Using Web Data to Investigate Antonym Canonicity

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

1. Introduction

In the literature, some researchers (e.g. Gross, Fischer and Miller 1989, Charles, Reed and Derryberry 1994) treat antonym pairs as 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, Chaffin, Conti, Peters and Robbins 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, Chaffin, Daniel and Wool 1986) and elicitation experiments (Paradis, Willners, Murphy and Jones, forth.). 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). We argue that a language's most canonical antonym pairs can reasonably be expected to co-occur with highest fidelity in such constructions (*fidelity* here refers to the tendency of words to co-occur with each other, in preference to other semantically plausible pairings, across the widest possible range of appropriate contexts) and that, given their relatively low frequency in language, an extremely large corpus is needed in order to identify such patterns. The specific aims of this paper are therefore (a) to assess the degree to which a series of lexico-grammatical constructions can be used as a diagnostic of antonymy; (b) to measure the strength of antonym pairs belonging to ten semantic scales by examining their co-occurrence fidelity within these constructions; and (c) to evaluate the usefulness of the World Wide Web as a corpus for research into certain types of low-frequency phenomena in language.

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,

[1] School of Education, University of Manchester
e-mail: stevejones@manchester.ac.uk
[2] Department of Linguistics and English Language, University of Sussex
e-mail: m.l.murphy@sussex.ac.uk
[3] Centre for Languages and Literature, Lund University
e-mail: carita.paradis@englund.lu.se
[4] Department of Linguistics, Lund University
e-mail: Caroline.Willners@ling.lu.se

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*, with pairs like *night/day* and *good/bad* following. A less direct approach had been previously taken by Deese (1965) and Clark (1970), who used word association tests to tap into intuitions about the relation. In such tests, informants were invited to say or write the first word that comes into their heads upon hearing or reading a stimulus word. Among those words most frequently elicited by one another were *inside/outside* and *right/wrong*, providing evidence that responses were "overwhelmingly contrastive or antonymic to the stimulus" (Deese 1965: 347). However, because judgement tests and elicitation experiments are metalinguistic in 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). However, co-occurrence alone is not a reliable criterion for identifying antonyms because many pairs of words co-occur (e.g. *fish/chips*, *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, in doing so, they tend to serve one of a small number of discourse functions in text (Jones 2002).

2. 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 essentially involved three steps: step one was to identify several grammatical constructions within which antonym pairs are known to co-occur frequently (e.g. *X and Y alike*); step two was to search those frames with a range of single adjectives inserted in either X- or Y-position (e.g. "thin and * alike"); and step three was to examine 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 to distilling 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 Selecting Suitable Frames

The present study reverses the approach taken in many previous studies because instead of searching for antonym pairs in order to identify the phrases in which they co-occur, we searched the phrases 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, Paradis, Willners and Jones, forthcoming). 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. These constructions can be described in the abstract using X and Y as placeholders for the antonymous words—for example *neither X nor Y* and *X rather than Y*.

To make this study manageable, each chosen construction was 'seeded' with one of a range of adjectives in X- or Y-position. For example, the construction *both X and Y*, when seeded with *open*, would involve two searches: "both open and *" and "both * and open". In each case, * is a wildcard which allows for an unspecified string of characters.

When deciding which contrastive constructions would be most appropriate for this study, any constructions that were less than four words long were initially ruled out, as search strings composed of two words and a wildcard often give non-constituent results, a hazard compounded by using a corpus without grammatical tags. For example, a search for *fat* in the *X, not Y* construction 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 involved 11 four-word constructions that were seeded with a variety of adjectives in order to rule out any that generated large amounts of non-constructional 'noise'. The most productive frames were generally found to associate with the functional category of Coordinated Antonymy, for example *X and Y alike, X as well as Y, both X and Y, either X or Y, neither X nor Y, 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). Because the aim was to assess antonym co-occurrence across a range of functional categories, the number of coordinated constructions was limited to four. In addition, we chose *from X to Y, between X and Y* and *X versus Y*. The last-mentioned differs because it is a three-word, rather than a four-word construction. However, as the pilot tests 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, as shown in Table 1.

wildcard-first frame	wildcard-second frame
* *and* Adj *alike*	Adj *and* * *alike*
between * *and* Adj	*between* Adj *and* *
both * *and* Adj	*both* Adj *and* *
either * *or* Adj	*either* Adj *or* *
from * *to* Adj	*from* Adj *to* *
* *versus* Adj	Adj *versus* *
whether * *or* Adj	*whether* Adj *or* *

Table 1. The fourteen search frames used in this study

2.2 Selecting adjectives 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, Willners, Löhndorf and Murphy (2006). This study invited fifty informants to state the 'opposite' of a series of given adjectives.[5] 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 29 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.

