



LUND UNIVERSITY

Swedish opposites - a multi-method approach to antonym canonicity

Willners, Caroline; Paradis, Carita

Published in:

Lexical-semantic relations from theoretical and practical perspectives

2010

[Link to publication](#)

Citation for published version (APA):

Willners, C., & Paradis, C. (2010). Swedish opposites - a multi-method approach to antonym canonicity. In P. Storjohann (Ed.), *Lexical-semantic relations from theoretical and practical perspectives* (Vol. Lingvisticæ Investigationes Supplementa). John Benjamins Publishing Company.

Total number of authors:

2

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Swedish opposites

A multi-method approach to ‘goodness of antonymy’*

Caroline Willners and Carita Paradis

This is an investigation of ‘goodness of antonym pairings’ in Swedish, which seeks answers to why speakers judge antonyms such as *bra-dålig* ‘good-bad’ and *lång-kort* ‘long-short’ to be better antonyms than, say, *dunkel-tydlig* ‘obscure-clear’ and *rask-långsam* ‘speedy-slow’. The investigation has two main aims. The first aim is to provide a description of goodness of Swedish antonym pairings based on three different observational techniques: a corpus-driven study, a judgement experiment and an elicitation experiment. The second aim is to evaluate both converging and diverging results on those three indicators and to discuss them in the light of what the results tell us about antonyms in Swedish, and perhaps more importantly, what they tell us about the nature of antonymy in language and thought more generally.

1 Introduction

In spite of the widespread consensus in the linguistic literature that contrast is fundamental to human thinking and that antonymy as a lexico-semantic relation plays an important role in organising and constraining the vocabularies of languages (Lyons 1977, Cruse 1986, Fellbaum 1998, Murphy 2003), relatively little empirical research has been conducted on antonymy, either using corpus methodologies or experimental techniques. No studies have been conducted using a combination of both methods.

* Thanks to Joost van de Weijer for help with the statistics, to Anders Sjöström for help with producing figures and to Simone Löhndorf for help with data collection.

The general aim of this article is to describe a combination of methods useful in the study of antonym canonicity, to summarise the results and to assess their various advantages and disadvantages for a better understanding of goodness of antonymy as a lexico-semantic construal. By combining methods, we hope to contribute to the knowledge about the nature of antonymy as a relation of binary contrast. A mirror study has been performed for English and is reported on in Paradis et al. (submitted).

Antonyms are at the same time minimally and maximally different from one another. They activate the same conceptual domain, but they occupy opposite poles/parts of that domain. Due to the fact that they are conceptually identical in all respects but one, we perceive them as maximally similar, and, at the same time, due to the fact that they occupy radically different poles/parts, we perceive them as maximally different (Cruse 1986, Willners 2001, Murphy 2003). Words that we intuitively associate with antonymy are adjectivals (Paradis & Willners 2007).

Our approach assumes antonyms, both more strongly canonical and less canonical, to be conceptual in nature. Conceptual knowledge reflects what speakers of languages know about words, and such knowledge includes knowledge about their relations (Murphy 2003: 42-60, Paradis 2003, 2005, Paradis et al. submitted). Treating relations as relations between concepts, rather than relations between lexical items is consistent with a number of facts about the behaviour of relations. Firstly, relations display prototypicality effects, in that there are better and less good relations. In other words, not only is *torr* 'dry' the most salient and well-established antonym of *våt* 'wet', but the relation itself may also be perceived as a better antonym relation than, say, *seg-mör* 'tough-tender'. When asked to give examples of opposites, people most often offer pairs like *bra-dålig* 'good-bad', *svag-stark* 'weak-strong', *svart-vit* 'black-white' and *liten-stor* 'small-large', i.e. common lexical items along salient (canonical) dimensions. Secondly, just like non-linguistic concepts, relations in language are about

Comment [MSOffice1]: Is this page reference needed?

Comment [CW2R1]: Yes, we think it is helpful with page references when referring to books.

construals of similarity, contrast and inclusion. For instance, antonyms may play a role in metonymisation and metaphorisation. At times, new metonymic or metaphorical coinages seem to be triggered by relations. One such example is *slow food* as the opposite of *fast food*. Thirdly, lexical pairs are learnt as pairs or construed as such in the same contexts. Canonicity plays a role in new uses of one of a pair of a salient relation. For a longer introduction to this topic, see Paradis et al. (forthcoming).

The central issue of this paper concerns ‘goodness of antonymy’ and methods to study this. Like Gross & Miller (1990), we assume that there is a small group of strongly antonymic word pairs (Canonical antonyms) that behave differently from other less strong (non-canonical) antonyms. (Direct/indirect and lexical/conceptual are alternative terms for the same dichotomy.) For instance, it is likely that speakers of Swedish would regard *långsam-snabb* ‘slow-fast’ as a good example of canonical antonymy, while *långsam-kvick* ‘slow-quick’, *långsam-rask* ‘slow-rapid’ and *snabb-trög* ‘fast-dull’ are perceived as less good opposites. All these antonymic pairs in turn will be different from unrelated pairs such as *långsam-svart* ‘slow-black’ or synonyms such as *långsam-trög* ‘slow-dull’.

As for their behaviour in text, Justeson & Katz (1991, 1992) and Willners (2001) have shown that antonyms co-occur in the same sentence at higher than chance rates, and that canonical antonyms co-occur more often than non-canonical antonyms and other semantically possible pairings (Willners 2001). These data support the dichotomy view of the Princeton WordNet and Gross & Miller (1990).

