Using web data to investigate antonym canonicity

Jones, Steven; Murphy, Lynne; Paradis, Carita; Willners, Caroline

2007

Googling for 'opposites': a web-based study of antonym canonicity

Steven Jones,[1] Carita Paradis,[2]
M. Lynne Murphy[3] and Caroline Willners[4]

Abstract

This paper seeks to explain why some semantically-opposed word pairs are more likely to be seen as canonical antonyms (for example, *cold*/*hot*) than others (*icy*/*scorching*, *cold*/*fiery*, *freezing/hot*, etc.). Specifically, it builds on research which has demonstrated that, in discourse, antonyms are inclined to favour certain frames, such as '*X* and *Y* alike', 'from *X* to *Y*' and 'either *X* or *Y*' (Justeson and Katz, 1991; *etc*.), and to serve a limited range of discourse functions (Jones, 2002). Our premise is that the more canonical an antonym pair is, the greater the fidelity with which it will occupy such frames. Since an extremely large corpus is needed to identify meaningful patterns of co-occurrence, we turn to Internet data for this research. As well as enabling the notion of antonym canonicity to be revisited from a more empirical perspective, this approach also allows us to evaluate the appropriateness (and assess the risks) of using the World Wide Web as a corpus for studies into certain types of low-frequency textual phenomena.

1. Introduction

More than members of any other semantic relation (synonyms, hyponyms, *etc*.), antonym pairs are able to achieve special, 'canonical' status in a

[1] School of Education, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom
Correspondence to: Steven Jones, *e-mail*: stevejones@manchester.ac.uk
[2] School of Humanities, Växjö University, 351 95 Växjö, Sweden
[3] Department of Linguistics and English Language, University of Sussex, Falmer, Brighton, BN1 9QN, United Kingdom
[4] Department of Linguistics and Phonetics, Lund University, Lund, Sweden, SE– 221 00

Corpora Vol. 2 (2): 129–154

language.[5] In the literature, some (e.g., Gross *et al*., 1989; Charles *et al*., 1994) assume that antonym pairs are either canonical (for example *old*/*young*, *cold*/*hot* and *happy*/*sad*) or non-canonical (*aged*/*youthful*, *cool*/*hot*, *happy*/*miserable*), while others assume or argue for a continuum between the two categories (e.g., Herrmann *et al*., 1979; Murphy, 2003). Among the methods that have been used to investigate antonym canonicity are word association tests (Deese, 1965; Clark, 1970), judgement tests (Herrmann *et al*., 1986) and elicitation experiments (Paradis *et al*., forthcoming). This paper approaches the issue by building specifically on research that has demonstrated the tendency of antonyms to favour certain lexico-grammatical constructions in discourse, such as 'both *X* and *Y*', 'from *X* to *Y*' and 'whether *X* or *Y*' (Justeson and Katz, 1991; Mettinger, 1994; Fellbaum, 1995; Jones, 2002). In this paper, we argue that a language's most canonical antonym pairs can reasonably be expected to co-occur with highest fidelity in such constructions – that is, they will co-occur with each other, in preference to other semantically-plausible pairings, across the widest possible range of appropriate contexts. Given the relatively low frequency of such phrases in language, an extremely large corpus is needed in order to identify such patterns. The specific aims of this paper are, therefore:

- To assess the degree to which a series of lexico-grammatical constructions can be used as a diagnostic of antonymy;
- To measure the strength of antonym pairs belonging to ten semantic scales by examining their co-occurrence fidelity within these constructions; and,
- To evaluate the usefulness of the World Wide Web, as accessed through a freely-available search engine, as a corpus for research into certain types of low-frequency phenomena in language.

In addressing these specific aims, the more general issue of antonym canonicity is also dealt with. This issue is important because canonical antonyms are central to the organisation of adjectival meaning in those theories for which paradigmatic semantic associations between words contribute to the words' semantic value – for example, WordNet (Gross and Miller, 1990) and Meaning-Text Theory (Mel'čuk, 1996) – and because canonical antonyms are often needed for language applications, such as dictionaries and thesauri, computational lexicons and psychological/psycholinguistic experiments.

[5] See, for example, Cruse (1986: 197), who notes that 'of all the relations … oppositeness is probably the most readily apprehended by ordinary speakers', Jones (2002: 117) or Murphy (2003: 26).

2. Measuring antonym canonicity

For the purposes of this article, *antonym pair* refers to any two words that are semantically opposed and incompatible with respect to at least one of their senses, for example *chilly*/*warm*. An antonym pair is said to be *canonical* if the two words are associated by 'convention' as well as by semantic relatedness, for example, *private/public*. In other words, canonical antonym pairings have been learnt as pairings of lexical units (i.e., pairings of form-sense combinations), not just derived by semantic rules (i.e., sense-sense pairings). The notion of 'conventionality', however, is difficult to pin down; this paper assumes that more conventional pairings will be found to co-occur in a wider range of phrasal contexts; that is, they are not opposed just by virtue of being in one set phrase. By this criterion, *rich* and *poor* are more likely to have canonical status than *rags* and *riches*. Reciprocity of the relation is also assumed to be an indicator of canonicity. For example, searches may point to the 'best opposite' of both *fast* and *rapid* being *slow*. However, *slow* may only reciprocate this antonymy in the case of *fast*, not *rapid*. We claim, therefore, that the strength of antonym canonicity can be measured in terms of the reciprocal frequency of association between two words, and, more importantly, by the fidelity of the pairing.[6]

In general, studies into antonym canonicity have been based on either the results of metalinguistic activities or on corpus-based searches. To begin with the former, it has been noted that, 'language users can intuitively sort 'good' (or prototypical) antonyms from not-so-good ones and downright bad ones' (Murphy, 2003: 11). This is often referred to as the 'clang phenomenon' – a term used to describe the reaction to those pairs that intuitively strike the hearer as being good 'opposites' (Charles and Miller, 1989; Muehleisen, 1997). One example of a metalinguistic approach is supplied by Herrmann *et al*. (1986), who asked informants to judge the antonymy of 100 test pairs on a scale from one to five. The highest scoring pair was *maximize*/*minimize*, followed by pairs like *night*/*day* and *good*/*bad*. A less direct approach had been taken previously by Deese (1965) and Clark (1970), who used word-association tests to tap into intuitions about the relation. In such tests, informants are invited to say or write the first word that comes into their heads on hearing or reading a stimulus word. Among those words most frequently elicited by one another were *inside*/*outside* and *right*/*wrong*, providing

[6] In some cases, this can involve the extension of the antonym relation to other senses of the word, for example the use of *cold* to mean 'legally obtained' in contrast to the 'stolen' sense of *hot* (Lehrer, 2002) and the use of *white* to mean 'with milk' in contrast to black coffee (Murphy, 2006). In these cases, awareness of the canonical relation encourages the application of the words as a pair in semantic domains to which only one of the words has previously been applied.

evidence that responses to adjectival stimuli were, 'overwhelmingly contrastive or antonymic to the stimulus' (Deese, 1965: 347). However, because judgement tests and elicitation experiments are metalinguistic by nature, they assess not how language is used, but how informants reflect on the meaning(s) of given words and the relations that hold between them.

