

## What a Corpus-based Dictionary Tells us about Antonymy<sup>1</sup>

Carita Paradis & Caroline Willners

carita.paradis@vxu.se

caroline.willners@ling.lu.se

School of Humanities

Växjö University

351 95 Växjö

Sweden

### Abstract

This paper investigates the treatment of antonymy in *Collins COBUILD Advanced Learner's English Dictionary* (2003) in order to find out what kinds of headwords are provided with antonyms as part of their definitions and also discusses the principles for antonym inclusion in the entries. CCALED includes canonical antonyms such as *good/bad* and *dead/alive*, as well as more contextually restricted pairings such as *hot/mild* and *flat/fizzy*. The vast majority of the antonymic pairings in the dictionary are adjectives. Most of the antonyms are morphologically different from the headwords they define and typically do not involve antonymic affixes such as *non-*, *un-* or *-less*. Only one-third of the total number of pairs are given in both directions. The principles according to which antonyms are included in CCALED are not transparent to us.

### 1 Introduction

Dictionaries in general and learners' dictionaries in particular are important tools in the process of acquiring foreign languages. We take it for granted that the main goal of a corpus-based learner's dictionary is to provide learners with relevant, idiomatic and useful information that will help them setting up native-like links between words and meanings. It is natural to think that lexicographers are keen to include corpus information about lexico-semantic relations such as synonyms, antonyms, hyponyms and superordinates in pursuit of this goal.

This paper explores the use of antonyms in the definitions of headwords in the 4th edition of *Collins COBUILD Advanced Learner's English Dictionary* (Sinclair et al. 2003), henceforth CCALED. It raises the question of what the principled basis for antonym inclusion is, could or should be. The term *antonym* in this study is equivalent to 'opposite' as defined by the dictionary.<sup>2</sup> Three questions are central to the study of CCALED. They are:

---

<sup>1</sup> We would like to thank Anna Nilsson-Drake for retrieving all the antonyms manually from the dictionary and Lynne Murphy and Steven Jones for comments on an earlier version of this paper.

<sup>2</sup> For various definitions and studies of antonymy see Lehrer & Lehrer 1982, Muchleisen 1997; Paradis 2001, Willners 2001; Jones, 2002, Murphy 2003, Croft & Cruse 2004, Paradis & Willners 2006 as well as <http://www.f.waseda.jp/vicky/complexica/index.html>.

- (i) What kinds of headwords are provided with antonyms?
- (ii) Do the meanings of the antonymic pairs tell us something about the lexical structure of antonymy in English?
- (iii) What do the antonymic pairings tell us about the lexicographic principles involved in selecting antonyms for inclusion?

There are several reasons for selecting CCALED. The first and most important reason is that CCALED is corpus-based. The dictionary has a comparatively long tradition of approximately two decades of using real text as a basis for the compilation of the dictionary, and the corpus aspect plays an important role in the promotion of the dictionary. It is therefore interesting to see how this might be reflected in the selection and inclusion of antonyms. The dictionary takes pride in making principled use of the gigantic 520-million corpus, the *Bank of English*. The founding Editor-in-Chief John Sinclair points out that 'decisions about which words to include as headwords in the dictionary, which meanings to draw attention to, which phrases to recognize as settled expressions in the language, and many other issues, are directly informed by the *Bank of English*' (CCALED vii-x). It is also stated in the introduction to the dictionary that the corpus information is at the heart of each entry and special software has been developed to help the lexicographers to make decisions about different senses of words, the language of the definitions, the choice of examples and the grammatical information, i.e. the information given in the margins. Furthermore, it is pointed out that the corpus enables the lexicographers to make decisions with confidence and accuracy (2003: ix-x). As dictionary users we take this information to mean that the lexicographers are dealing with lexico-semantic structures such as antonymy in naturally occurring contemporary language in a principled text-informed way. However, since we have not been able to find any explicit information about the principles for selecting and including antonyms, neither in the introduction to CCALED nor in *Looking up*, the manual for the first edition (1987), we set out to examine the choice of antonyms in order to uncover the working methods.

The second reason for choosing CCALED is that it is a learner's dictionary, and learners of languages are eager to learn lexical antonyms in pairs (e.g. *full-empty*, *light-dark*). Antonym drills are common in language learning curricula and knowledge of antonymy is necessary for textual competence (Halliday & Hasan 1976). Therefore, it seems reasonable to assume that lexicographers who compile learners' dictionaries give antonymy special attention. Thirdly, the COBUILD project is couched in the structuralist framework for which lexical relations, both paradigmatic and syntagmatic, between words are foundational for the theoretical approach to meaning. Finally, CCALED has a practical advantage over most other dictionaries in that lexical relations are specified in the margin and therefore easy to spot and retrieve from the book.