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

Table 2: Stimulus and response adjectives (emboldened words used as initial seed words in present study)

[5] The stimuli for that experiment were, in turn, selected from a range of frequently co-occurring adjective pairs in the British National Corpus.

All of the adjectives listed in the first column of Table 2 were found to be strongly, uni-directionally associated with one particular antonym by informants (Paradis et al., 2006).[6] 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 considered 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.

2.3 Selecting a corpus

Piloting a similar approach to the one taken here, Jones demonstrated that the favoured antonyms of *natural* in text are *artificial* and *man-made* (but not *unnatural*), and that *style* tends to be placed in opposition against *substance* most commonly in contemporary English (2002: 154-167). 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 text 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-designed corpus-searching software. For example, web-pages are not selected at random but rather sorted according to extraneous criteria (relevance of topic, popularity of web-site, etc.) via Google's PageRank algorithm (see Ciaramita and Baroni 2006: 145). 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 * position as well as single-word items. As Sharoff 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

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

lemmas (as opposed to word forms), restricting the POS or specifying the distance between components in the query in less than crude ways" (2006: 64).

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 count the frequency in which individual words occupy particular slots within these phrases. Indeed, this and other drawbacks are heavily counter-balanced by the major advantage to using the web as 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.[7] 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, as discussed later, 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, up to 990 contexts (the maximum number posted by Google) were retrieved for each of the 20 adjectives in each of the 14 frames.[8] 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 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.

3. Results

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.

[7] The phrase *female and male alike* does not occur at all in the BNC but generates about 600 Google hits.
[8] Because Google extrapolates the number of hits found for any search, the maximum retrievable number is 990 even if the claimed number of hits found is significantly higher.

4.1 A case study of *dull*

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 5,000sq.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.

(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[9] 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.

[9] A British National Corpus search on those fourteen dull frames yields a total of only two hits.

		freq	%	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
	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 to retrieve 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% of informants identified this word *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% 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)-(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, contexts (16) to (18) 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 coordinating constructions like these (Haspelmath, forth.), 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 rogue pairings are comparatively rare in these contexts: among the ten adjectives retrieved most often by

dull, only *boring* is clearly non-antonymous, and this adjective was found in only four of the fourteen frames.

3.2 Towards a New Definition of Canonicity

The results generated by searches on *dull* show that, while frequency of co-occurrence in contrastive constructions may be indicative of 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 83 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 11 of the 14 frames; *dynamic* in only 3). 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 coordination), 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), a 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 (15) – (17), would be reduced, and (c) idiomatic expressions and fixed contexts would be less likely to skew findings. To give an extreme example of last-mentioned issue, our searches found that the third most common 'opposite' of *rich* was *roach*, accounting for 0.81% 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 skew the results.

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 *small* tops the list because it appeared in 78.76% of the contexts identified in the fourteen searches for *large*). Also recorded is the number of frames in which each pair co-occurred and the total number of contexts identified (14 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% of the output.

	seed word		retrieved adjective	%	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: Adjectives retrieved by seed word in ten frames or more

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 uni-directional, not a bi-directional, relationship. The failure of adjective A to retrieve adjective B as often as B retrieves A is indicative of asymmetry within the relation. A's antonymy is unrequited either because B shows a stronger preference for a third adjective (as *small* favours *large* more than *big*) or because B 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% of informants offered *slow* as the 'opposite' of *rapid*, but none offered *rapid* when given *slow* as a stimulus.

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 bi-directional. As the rightmost column of Table 4 shows, ten of the eleven new pairings were indeed found to be bi-directional. 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% of all contexts. The only newly identified adjective found not to be bi-directional was *interesting*, 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: 120-137). 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 coordinated 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. The first two pairs were subsequently deemed non-canonical because the relation was uni-directional. 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* (13 frames; 22.03% of contexts).[10] This pair was therefore added to the list of canonical antonyms, all of which are recorded in Table 5.

[10] 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% of all hits).

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: canonical pairs identified in this study

While most of the pairs recorded in Table 5 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, *laparoscopic* retrieved *open* in thirteen of the fourteen frames and accounted for nearly 60% of all contexts (only three adjectives retrieved any antonym 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% of the output, far more than more conventional contrast words such as *enclosed* (1.69%) and *secret* (0.70%).[11] 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 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.