Corpus-based studies examine antonyms in natural language use and many have treated co-occurrence as a key indicator of canonicity (Charles and Miller, 1989; Justeson and Katz, 1991, 1992; Willners, 2001). This starting point seems reasonable given that antonyms co-occur within sentences 6.6 times more often than chance would allow (Jones, 2002: 115). Furthermore, 'direct' antonyms have been shown to co-occur three to twelve times more often than expected,[7] while other semantically-possible pairings from the same scales co-occur only 1.45 times more often than expected (Willners, 2001: 78). However, co-occurrence alone is not a reliable criterion for identifying antonyms because many pairs of words co-occur (e.g., *surf/net*, *climate*/*change*, etc.) without being in an opposite relation. Antonyms are distinguishable from other collocates because they tend to be distributed in a range of particular lexico-grammatical constructions and so tend to serve one of a small number of discourse functions in text (Jones, 2002).

Neither metalinguistic exercises nor co-occurrence criteria are ideal for assessing the canonicity of antonyms. The former are often biased towards basic, high frequency lexical pairings and, moreover, by the notion that words can only have one 'best' antonym. For instance, ask someone for the 'opposite' of *hot* and they are most likely to choose *cold* without considering that *cold* is not the antonym of *hot* in its 'spicy' sense. Corpus studies are better able to deal with a word having multiple antonyms, but most to date have searched for known canonical antonym pairs and compared them to pairs that are perceived as less canonical. Thus, they have not provided a means for discovering antonym pairs so much as a way to confirm existing intuitions regarding the antonymic relation. Since they measure frequency of co-occurrence, they are also more likely to treat as canonical those pairs that have more common words and senses. As we explain in the next section, this study combines the best aspects of both elicitation and corpus methods, but avoids some of their associated problems.

3. Methodology

The approach adopted here can be thought of as an antonym elicitation task that elicits antonyms from a corpus of natural language. This process

[7] Source: The Swedish Stockholm-Umeå Corpus (see Willners, 2001).

essentially involved three steps: step one involved identifying several grammatical constructions (which we refer to in this paper as 'frames') within which antonym pairs are known to co-occur frequently (e.g., '*X* and *Y* alike'); step two involved searching those frames 'seeded' with a range of single adjectives in either the *X*- or the *Y*-position (e.g., 'thin and * alike'); and step three involved examining which adjectives were retrieved most commonly in the wildcard '*' position (in this case, *thick*, *fat*, *heavy* and *overweight*). Therefore, while co-occurrence criteria were applied, a more fine-grained approach was adopted in order to distil those word pairs with the strongest claims to canonicity from those that contrast (or merely co-occur) in a restricted range of (possibly idiomatic) contexts. The results of metalinguistic experiments (e.g., Paradis *et al*., 2006) are here used for comparative purposes only; our aim is to privilege the evidence provided by natural language usage instead.

3.1 Selection of frames

The present study reverses the approach taken in many previous studies because instead of searching for antonym pairs in order to identify the types of phrases in which they co-occur, we searched the phrase-types[8] in order to identify antonym pairs. We chose to explore a wide range of antonym frames (as identified by Justeson and Katz, 1991; Mettinger, 1994; *etc*.) that reflect a wide range of discourse functions (Jones, 2002). These discourse functions were initially developed using newspaper corpora, but have also been found to account for antonym use in spoken English (Jones, 2006), in child-produced and child-received language (Jones and Murphy, 2005), and in Swedish (Murphy *et al*., forthcoming).

When deciding which contrastive constructions would be most appropriate for this study, any constructions that were less than four words long were ruled out initially. This was because search strings composed of two words and a wildcard often gave non-constituent results – a hazard compounded by using a corpus without grammatical tags. For example, a search for the '*X*, not *Y*' construction, in which *fat* is the seed word used in *X*-position, finds many examples in which *fat* and *not* occur in different clauses (e.g., 'it is the type of fat, not the amount, that is most important'). Pilot tests were carried out on eleven four-word constructions that were seeded with a variety of adjectives in order to rule out any that generated large amounts of

[8] Murphy (2006) proposes that the phrasal patterns associated with antonymy should be regarded as constructions, i.e., pairings of the partially lexicalised phrasal forms with particular contrastive meanings.

non-constructional 'noise'. The most productive frames were generally found to associate with the functional category of Co-ordinated Antonymy, for example, '*X* and *Y* alike', '*X* as well as *Y*', 'both *X* and *Y*', 'either *X* or *Y*', 'neither *X* nor *Y*' and 'whether *X* or *Y*'. In these constructions, the antonyms' inherent opposition is not activated, and the pair are united in order to exhaust a particular semantic scale (for example, 'he is neither optimistic nor pessimistic about his prospects', Jones, 2002: 71). Since the aim was to assess antonym co-occurrence across a range of functional categories, the number of co-ordinated constructions was limited to four. In addition, we chose 'from *X* to *Y*', 'between *X* and *Y*' and '*X* versus *Y*'. The last of these differs because it is a three-word, rather than a four-word construction. However, as pilot-testing confirmed, it is strongly associated with contrast and generates much less 'noise' than other three-word phrases. Placing the wildcard * alternately in the first and second adjective positions in the seven constructions, results in fourteen searchable frames, shown in Table 1.

Wildcard-first frame	**Wildcard-second frame**	**Functional type**
* *and* Adj *alike*	Adj *and* * *alike*	CO-ORDINATED
between * *and* Adj	*between* Adj *and* *	VARIOUS
both * *and* Adj	*both* Adj *and* *	CO-ORDINATED
either * *or* Adj	*either* Adj *or* *	CO-ORDINATED
From * *to* Adj	*from* Adj *to* *	TRANSITION / CO-ORDINATED
* *versus* Adj	Adj *versus* *	CONFLICT
whether * *or* Adj	*whether* Adj *or* *	CO-ORDINATED

Table 1: The fourteen search frames used in this study

As can be seen from the right-hand column of Table 1, there is not a one-to-one correspondence between the search frames and Jones's discourse-functional categories. For example, while 'from *X to Y*' is often transitional in nature (e.g., 'it turned from *X* to *Y*'), it can also occur within larger constructions that signal the entirety of a scale (e.g., 'run the gamut from *X* to *Y*'), in which case it serves a co-ordinated function. Likewise, 'between *X* and *Y*' can serve multiple functions in discourse. The main reason for varying the functional categories of the search frames was to include as broad a range of categories as possible so that that the antonym pairs identified through this method would be

valid 'opposites' and not, for example, words that just often happen to be conjoined.

