* 8 (10): 627–633.

Schmid, Hans-Jörg. 1993. Cottage *und Co.,* idea, start *vs.* begin. Tübingen: Max Niemeyer.

Wang, Chueh-chen. 2006. What distinguish synonymous constructions: A corpus-based study of Mandarin *lian … dou* and *lian … ye* constructions. Paper presented at GCLA 2006.