The purpose of this paper is thus to give a short description of the treatment of antonyms in this corpus-based dictionary, to raise the question of how lexicographers use or could use a huge corpus as a guide to the selection of antonyms and more generally to encourage a discussion of the nature and structure of antonymy in language.

## 2 Collins COBUILD Advanced Learner's English Dictionary (4th edition)

CCALED contains more than 110 000 words, selected from *The Bank of English*. The meanings and uses of every headword are presented through definitions and real examples from the corpus. The dictionary also gives additional information about synonyms, antonyms, superordinates and grammatical patterns in a separate column. For instance, the meaning of the headword *hazardous* is defined as follows 'something that is hazardous is dangerous, especially to people's health or safety'. The example from the corpus is: *They have no way to dispose of the hazardous waste they produce*. The definition in the separate column says that *hazardous* is an adjective. *Safe* is offered as the antonym of *hazardous* and *dangerous* as its synonym.

In contrast to *hazardous*, there are words that take up more space in the dictionary because they have many senses. *Light* is an example of a word that has more than one headword, each with several senses provided with different antonyms in the margin, *heavy*, *dark*, *deep* and *serious*. Together with the definitions, the example sentences and possible synonyms and grammatical patterns, these antonyms are there to account for the meanings and uses of *light* and to guide learners in their attempts to get a good grasp of the structure of the vocabulary of English as a foreign language.

### 3 Antonyms in CCALED

All in all, we found and investigated 1750 antonym pairs in CCALED. The headwords that have antonyms were examined with respect to (i) what word class they belong to, (ii) what their semantic characteristics are, (iii) what the distribution of affixal antonyms are, (iv) whether both members of the antonym pairs are presented as each other's antonyms and, finally, (v) whether the principles for the choice of antonym pairs are transparent and shed light on the structure of antonymy in the English vocabulary. This section deals with the above five issues in turn.

Firstly, the distribution of headwords with antonyms across word classes is shown in Table 1. Antonyms are most often given for adjectives. More exactly, 1 031 out of the 1 750 (59%) headwords are adjectives. Within the group of adjectives with antonyms, 95% (977 out of 1 031) are gradable, either scalar adjectives such as *big/small* or non-scalar adjectives such as *dead/alive* (Paradis 2001). The remaining 5% (54 out of 1 031) of the adjectives are non-gradable such as *abstract/concrete* and *female/male*.

Word class	Antonym given	%
Adjectives	1 031	59
Nouns	317	19
Verbs	220	13
Others	182	9
Total	1 750	100

Table 1. The distribution of antonyms across word classes

It is hardly surprising that adjectives are the most common headwords for which antonyms are given. The reason is that a large number of adjectives typically denote single properties, whereas many nouns typically signify complex meanings with many properties. Typically antonymous adjectives are thus maximally similar in their meanings but differ in signifying opposite aspects or two directions on the same dimension. For instance, *big* and *small* are both associated with the content domain of SIZE on a SCALE, and *dead* and *alive* are associated with EXISTENCE construed on either side of a BOUNDARY. The conceptual simplicity of the content expressed in combination with a configuration of SCALE or BOUNDARY invokes binary contrast and makes it a prominent mode of construal. It is not equally natural for most non-gradable adjectives to form pairs, since many of them are derived from nouns and thereby inherit complex meaning structures. For instance, what would be a natural antonym of *financial*, *linguistic*, *pictorial* or *dental* from a lexico-semantic point of view? A possibility would of course be lexicalizations with the affix *non-*, which turn meanings into their mirror images 'not being X'. However, the productivity of the *non-* prefix in word formation makes it less useful and less informative in dictionary entries. Furthermore, there are also non-gradable meanings that readily lend themselves to binary contrast. *Abstract/concrete* and *female/male* are examples of such conventionalized lexical binarity. Both pairs indicate how people categorize phenomena in the world and/or how the nature of the world forces us to categorize things accordingly.

With respect to the principles for antonym inclusion in the entries, it deserves to be mentioned that many of the most obvious pairs are included, such as *big/small*, *strong/weak* and *bad/good*. Some of them are given in both directions in a symmetrical fashion such as *strong* for *weak* and *weak* for *strong*. *Small* is given as the antonym of *big* and *large*, while only *large*, but also *major*, are given as antonyms of *small*. Such apparent discrepancies made us wonder whether the corpus is the source of information about there being some kind of stronger relationship between *small* and *large* and *major* than between *small* and *big*. Among other antonyms, *strong* is given *weak* and *slight*. *Weak* returns *strong*, while *slight* does not. We fail to see a clear pattern in the choice of antonyms and the symmetry of presentation and will therefore come back to this problem of reversals and symmetry later in this section.