3.2 Adjectives selected as seed words

The list of adjectives used to 'seed' the frames was taken from the stimuli and responses generated by an antonym elicitation task, conducted by Paradis *et al*. (2006). This study invited fifty informants to state the 'opposite' of a series of given adjectives.[9] These adjectives were then ranked according to the lowest number of different antonyms elicited. Some words were found to elicit the same antonym from all fifty informants (e.g., *clean* → *dirty*), while other words elicited as many as twenty-nine different responses (e.g., *calm* → *stressed*, *stormy*, *rough*, *agitated*, etc.). For this study, we randomly selected ten of the top forty adjectives on the list compiled by Paradis *et al*. (2006). These are recorded in Table 2, together with the 'opposites' elicited for each word. All of the adjectives listed in the first column of Table 2 were found to be strongly, uni-directionally associated by informants with one particular antonym (Paradis *et al*., 2006).[10] In order to learn more about entire semantic scales, not just individual adjectives, both the stimulus word and its majority 'opposite' were used as seed words in the present study. However, it should be noted that not all of these ten pairs would necessarily be regarded as canonical antonyms (e.g., *rapid/slow*).

In addition to the twenty adjectives emboldened in Table 2, post-hoc searches were conducted on any word that was not part of the original search list, but that was subsequently identified as a potential canonical antonym of one of those twenty adjectives. For example, since *fast* was returned by searches on *slow*, we later executed searches for *fast* in the fourteen frames.

3.3 Selecting a corpus

Piloting a similar approach to the one taken here, Jones (2002: 154–67) demonstrated that the favoured antonyms of *natural* in text are *artificial* and

[9] The stimuli for that experiment were, in turn, selected from a range of frequently co-occurring adjective pairs in the British National Corpus (see Paradis and Willners, forthcoming).

[10] The most common 'opposite' elicited for each stimulus word was suggested by at least 50 percent of all informants and, in each case, the popularity of this 'opposite' was more than double that of the second most common response.

Stimulus Word	Response Word(s)								
beautiful	**ugly** (50)								
poor	**rich** (50)								
open	**closed** (40)	shut (10)							
large	**small** (48)	little (1)							
rapid	**slow** (47)	sluggish (2)	slim (1)						
exciting	**boring** (36)	dull (13)	fast (1)						
strong	**weak** (47)	feeble (1)	unexciting (1)						
wide	**narrow** (45)	thin (3)	mild (1)	slight (1)					
thin	**fat** (35)	thick (13)	skinny (1)	slim (1)					
dull	**bright** (28)	exciting (10)	overweight (1)	wide (1)					
		interesting (8)	shiny (2)	lively (1)	sharp (1)				

Table 2: Stimulus and response adjectives (words in boldface were used as initial seed words in this study)

man-made (but not *unnatural*), and that *style* tends to be placed in opposition to *substance* most commonly in contemporary English. However, although these findings were based on a sizeable corpus (280 million words), only the relatively conventionalised antonyms of relatively frequent seed words could

be identified and, at lower levels of frequency, output was not always found to be contrastive. This study therefore required a much larger corpus to allow for the development of a more accurate and detailed antonym profile of many more adjectives. For this reason, we turned to the World Wide Web.

Of course, whether the web should be regarded as a 'corpus' remains open to debate. The web is not a structured collection of texts specifically compiled for linguistic analysis, nor is it representative of language in general — criteria that Kennedy (1998: 3) and Biber *et al*. (1998: 246) apply in their definition of corpora. Though some recent studies note elements of comparability between the web and traditional 'balanced' corpora (see Fletcher, 2004; Sharoff, 2006; or, for a more detailed review, Kilgarriff and Grefenstette, 2003), using the Internet for linguistic analysis remains problematic. For example, because data are not collated according to any sociolinguistic principles, issues arise concerning consistency (American vs. British English, unorthodox spelling, *etc*.) and duplication of the same texts (song lyrics, political speeches, *etc*.). Furthermore, our chosen search engine, Google, lacks the sophistication of purpose-built corpus-searching software: web-pages are not selected at random but rather sorted according to extraneous criteria (relevance of topic, popularity of web-site, *etc*.) through Google's PageRank algorithm (see Ciaramita and Baroni, 2006: 145); reported frequency counts are often inaccurate. Furthermore, pages from the same source (or even repetitions of text within the same page) are often retrieved by a single search, and wildcard searches (a necessity for studies of this kind) automatically find examples of multi-word phrases in the * position as well as single-word items. As Sharoff (2006: 64) notes:

> Google is a poor concordancer. It provides only limited context for results of queries, cannot be used for linguistically complex queries, such as searching for lemmas (as opposed to word forms), restricting the POS or specifying the distance between components in the query in less than crude ways.

However, Google's limited searching and concordancing sophistication is a less significant disadvantage for a study of this nature (see also Robb, 2003). The goal here is not to examine the wider context in which the search phrases occur but rather to measure the relative frequency in which individual words occupy particular slots within these phrases. Indeed, this and other drawbacks are heavily counter-balanced by the major advantage of using the web as a corpus: its size. Many of the word-strings that we want to search for are too low in frequency to occur in more conventional corpora. To give one example, the phrase 'male and female alike' appears only once in the 100-million-word British National Corpus but generates an estimated frequency of

over 45,000 on Google.[11] This clearly widens the scope of phrasal searching far beyond that which is possible using more typical corpora. The Internet also provides a 'democratic' representation of both formal and informal styles (see Santini, 2005, for more details about 'Web genre') and allows us to revisit the antonym relation using the most contemporary English available. Furthermore, some of the pitfalls associated with using Internet search engines were avoided by the approach taken to analysing the data. For example, multiple hits from the same website were ignored, thus increasing the likelihood that data would originate from different authors. Also, our results were calculated in proportional terms only, and were, therefore, unable to be skewed by false frequency counts. Finally, pairs of words were only considered canonical if retrieved by one another on two or more occasions in ten or more frames, a practice that minimised the distortion caused by text being duplicated across different sites.

In this study, the output for each string was searched by way of Python software developed by Johan Dahl at Lund University. This allowed for up to 990 contexts (the maximum number posted by Google) to be retrieved for each of the twenty adjectives in each of the fourteen frames.[12] The number of usable contexts was often smaller than 990 because Google's wildcard * allows results of more than one word. The files of sentences were automatically searched (using software developed by Lisa Persson at Sussex University) and sorted according to the word occurring in the wildcard position. This procedure ignored any results in which the wildcard consisted of more than one orthographic word, and tabulated the number of tokens of any word found in the wildcard position within the frame. Words occurring only once in any given frame were counted together as one type and subjected to no further individual analysis. These results were recorded in a spreadsheet so that comparisons across search frames could be made.

Throughout the data collection stage, relevant published standards were followed, namely those cited by Linguists for the Responsible Use of Internet Data,[13] as far as possible. For example, all retrieved contexts were saved into files, and two manual spot-checks of the data were conducted to check for typographical errors and repetitions. These spot-checks (each drawing on a random sample of 100 contexts) found no examples of typographical errors among words retrieved in wildcard position, but identified

[11] The phrase 'female and male alike' does not occur at all in the BNC but generates about 600 Google hits.
[12] Because Google extrapolates the number of hits found for any search, the maximum retrievable number is 990 even if the stated number of hits is significantly higher.
[13] http://www.unc.edu/~lajanda/responsible.html

three examples of repetitions.[14] Unfortunately, it was not possible to cross-check web frequencies with those found in standard corpora because, in such corpora, many of our word-strings would occur with negligible frequency.