[11] Among the other adjectives retrieved by *open* were *arthroscopic* (5th most common) and *endovascular* (6th), both of which also contrast with *open* with reference to surgery.

We argue that their repeated co-occurrence within a large proportion of antonymic frames make *laparoscopic*/*open* a very strong candidate to be considered 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 extensively 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 give the same examples of *odd number*—3, 5, 7—not because they are 'more odd' but because they are more familiar, generalisable and approachable examples (Armstrong, Gleitman and Gleitman, 1985). The results presented here confirm that strength of association is separable from plain word/sense frequency and that less common pairs may be as 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 with *thick*). For words that are polysemous, this means that elicitation reveals only a fraction of what subjects know about the stimulus words.

4. Conclusions

This research 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, results indicate 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 6 shows, in the case of nine of the ten words randomly chosen as a starting point for this research, the adjective retrieved most commonly was the same as that intuitively paired with the seed word by the highest proportion of informants.

seed word	top textual antonym	top intuitive antonym
beautiful	ugly (11%)	ugly (100%)
poor	rich (44%)	rich (100%)
open	closed (38%)	closed (80%)
large	small (79%)	small (96%)
rapid	slow (13%)	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 6: comparison of textual and intuitive antonyms

Although some of the web-searched antonyms were retrieved at extremely low proportions (*exciting* retrieved *boring* in only 2.29% of contexts), these proportions were still higher than any other adjective identified and therefore 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 safely be 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. Commentators 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.

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 currently available corpora 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 functions. Nevertheless, the disadvantages to 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 therefore 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.[12] 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. Additionally, 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: 263–308.

Biber D., S. Conrad and R. Reppen (1998) Corpus Linguistics: investigating language structure and conceptual interpretation. Berlin: Springer Verlag.

Ciaramita M. and M. Baroni (2006) Measuring Web Corpus Randomness: a progress report. In Baroni, M. and S. Bernardini (eds.) 2006. *Wacky! Working papers on the Web as Corpus*. Bologna: GEDIT.

Charles W.G. and G.A. Miller (1989) Contexts of antonymous adjectives. *Applied Psycholinguistics*. 10: 357–75.

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

Clark H.H. (1970) Word Associations and Linguistic Theories. In J. Lyons (ed.) New Horizons in Linguistics. Baltimore: Penguin 271–86.

Cruse D.A. (1986) Lexical Semantics. Cambridge University Press, Cambridge.

Cruse D.A. (1994) Prototype Theory and Lexical Relations. *Rivista di Linguistica*. 6: 167–88.

Deese J. (1965) The structure of associations in language and thought. Baltimore: Johns Hopkins University Press.

[12] For example, a more advanced mechanism for identifying canonicity might place less emphasis on bi-directionality, 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: 120-137)

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

Fletcher B. (2004) Making the Web more Useful as a source for Linguistic Corpora. In T. Upton (ed.) Corpus Linguistics in North America. Rodopi, Amsterdam.

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: 92–106.

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

Haspelmath, M. (forth.) Coordination. In T. Shopen (ed.), Language typology and syntactic description (2nd edn). 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: 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: 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): 401–422.

Jones, S. (2006) A lexico-syntactic analysis of antonym co-occurrence in spoken English. *Text and Talk 2*6:2. 191-216.

Justeson J.S. and S.M. Katz (1991) Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics*. 17: 1–19.

Justeson J.S. and S.M. Katz (1992) Redefining antonymy. *Literary and Linguistic Computing*. 7: 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), 333–47.

Lehrer A. (2002) Paradigmatic relations of exclusion and opposition I: gradable antonymy and complementarity. In Cruse D.A., F. Hundsnurscher, M. Job and P.R. Lutzeier (eds.), Handbook of Lexicology. Berlin: de Gruyter, 498–508.

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

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 Schönefeld, D. (ed.) Constructions all over**:** case studies and theoretical implications. Special volume of Constructions, SV1-8/2006. http://www.constructions-online.de/

Paradis, C., C. Willners, M.L. Murphy and S. Jones (forth.). 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 Baroni, M. and S. Bernardini (eds.) 2006. *Wacky! Working papers on the Web as Corpus.* Bologna: GEDIT.

Willners C. (2001) Antonyms in context. Lund: Lund University.