The nominal meanings can be grouped into abstract and concrete notions. Most of the nouns (71%) denote abstract meanings and the rest (29%) denote concrete meanings. There are abstract pairs such as *victory/defeat*, *advantage/disadvantage*, *aggression/gentleness*, *pessimism/optimism*, *absence/presence* and there are concrete pairs such as *borrower/lender*, *buyer/seller*, *hero/villain*, *highbrow/lowbrow*, *(big) fish/(small) fry* and *dog/bitch*. Again, the majority of the antonymic nouns are associated with simple content structures, which point up binarity, just like most of the adjectives do.

The majority of the verb meanings refer to bounded events and actions, e.g. *accept/reject*, *agree/disagree*, but there are also scalar ones such as *diminish/increase*, *criticize/praise*. Only a few of the verbs have stative meanings, e.g. *hate/love*, *like/dislike* and *dread/look forward to*. The semantic patterns are again similar to the adjectival and nominal meanings in being conceptually simple with a natural tendency to bisect the domain or form opposite poles on a scale. The final category, named 'others' which mainly contain temporal, direc-

tional and locative prepositions and adverbs such as *in/out*, *up/down*, *before/after* has not been given any attention in this study for reasons of space limitation.

Furthermore, we investigated how many of the antonym relations are given in both directions. Of all the antonym relations in the dictionary, only 37% are given in both directions. Examples of pairs that occur in both directions are *dead*↔*alive*, *bad*↔*good*, *broad*↔*narrow*, *clean*↔*dirty*, *dark*↔*light*, *dry*↔*wet*, *hard*↔*soft*, *heavy*↔*light*, *large*↔*small*. But quite unexpectedly, we found *big*→*small* but not *small*→*big* and *little*→*big* but not *big*→*little* and we found only *cheap*→*expensive*, *cordial*→*hostile*, *cruel*→*kind*, *difficult*→*easy*, *dry*→*sweet*, *dusk*→*dawn*, *old*→*new*, *hate*→*love*, *nasty*→*nice*, *dull*→*interesting*, *dull*→*sunny*, *dull*→*sharp*, *false*→*true*, *full*→*empty*, *odd*→*even*, *sad*→*happy*, *safe*→*dangerous*, *short*→*tall*, *profound*→*shallow*, *stale*→*fresh*, *dog*→*bitch* in that order. In our opinion as non-native speakers of English, all these pairs deserve to be reversed. Again, it is not clear to us why these pairings should be helpful for the learner in the above directions only. One reason may be that antonyms are given to disambiguate uses of an entry. For instance, the reason why *interesting* and *sunny* are offered as antonyms of *dull* may be to distinguish the two senses of *dull*, and this disambiguating function is not considered to be necessary for *sunny* and *interesting*. Also, there are cases where the reason for the unidirectionality is understandable. For instance, *underwhelmed*→*overwhelmed* are given in this direction only. The reason is likely to be that *underwhelmed* is a comparatively new coinage, which plays on the relation of antonymy and is possible just because such relations are conventionalized modes of construals. This is also an argument that such contrast relations exist above and beyond words at a more abstract level of relations of thought.

Morphologically derived antonym relations are rarely reversed, but there are differences within this category too. Antonyms of headwords containing the prefix *in-* are reversed in 35% of the cases and *un-* in 16% of the cases, but *non-* entries are never reversed. There is no lexicographical need for *non-*prefixed words to be reversed, since the prefix *non-* operates in a similar fashion to the logical negator, i.e. without any collocational restrictions or constrained interpretations. Intrinsic binarity in a domain opens up for two possibilities only. This is clearly the case for affixed antonyms. *Un-* as a prefix is almost always the opposite of the root that follows *un-*, except for words such as *uneasy* and *uncouth*.

Out of the total number of headwords with antonyms, 638 involve a prefixed word. Apart from truly sublexical prefixes, such as the ones mentioned above, there are also prefixes that are lexical such as *left-click/right-click* and *overground/underground*.