4. Results and analysis

This section begins with a close look at the results for one of the twenty seed words, *dull*, before looking more broadly at patterns of co-occurrence along all ten of the semantic scales examined. Through reviewing the results for *dull*, the means of analysis and the thresholds for determining canonicity are further explained.

4.1 Dull: a case study

Listed below are fourteen contexts retrieved for *dull* (one example from the output generated by each of the fourteen searches). The search-phrase appears in bold and the word retrieved in wildcard position is italicised.

(1) I would gladly hear your musings, **dull and *dreary* alike**.
(2) Most young women, ***intelligent* and dull alike**, feel the same way.
(3) **Both dull and *bright*** colors are used in impressionistic paintings.
(4) Senses become **both *acute* and dull** at the same time
(5) The outer surface of the shell may be **either dull or *shiny*.**
(6) You'll probably find this **either *amusing* or dull**, depending on your politics.
(7) Intensity refers to a color's strength **whether dull or *bright***.
(8) Other art meetings, **whether *fun* or dull**, were strained.
(9) The 5,000 sq.km salt lake ranges **from dull to *technicolour*** depending on the weather.
(10) The amethyst surface luster varies **from *glassy* to dull**.
(11) It's **dull versus *bright***, what with bland hues thrown in.
(12) Choose between types of pain: new versus old, ***sharp* versus dull**, local versus radiating.

[14] None of the web-sites appeared to be written by non-native speakers or was composed of the highly stylised use of language associated with chatrooms or other web 'spaces'.

(13) "The Three Sisters" precariously walks the line **between dull and *compelling***.

(14) For me the difference **between *interesting* and dull** is the sincerity of the preacher.

In total, 2,760 contexts[15] were retrieved for *dull* and, as those above indicate, many different words were found in the wildcard position. The next step was therefore to combine the frequencies for the fourteen frames and create a ranked list. The ten most commonly-retrieved adjectives for *dull* are recorded in Table 3, together with their frequency (expressed both in absolute terms and as a percentage of all output) and the number of frames in which each adjective appeared.

Rank	Adjective	Freq.	Percent	Frames
1	Bright	103	3.73	11
2	Dynamic	83	3.01	3
3	Sharp	73	2.64	8
4	Dazzling	60	2.17	2
5	Shiny	50	1.81	8
6	Boring	28	1.01	4
7	Brilliant	22	0.80	5
8	Delightful	21	0.76	1
9	Exciting	19	0.69	6
10	Interesting	19	0.69	6

Table 3: Top ten adjectives retrieved by searches on *dull*

The first thing to note here is that searching for 'seeded' contrastive constructions works very well as a means of retrieving antonymous adjectives in a corpus. Although not every adjective found with *dull* was a possible antonym (e.g., *dreary* in (1)), those found to co-occur repeatedly are clearly the most semantically incompatible. This supports the contention that these constructions are themselves contrastive (Murphy, 2006) and justifies their use as an antonym-discovery methodology (Jones, 2002).

It is no surprise to find that *bright* is the most frequent textual antonym of *dull*. In elicitation experiments (Paradis *et al*., 2006), 56 percent of

[15] A British National Corpus search on those fourteen dull frames yields only two hits.

informants volunteered this word as *dull*'s best 'opposite'. Furthermore, of the five other words suggested by informants in that survey, only *lively* (offered by one of the fifty informants only) does not appear in Table 1. This confirms that there is a high degree of correlation between elicited antonyms and those found to co-occur repeatedly in contrastive constructions on the web. Nevertheless, it should be noted that *bright* accounts for only 3.73 percent of the adjectives placed in opposition to the seed word. That *dull* is able to contrast with a variety of items in text is partly a result of the word's polysemy, with adjectives such *dazzling* and *shiny* reflecting one sense of *dull*, while another sense is mirrored by *exciting* and *interesting*.

Of the ten adjectives listed in Table 3, nine can safely be regarded as semantically contrastive. The only exception is *boring*, which ranks relatively highly because of output exemplified by contexts (15) to (17).

(15) Kidman and Baldwin act well, but Pullman is **both dull and *boring***.
(16) They're **both *boring* and dull** words, and it's no wonder we all mix them up all the time.
(17) If you thought that being a Samaritan would be **either *boring* or dull** then think again!

Because *dull* and *boring* clearly make better candidates for synonymy than antonymy, the contexts in (15) to (17) raise questions about whether the constructions used in this research are truly contrastive. However, any constructional form (see, for example, Goldberg, 1995), and in particular co-ordinating constructions like these (Haspelmath, forthcoming), may be associated with more than one meaning. Thus, not every instance of 'both *X* and *Y*' carries inherently contrastive semantics. Also, it should be noted that non-antonymous pairings are comparatively rare in these contexts: among the ten adjectives retrieved most often by *dull*, only *boring* is not contrastive, and this adjective was found in only four of the fourteen frames.

4.2 Canonicity criteria

The results generated by searches on *dull* show that, while frequency of co-occurrence in contrastive constructions may indicate canonical antonymy, *breadth* of co-occurrence is a more reliable diagnostic. For example, in terms of raw frequency, *bright* and *dynamic* were placed in opposition against *dull* at relatively similar rates (103 hits and eighty-three hits respectively). However, in terms of the number of frames in which the words co-occurred, the difference was much greater: *bright* contrasted with *dull* in eleven of the

fourteen frames; *dynamic* in only three. As canonical pairs are paradigmatically related, not just related as co-members of a particular phrase (or, indeed, a particular group of phrases that express a single type of function, such as co-ordination), we took the view that a threshold was necessary. The term 'canonical' was therefore reserved for word pairs that were found to retrieve one another on at least two occasions in at least ten of the fourteen frames. Though results confirm that canonicity operates along a continuum (because some pairs retrieve one another with greater fidelity than others), this threshold was introduced because, (a) it is reasonable to expect that any strong paradigmatic relation will manifest itself in a wide range of appropriate frames, (b) the impact of noise caused by homologous but non-contrastive constructions (as exemplified in examples (15) to (17)) would be reduced, and (c) idiomatic expressions and fixed contexts would be less likely to skew distributions. To give an extreme example of the issue mentioned last, our searches found that the third most common 'opposite' of *rich* was *roach*, accounting for 0.81 percent of all usage. However, these words co-occurred in one of the fourteen frames only, and this, we discovered, was a consequence of a recent album entitled, *Rich versus Roach*. Requiring co-occurrence across a large number of frames reduces the chance that such examples would distort findings.

All of the oppositions that met the ten-frame canonicity threshold are recorded in Table 4, which lists each retrieved adjective according to the proportion of all relevant output that it accounted for (so *large* → *small* tops the list because *small* appeared in 78.76 percent of the output generated by searches on *large*). Also recorded is the number of frames in which each pair co-occurred and the total number of contexts identified (fourteen and 4,361 respectively in the case of *large* → *small*). The final column indicates whether the relation is reciprocal (i.e., whether the retrieved antonym itself retrieves its seed word in ten frames or more), and provides details for those words that were not part of the original study. For example, *lean* was not used as a seed word, but subsequent searches showed that it does reciprocate its antonymy with *fat*, co-occurring in twelve frames and accounting for 8.61 percent of the output. All twenty of the initial seed words retrieved at least one adjective often enough for the pairing to be deemed canonical. Two of the seed words each retrieved three adjectives: *bright* (*dark*, *dim* and *dull*) and *fat* (*thin*, *lean* and *skinny*). A further five seed words each retrieved two antonyms (*narrow*, *open*, *poor*, *small* and *thin*); while the remaining thirteen seed words retrieved one adjective only.

In terms of assessing canonicity, the next step was to discount those pairs found to be in a non-reciprocal relationship. The failure of adjective *X* to retrieve adjective *Y* as often as *Y* retrieves *X* is indicative of asymmetry within the relation. *Y*'s antonymy is unrequited either because *Y* shows a stronger

preference for a third adjective (as *small* favours *large* more than *big*) or because *Y* contrasts more promiscuously with a wider range of contrast items (thereby increasing competition for the wildcard slot in each search). Of the ten pairs we began with, two were discounted because they failed to meet this criterion: *rapid*/*slow* and *boring*/*exciting* (as *slow* retrieves *rapid* in seven frames only, and *boring* retrieves *exciting* in nine). This is consistent with the results of elicitation tests because, for example, 94 percent of informants offered *slow* as the 'opposite' of *rapid*, but none offered *rapid* when given *slow* as a stimulus.

	Seed word		Retrieved adjective	Per-cent	Frames	Contexts	Reciprocal?
1	**large**	→	small	**78.76**	14	4361	Y
2	**rich**	→	poor	**67.94**	13	4209	Y
3	**closed**	→	open	**57.13**	12	2271	Y
4	**small**	→	large	**53.55**	14	3001	Y
5	**weak**	→	strong	**48.41**	13	2019	Y
6	**poor**	→	rich	**44.02**	14	2193	Y
7	**slow**	→	*fast*	**43.65**	13	1625	Y (**50.00**; 13; 1781)
8	**open**	→	closed	**37.45**	10	2240	Y
9	**strong**	→	weak	**36.06**	12	1504	Y
10	**narrow**	→	wide	**34.76**	13	918	Y
11	**thin**	→	*thick*	**33.60**	14	994	Y (**69.72**; 13; 2229)
12	**bright**	→	*dark*	**27.02**	12	861	Y (**4.24**; 10; 186)
13	**wide**	→	narrow	**26.04**	13	887	Y
14	**narrow**	→	*broad*	**17.42**	11	460	Y (**21.71**; 12; 705)
15	**rapid**	→	slow	**12.99**	10	346	N (**5.24**; 7; 195)
16	**ugly**	→	beautiful	**10.95**	14	323	Y
17	**beautiful**	→	ugly	**10.87**	14	374	Y
18	**thin**	→	fat	**9.13**	11	270	Y
19	**small**	→	*big*	**8.87**	12	497	Y (**53.64**; 13; 2856)
20	**bright**	→	*dim*	**8.25**	11	263	Y (**27.73**; 13; 475)
21	**open**	→	*laparoscopic*	**7.56**	10	452	Y (**59.98**; 13; 1175)
22	**fat**	→	thin	**5.65**	11	246	Y
23	**fat**	→	*lean*	**3.79**	10	165	Y (**8.61**; 12; 210)
24	**dull**	→	bright	**3.73**	11	103	Y
25	**poor**	→	*wealthy*	**3.27**	10	163	Y (**37.88**; 11; 899)
26	**bright**	→	dull	**3.11**	11	99	Y
27	**exciting**	→	boring	**2.29**	10	54	N (**1.53**; 9; 63)
28	**fat**	→	*skinny*	**1.63**	11	71	Y (**13.15**; 11; 88)
29	**boring**	→	*interesting*	**1.53**	12	63	N (**1.69**; 7; 53)

Table 4: Ranked list of adjectives retrieved by seed word in ten frames or more

The eleven italicised adjectives in Table 4 were not part of the original search list, but were subjected to post-hoc searches in order to determine

whether their relation with the seed word was reciprocal. As the righthand column of Table 4 shows, ten of the eleven new pairings were indeed found to be reciprocal. For example, *thick* was the adjective retrieved most commonly by *thin*, and Table 4 shows that relation is indeed reciprocal, as *thick* retrieves *thin* in thirteen of the fourteen frames and in 69.72 percent of all output. The only newly-identified pair found not to be in a reciprocal relation was *interesting*/*boring*. *Interesting* was retrieved by *boring* in twelve frames, but only able to reciprocate in seven. Discarding pairs such as *boring*/*exciting* and *boring*/*interesting* raises the question of whether the ten-frame threshold was set too high, especially as many antonyms, especially morphological pairs, show a strong preference towards one particular sequence in text (see Jones, 2002: 120–37). However, the high threshold was necessary in order to ensure that the range of frames returned included some non-coordinated frames (see Table 1). It was important to include a range of discourse functions among the frames because the co-ordinated frames were more likely to return synonyms (e.g., *dull*/*boring*) than the other frames.

A secondary reason for conducting post-hoc searches on the eleven new adjectives was to flag up any further possible canonical pairings that might operate along each scale. In the case of three of the eleven adjectives, the searches were successful in identifying new potential pairings: *broad* retrieved *specific* in ten frames; *lean* retrieved *rich* in eleven; and *big* retrieved *little* in thirteen. Subsequently, the first two pairs were deemed non-canonical because the relation was non-reciprocal. In other words, *specific* failed to retrieve *broad* in ten frames or more and, similarly, *rich* failed to retrieve *lean*. However, the 'post post hoc' searches on *little* showed that this adjective does indeed hold a reciprocal relation with *big* (thirteen frames; 22.03 percent of output).[16] This pair was therefore added to the list of canonical antonyms, all of which are recorded in Table 5. Table 5 raises some key questions about antonym canonicity and our mechanism for identifying it, especially in regard to one of the pairs: *laparoscopic*/*open*. The next section discusses this pairing in closer detail and considers the implications of the findings presented above.

5. Rethinking canonicity

While most of the pairs recorded in Table 4 are familiar antonyms, *laparoscopic*/*open* – a pair describing two types of surgery, one less invasive than the other – is much less commonplace. When used as a seed word,

[16] The further searches conducted for *little* identified no other possible antonyms (*large* was retrieved by *little* in seven frames only and accounted for only 1.97 percent of all hits).

laparoscopic retrieved *open* in thirteen of the fourteen frames and accounted for nearly 60 percent of all output (only three adjectives retrieved any antonym

Scale	Canonical pair(s)
BEAUTY	*beautiful* / *ugly*
WEALTH	*poor* / *rich*, *poor* / *wealthy*
OPENNESS	*closed* / *open*, *laparoscopic* / *open*
SIZE	*large* / *small*, *big* / *small*, *big* / *little*
SPEED	*fast* / *slow*
INTERESTINGNESS	[no canonical pairs identified]
STRENGTH	*strong* / *weak*
WIDTH	*narrow* / *wide*, *broad* / *narrow*
THICKNESS/FATNESS	*thick* / *thin*, *fat* / *thin*, *fat* / *skinny*, *fat* / *lean*
LUMINOSITY	*bright* / *dull*, *bright* / *dim*, *bright* / *dark*

Table 5: Full list of canonical pairs identified in this study

at a higher rate in this entire study, as Table 4 shows). However, it is perhaps more surprising still that, when the same frames were seeded with *open*, *laparoscopic* accounted for as much as 7.56 percent of the output – far more than more conventional contrast words such as *enclosed* (1.69 percent) and *secret* (0.70 percent).[17] Three examples from the data are given below, the last of which includes two repetitions of the phrase *both open and laparoscopic* (counted only once in calculations).