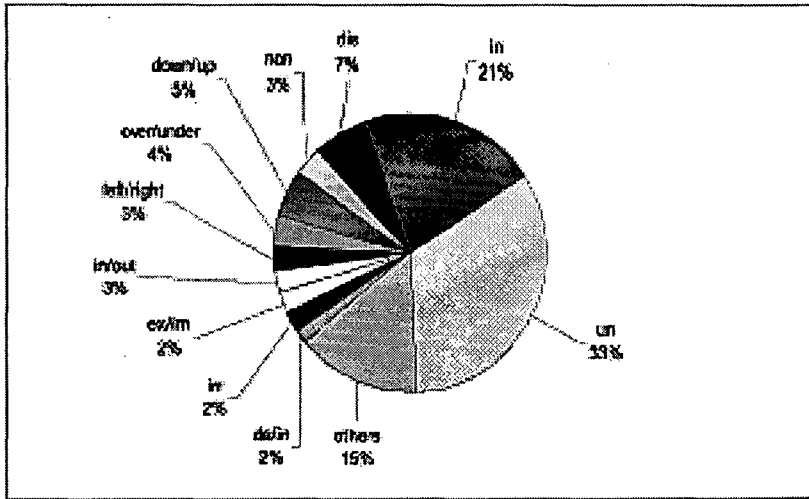


Figure 1. The distribution of prefixes in the entries with antonyms

As Figure 1 shows, the prefix *un-* is the most commonly used prefix with accompanying antonyms – one third of all prefixed antonyms are formed with that prefix. Some antonymous pairs require the attachment of only one prefix to create an opposite meaning: *paid/unpaid*, whereas other pairs demand a prefix for both words, such as *down-river/up-river*, *over-ground/underground*, as Figure 1 shows. Antonyms with suffixes are less common. We found ninety-nine pairs altogether. Seventy-five of the pairs have preposition-like additions, e.g. *check in/check out*, *mark down/mark up*, *stay in/go out* and *turn on/turn off*. Among the other twenty-four pairs, many are of the *-ful/-less* type, e.g. *careful/careless*, *emotionless/emotional*, *joyless/joyous*, *noiseless/noisy*.

## 5 Conclusion

The potential outcome of this investigation is that there are clear similarities across the meanings of the headwords that are defined by antonyms. They are all what we might call inherently binary because they map on to simple content structures and they are construed according to a scale or a boundary. The majority of the headwords with antonyms are adjectives. The principles for what antonyms are included in the dictionary are not transparent to us in spite of the fact that we have scrutinized the dictionary manually from cover to cover. One of the main purposes for a learner’s dictionary such as CCALED is to guide learners in their attempts to get a good grasp of the structure of the vocabulary of English. Being a corpus-based dictionary, lexicographers should make principled use of the huge text corpora they have at their disposal nowadays.<sup>3</sup>

<sup>3</sup> Paradis & Willners (forthcoming) suggest a corpus-based method that could potentially be useful for selecting antonyms for inclusion in dictionaries.

## References

### A. Dictionaries

Sinclair, J. (ed.) (2003), *Collins COBUILD Advance Learner's English Dictionary (4th edition)*. Glasgow, Harper Collins Publishers.

### B. Other literature

*The Comparative Lexical Relations Group*. 24 November 2005. <http://www.f.waseda.jp/vicky/complexical/index.html>

Croft, W., Cruse, A. (2004), *Cognitive linguistics*, Cambridge, Cambridge University Press.

Halliday, M.A.K., Hasan, R. (1976), *Cohesion in English*, Harlow, Longman.

Jones, S. (2002), *Antonymy: a corpus-based perspective*, Amsterdam, Routledge.

Lehrer, A., Lehrer, K. (1982), 'Antonymy.' *Linguistics and philosophy* 5, pp. 483-501.

Muehleisen, V. (1997), *Antonymy and Semantic Range in English*, Evanston, IL, Northwest University. PhD.diss.

Murphy, L. (2003), *Semantic relations and the lexicon: antonyms, synonyms and other semantic paradigms*, Cambridge, Cambridge UP.

Paradis, C. (2001), 'Adjectives and boundedness', *Cognitive Linguistics* 12.1, pp. 47-65.

Paradis, C., Willners, C. (2006), 'Antonymy and negation: the boundedness hypothesis', *Journal of Pragmatics*.

Paradis, C., Willners C. (forthcoming), 'Selecting antonyms for dictionary entries: methodological aspects' in Klingvall, E., Manninen, S. (eds.), *The Department of English: working papers in English linguistics*, vol. VI (<http://www.englund.lu.se/research/workingpapers/index.phtml>).

Sinclair, J. (1987), *Looking up an account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*, London, HarperCollins.

Willners, C. (2001), *Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives*, Travaux de Institut de Linguistique de Lund 40. Lund, Sweden: Dept. of Linguistics, Lund University.