(18) The surgery has moved **from open to laparoscopic**.

(19) These uncommon but potentially serious complications may occur after **either open or laparoscopic** techniques.

(20) Combining the telescopic surgical device with the automatic fluid control system of the present invention will result in making the

[17] Among the other adjectives retrieved by *open* were *arthroscopic* (fifth most common) and *endovascular* (sixth most common), both of which also contrast with *open* with reference to surgery.

> telescopic surgical device a multi functional hand piece thereby enabling it to handle **both open and laparoscopic** electrosurgery, **both open and laparoscopic** argon beam coagulation, and suction/irrigation for **both open and laparoscopic** procedures.

We argue that their repeated co-occurrence within a large proportion of antonymic frames make *laparoscopic*/*open* a very strong candidate to be regarded as a canonical pair, even if most English speakers would be very unlikely to volunteer it in an elicitation test.

Indeed, this pair provides evidence that those 'opposites' intuitively favoured in artificial experiments are not necessarily the same as those that are coupled most reliably in naturally-occurring language. Context-free elicitation and judgement tests reflect both frequency and associative strength because informants are attempting to supply familiar opposites. Therefore, the more well-known the pair, the more often it will be volunteered. However, antonyms operating in more restricted contexts (such as *laparoscopic*/*open* or, to use Murphy's (2003: 178) morphological illustration, *derivational*/ *inflectional*) may well be stronger in terms of their opposability, even though they are less widely known. This may be compared to the problem of *odd number* in testing prototype notions. Research shows that people offer the same examples of an odd number – 3, 5, 7 – not because they are 'more odd' but because they are more familiar, generalisable and approachable examples (Armstrong *et al*., 1983). The results presented here confirm that strength of association is separable from plain word/sense frequency and that less common pairs may be equally canonical within their particular register/jargon as the everyday pairs identified in elicitation tests.

Indeed, the fact that *open* retrieved both *laparoscopic* and *closed* as canonical antonyms highlights another weakness of the elicitation method. Not only do subjects tend to think of high-frequency senses of the stimulus words, they are also usually asked to provide one antonym only. Even if allowed to give multiple responses, the first response may block access to other candidates (as a first response of *fat* to the stimulus *thin* may prime an informant to think of *thin* in its gestalt size sense, and thus block its one-dimensional opposition to *thick*). For words that are polysemous, this means that elicitation reveals only a fraction of what subjects know about the stimulus words.

Not all cases of multiple pairings in Table 5 are explicable in terms of polysemy, however. For instance, *narrow* is canonically opposed to both *broad* and *wide*, but does not have two different senses meaning 'not broad' and 'not wide'. Indeed, some adjectives yield multiple antonyms both in text and across subjects in elicitation tests. This is illustrated in Table 6, which compares those words offered as 'opposites' of *bright* by informants (Paradis *et al*., 2006) with those retrieved by *bright* in the present study

Intuitive antonyms of *bright*	Textual antonyms of *bright*
1. *dark* (42 percent)	1. *dark* (27 percent)
2. *dull* (28 percent)	2. *dim* (8 percent)
3. *dim* (14 percent)	3. *dull* (3 percent)
others (16 percent)	others (62 percent)

Table 6: A comparison of the intuitive and textual antonyms of *bright*

The three canonical antonyms of *bright*, as identified in this study, are the same as the three 'opposites' that ranked most highly in the elicitation experiment. The two main differences between the results are that, (a) the seeded-construction corpus method identifies oppositions at lower proportions because of the greater breadth of output generated by the searches (comprised of both non-contrastive 'noise' and lower frequency oppositions), and (b) *dull* and *dim* are ranked differently in the elicitation and corpus studies. The reason for the latter discrepancy is difficult to pinpoint, but it is possible that dialectal variation is partly responsible. The web data was mostly American, but the informants used in the elicitation tests were mostly British. Regardless, the fact that *bright* generates multiple antonyms according to both methods suggests that it is indeed possible for a single word to have more than one antonym, and that antonym canonicity does not demand exclusivity.

The issue of multiple antonyms is also raised by output generated within the SIZE scale. This scale has been investigated in several studies, including Justeson and Katz (1991), who cite the instinct of most English speakers to pair *large*/*small* and *big*/*little* (but not *large*/*little*) as evidence that antonymy is a lexical as well as a semantic relation, and by Muehleisen (1997) who argues that these pairings are a consequence of the differing collocation profiles of each adjective. Figure 1 quantifies the relationships between the four key adjectives that operate along the scale of size according to their antonym preferences in text.

Figure 1 shows that the relations holding between key adjectives on the SIZE scale are complex and asymmetrical. For example, although the favoured antonym of *little* is *big*, this favouritism is not reciprocated because *big* retrieves *small* in a higher proportion of the output than it retrieves *little* (54 percent as opposed to 12 percent). Furthermore, the evidence shows that, just as *little* is not the most common antonym of *big*, *big* is not the most common antonym of *small*. In fact, *small* shows a clear preference for *large* (54 percent as opposed to 9 percent). This preference, unlike any other on the scale, *is*

reciprocated and, therefore, this pair has the strongest claim to being the most robustly canonical on the scale. As previous studies have indicated (e.g., Muehleisen, 1997), the antonymy holding between *large* and *little* is, conversely, particularly weak, with neither word retrieving the other in more than 2 percent of contexts.

These findings emphasise the importance of reciprocity to canonicity. The two pairings within the scale (*big*/*little* and *large*/*small*) that are considered most lexically parallel (Justeson and Katz, 1991) and most collocationally compatible (Mueleisen, 1997) are the two pairings which retrieve one another at the closest rates. The findings also confirm that elicitation tests are not always an accurate reflection of antonym usage. For example, when asked to supply the best 'opposite' of *small*, 68 percent of informants suggested *big* (and 30 percent suggested *large*). However, in the present study, *small* retrieved *big* in fewer than 9 percent of contexts (and retrieved *large* in over 53 percent).[18]

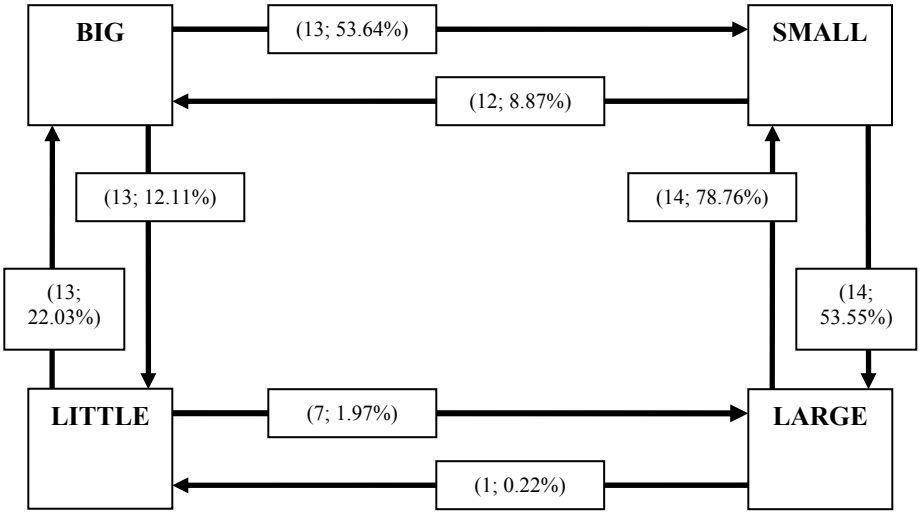

Figure 1: Strength of antonym relations on the SIZE scale (number of frames; percentage of all output)

[18] The tendency of *small* to retrieve *large* in this study is not a result of the corpus being skewed by clothing sizes in retail/catalogue web-sites; the nouns most commonly taking these adjectives are actually corporate in nature: *business*, *company*, *organisation*, etc.

One semantic scale that produced unexpected results was that of BEAUTY. *Ugly* generated a predictable set of potential antonyms; *beautiful*, *pretty*, *cute* and *handsome* were all among the top ten adjectives, though only *beautiful* appeared in ten frames or more. However, the output for *beautiful* was not comprised exclusively of words relating to unattractiveness. This is shown in Table 7, which lists all adjectives retrieved by *beautiful* in 0.5 percent of contexts or more.

	Frequency	Percent	Frames
ugly	374	10.87	14
functional	82	2.38	3
useful	62	1.80	5
practical	44	1.28	4
bizarre	35	1.02	5
durable	32	0.93	2
hideous	19	0.55	7
plain	18	0.52	3

Table 7: Adjectives retrieved most frequently by searches on *beautiful*

Apart from *ugly*, only two adjectives in Table 7 refer to a lack of beauty – *hideous* and *plain* – and they were both returned in a very low proportion of the output. Instead, the searches generated a series of words – *durable*, *functional*, *practical* and *useful* – with explicitly utilitarian qualities. Although it is possible for these adjectives to complement *beautiful* in non-contrastive contexts (like examples (16) to (18), in which *dull* and *boring* co-occurred), the output was mostly comprised of contexts in which the writer sought to correct a relevant presupposition:

(21) Most travel guides are **either beautiful or practical**; this one is both.

(22) Concrete can now be **both beautiful and durable**!

(23) Ornamental and Retention Ponds: Bridging the Gap **Between Functional and Beautiful**.

In contexts (22) to (24), the role of the antonym construction serves to negate the presumption that – when it comes to travel guides, concrete and ponds – beauty is incompatible with functionality. Therefore, even though *durable*,

functional, *practical* and *useful* are not listed as antonyms of *beautiful* by current dictionaries and thesauri, their repeated co-occurrence suggests that they are valid contrast items (though they are still far from being canonical antonyms). Utility, as well as ugliness, is perceived as being opposed to beauty in contemporary English-speaking culture.

6. Conclusion

The research presented here supports the view that contrastive word-pairs 'may be more or less antonymous rather than antonymous or not antonymous' (Justeson and Katz, 1991: 147) and has explored the potential for such relations to be identified using corpus methods and for the strength of their relation to be quantified. To begin with the question of whether a series of lexico-grammatical constructions can be used as an accurate diagnostic of antonymy and, therefore, as a reliable indicator of canonicity, our results suggest that the methodology tested was indeed highly appropriate. The seven constructions used in this research were successful in retrieving a range of contrast items for each seed word, and a strong correlation emerged between those items retrieved most frequently and those adjectives cited as 'good opposites' in elicitation experiments. Indeed, as the summary presented in Table 8 shows, in the case of nine of the ten words randomly chosen as a starting point for this research, the adjective retrieved most often in searches was the same as the adjective intuitively paired with the seed word by the highest proportion of informants.

Although some of the web-searched antonyms were retrieved at extremely low proportions (*exciting* retrieved *boring* in only 2.29 percent of output), these proportions were still higher than any other adjective identified and so remain indicative of the reliability of the constructions used. Only *thin* did not retrieve its intuitive antonym (*fat*) most frequently in this study, but the antonym that was retrieved most commonly (*thick*) ranked second in the elicitation experiment (and *fat* ranked second in the web search). Therefore, it can be reasonably concluded that the lexico-grammatical constructions used here are, collectively, an excellent diagnostic of the antonym relation, in that they tend to include co-occurring antonyms.

This research has also succeeded in its aim to shed further light on the antonym relation and on the phenomenon of canonicity itself. Researchers such as Charles and Miller (1989) treat co-occurrence as a cause of antonymy. However, this paper has shown that it can also be seen as a key symptom and used accordingly to gauge the strength of the antonym relation. The ways in which antonyms co-occur in text go beyond collocation, and we conclude that repeated co-occurrence across a wide range of antonym frames is a better

indicator of canonicity than either raw frequency counting or metalinguistic experimentation.

Seed word	Top textual antonym (percent)		Top intuitive antonym (percent)	
beautiful	*ugly*	(11)	*ugly*	(100)
poor	*rich*	(44)	*rich*	(100)
open	*closed*	(38)	*closed*	(80)
large	*small*	(79)	*small*	(96)
rapid	*slow*	(1)	*slow*	(94)
exciting	*boring*	(2)	*boring*	(72)
strong	*weak*	(36)	*weak*	(94)
wide	*narrow*	(26)	*narrow*	(90)
thin	*thick*	(9)	*fat*	(70)
dull	*bright*	(4)	*bright*	(56)

Table 8: Comparison of textual and intuitive antonyms for ten initial seed words

The final aim of this paper was to evaluate whether the web is a suitable corpus for research of this kind. The benefits are self-evident: even the largest of corpora currently available could not be used to draw meaningful conclusions about the tendency of low frequency antonyms to co-occur in low-frequency constructions. Had this methodology been applied to a conventional corpus, the canonicity threshold (at least two hits in at least ten of the fourteen frames) may not have been reached by any pair of words. For example, in the BNC, neither of the two most canonical pairs identified in this study meet the threshold: *large*/*small* retrieve one another in six of the fourteen frames only, and *poor*/*rich* in seven. Of course, this threshold was self-determined and could, therefore, have been lowered, but this would have compromised the reliability of the findings considerably, especially if the antonym co-occurrence could not be shown to cross different types of antonymic constructions and serve different discourse functions. Nevertheless, the disadvantages of using web data should not be underestimated. As discussed earlier, Internet search engines are idiosyncratic and limited in their retrieval methods (Ciaramita and Baroni, 2006), and the textual content of the web can be unbalanced, repetitive and unrepresentative (Kilgarriff and Grefenstette, 2003; Fletcher, 2004; Sharoff, 2006; *etc*.).

Indeed, this research is open to improvement and enlargement in several ways. After piloting dozens of potential constructions from those identified in previous corpus-based studies of antonymy (Justeson and Katz, 1991; Mettinger, 1994; Jones, 2002; *etc*.), we settled on the seven that retrieved contrastive items with maximum reliability and minimum 'noise'. However, it may be the case that, as antonym pairs change over time, so too do their favoured lexico-grammatical environments. Other textual constructions may thus need to be incorporated. Ideally, each frame would also be weighted according to the strength of its antonymic association so that a more sophisticated measurement of canonicity could be developed.[19] In terms of further research, the opportunity now arises to compare web-searched antonyms with those suggested by dictionaries or identified by lexical referencing systems such as WordNet. Furthermore, the authors of this paper are currently conducting new research to discover whether the methods used here can successfully retrieve antonyms in languages other than English, which have smaller representation on the web. Therefore, although this paper has succeeded in confirming that the textual behaviour of antonyms is predictable and has demonstrated that patterns of co-occurrence allow for pairings to be identified and levels of canonicity measured, it is no more than a preliminary step towards a fuller understanding of the antonym relation and its function in discourse.

References

Armstrong, S.L., L. Gleitman and H. Gleitman. 1983. 'What some concepts might not be', Cognition 13, pp. 263–308.

Biber, D., S. Conrad and R. Reppen. 1998. Corpus Linguistics: Investigating Language Structure and Conceptual Interpretation. Berlin: Springer Verlag.

Charles, W.G. and G.A. Miller. 1989. 'Contexts of antonymous adjectives', Applied Psycholinguistics 10, pp. 357–75.

Charles, W.G., M.A. Reed and D. Derryberry. 1994. 'Conceptual and associative processing in antonymy and synonymy', Applied Psycholinguistics 15, pp. 329–54.

[19] For example, a more advanced mechanism for identifying canonicity might place less emphasis on reciprocity, especially as the preference shown by some pairs for a particular sequence within a frame (e.g., *neither confirm nor deny*) can be so strong as to border on idiomaticity (Jones, 2002: 120–37).

Ciaramita, M. and M. Baroni. 2006. 'Measuring web corpus randomness: a progress report' in M. Baroni and S. Bernardini (eds) Wacky! Working Papers on the Web as Corpus, pp. 127–158. Bologna: GEDIT.

Clark, H.H. 1970. 'Word associations and linguistic theories' in J. Lyons (ed.) New Horizons in Linguistics, pp. 271–86. Baltimore: Penguin.

Deese, J. 1965. The Structure of Associations in Language and Thought. Baltimore: Johns Hopkins University Press.

Fellbaum, C. 1995 'Co-occurrence and antonymy', International Journal of Lexicography 8, pp. 281–303.

Fletcher, W. 2004. 'Making the web more useful as a source for linguistic corpora' in T. Upton (ed.) Corpus Linguistics in North America, pp. 191–205. Amsterdam: Rodopi.

Goldberg, A.E. 1995. Constructions: a construction grammar approach to argument structure. Chicago: University of Chicago Press.

Gross, D., U. Fischer and G.A. Miller. 1989. 'The organization of adjectival meanings', Journal of Memory and Language 28, pp. 92–106.

Gross D. and K.J. Miller. 1990. 'Adjectives in WordNet', International Journal of Lexicography 3 (4), pp. 265–77.

Haspelmath, M. Forthcoming. 'Coordination' in T. Shopen (ed.) Language Typology and Syntactic Description (second edition). Cambridge: Cambridge University Press.

Herrmann, D.J., R.J.S. Chaffin, G. Conti, D. Peters and P.H. Robbins. 1979. 'Comprehension of antonymy and the generality of categorization models', Journal of Experimental Psychology: Human Learning and Memory 5, pp. 585–97.

Herrmann, D.J., R.J.S. Chaffin, M.P. Daniel and R.S. Wool. 1986. 'The role of elements of relation definition in antonymy and synonym comprehension', Zeitschrift fur Psychologie 194, pp. 133–53.

Jones, S. 2002. Antonymy: A Corpus-Based Approach. London: Routledge.

Jones, S. and M.L. Murphy. 2005. 'Using corpora to investigate antonym acquisition', International Journal of Corpus Linguistics 10 (3), pp. 401–422.

Jones, S. 2006. 'A lexico-syntactic analysis of antonym co-occurrence in spoken English', Text and Talk 26 (2), pp. 191–216.

Justeson, J.S. and S.M. Katz. 1991. 'Co-occurrences of antonymous adjectives and their contexts', Computational Linguistics 17, pp. 1–19.

Justeson, J.S. and S.M. Katz. 1992. 'Redefining antonymy', Literary and Linguistic Computing 7, pp. 176–84.

Kennedy, G. 1998. An Introduction to Corpus Linguistics. London: Longman.

Kilgarriff, A. and G. Grefenstette. 2003. 'Introduction to the special issue on the Web as corpus', Computational Linguistics 29 (3), pp. 333–47.

Mel'čuk, I. 1996. 'Lexical functions: a tool for the description of lexical relations in a lexicon' in L. Wanner (ed.) Lexical Functions in Lexicography and Natural Language Processing, pp. 37–102. Amsterdam: Benjamins.

Mettinger, A. 1994. Aspects of Semantic Opposition in English. Oxford: Clarendon.

Muehleisen, V. 1997. Antonymy and Semantic Range in English, (unpublished PhD dissertation), University of Illinois.

Murphy, M.L. 2003. Semantic Relations and the Lexicon. Cambridge: Cambridge University Press.

Murphy, M.L. 2006. 'Antonyms as lexical constructions: or, why paradigmatic construction isn't an oxymoron' in D. Schönefeld (ed.) Constructions All Over**:** Case Studies and Theoretical Implications. Special volume of Constructions, SV1–8/2006. Available online at: http://www.constructions-online.de/

Paradis, C. and C. Willners. Forthcoming. 'Antonyms in dictionary entries: methodological aspects', Studia Linguistica.

Paradis, C., C. Willners, M.L. Murphy and S. Jones. Forthcoming. 'Good vs. bad antonyms: how to tell the difference'.

Paradis, C., C. Willners, S. Löhndorf and M.L. Murphy. 2006. 'Quantifying aspects of antonym canonicity in English and Swedish: textual and experimental', presented at Quantitative Investigations in Theoretical Linguistics 2, Osnabrück, Germany, 1–2 June.

Robb, T. 2003. 'Google as a quick 'n' dirty corpus tool', Teaching English as a Second or Foreign Language 7 (2).

Santini, M. 2005. 'Genres in formation? An exploratory study of Web pages using cluster analysis', Proceedings of the 8th Annual Colloquium for the UK Special Interest Group for Computational Linguistics.

Sharoff, S. 2006. 'Creating general-purpose corpora using automatic search engine queries' in M. Baroni and S. Bernardini (eds) Wacky! Working Papers on the Web as Corpus. Bologna: GEDIT.

Willners, C. 2001. Antonyms in Context. Lund: Lund University.