The test set used in the present study consists of Swedish word pairs of four different types: Canonical antonyms, Non-canonical antonyms, Synonyms and Unrelated word pairs (see Tables 4 and 5). The words in the Unrelated word pairs are always from the same semantic field but the semantic relation between them is not clear even though they might share certain aspects of meaning, e.g. *het-plötslig* ‘hot-sudden’. Synonyms and

Unrelated word pairs were introduced as control groups. While it is not possible to distinguish the four types using corpus methodologies, we expect significant results when judged for ‘goodness of oppositeness’ experimentally and in the number of unique responses when the individual words are used as stimuli in an elicitation test. All of the word pairs included in the study co-occur in the same sentence significantly more often than chance predicts.

An early study of ‘goodness of antonymy’ is to be found in Herrmann et al. (1979). They assume a scale of canonicity and use a judgement test to obtain a ranking of the word pairs in the test set. We include a translation of a subset of his test items in this study in an attempt to verify or disconfirm his results.

The procedure is as follows. Section 2 discusses some methodological considerations before the methods used are described in detail in following sections. Corpus-driven methods are used to produce the test set (Section 4) that is used in the elicitation experiment (Section 5) and the judgement experiment (Section 6). A general discussion of the results and an assessment of the methods are found in Section 7. Finally, the study is concluded in Section 8. Before going into details about our method and experiments, we give a short overview of previous work relevant to the present study.

2 Methodological considerations

In various previous studies, we explored antonymy using corpus-based as well as corpus-driven approaches¹ (e.g. Willners 2001, Jones et al. 2007,

¹ In current empirical research where corpora are used, a distinction is made between corpus-based and corpus-driven methodologies (Francis 1993, Tognini-Bonelli 2001: 65-100, Storjohann 2005, Paradis & Willners 2007). The distinction is that the corpus-based methodology makes use of the corpus to test hypotheses, expound theories or retrieve real examples, while in corpus-driven methodologies, the corpus serves as the empirical basis from which researchers extract their data with a minimum of prior assumptions. In the latter approach, all claims are made on the basis of the corpus evidence with the necessary

Murphy et al. 2009, Paradis et al.). Corpus data are useful for descriptive studies since they reflect actual language use. They provide a basis for studying language variation, and they also often provide metadata about speakers, genres and settings. Another, very important property of corpus data is that they are verifiable, which is an important requirement for a scientific approach to linguistics.

Through corpus-driven methods, it is possible to extract word pairs that share a lexical relation of some sort. However, there is no method available for identifying types of relation correctly. For instance, it is not possible to tell the difference between antonyms, synonyms and other semantically related word pairs (in this case word pairs from the same dimensions, which co-occur significantly at sentence level, but are neither antonyms, nor synonyms, e.g. *klen* ‘weak’-*kort* ‘short’). The answer(s) to the types of question we are asking are not to be found solely on the basis of corpus data. As Mönnink (2000: 36) puts it “The corpus study shows which of the theoretical possibilities actually occur in the corpus, and which do not.” The questions we are asking call for additional methods.

A combination of corpus data, elicitation data and judgement data is valuable in order to determine if and how antonym word pairs vary in canonicity. It also sheds light on different aspects of the issue. Like Mönnick (2000), we believe that a methodologically sound descriptive study of linguistics is cyclic and preferably includes both corpus evidence and intuitive data (psycho-linguistic experimental data).

3 Data extraction

3.1 Method

proviso that the researcher determines the search items in the first place. Our method is of a two-step type, in that we mined the whole corpus for both individual occurrences and co-occurrence frequencies for all adjectives without any restrictions, and from those data we selected our seven dimensions and all their synonyms.

Antonyms co-occur in sentences significantly more often than chance would predict and canonical antonyms co-occur more often than contextually restricted antonyms (Justeson & Katz 1991; Willners 2001). This knowledge helps us to decide which antonyms to select for experiments investigating antonym canonicity. Willners & Holtsberg (2001) developed a computer program called *Coco* to calculate expected and observed sentential co-occurrences of words in a given set and their levels of probability. An advantage of *Coco* was that it took variation of sentence length into account, unlike the program used by Justeson & Katz (1991).

Coco produces a table which lists the individual words and the number of individual occurrences of these words in the corpus in the four left-most columns. Table 1 lists 12 Swedish word pairs that were judged to be antonymous by Lundbladh (1988) from Willners (2001): *N1* and *N2* are the number of sentences respectively in which *Word1* and *Word2* occur in the corpus. *Co* is the number of times the two words are found in the same sentence and *Expected Co* is the number of times they are expected to co-occur in the same sentence if predicted by chance. *Ratio* is the ratio between *Observed* and *Expected* co-occurrences and *P-value* is the probability of finding the actual number of co-occurrences that was observed or more under the null hypothesis that the co-occurrences are due to pure chance only. All of Lundbladh's antonym pairs co-occurred in the same sentence significantly more often than predicted by chance.

Table 1. Observed and expected sentential co-occurrences of 12 different adjective pairs (from Willners 2001: 72).

<i>Word1</i>	<i>Word2</i>	<i>N1</i>	<i>N2</i>	<i>Co</i>	<i>Expected Co</i>	<i>Ratio</i>	<i>P-value</i>
bred	smal	113	55	2	0.12	17.39	0.0061
djup	grund	117	17	1	0.04	27.17	0.036
gammal	ung	1050	455	47	8.84	5.32	0
hög	låg	760	333	47	4.68	10.04	0
kall	varm	102	102	12	0.19	62.32	0

kort	lång	262	604	21	2.93	7.17	0
liten	stor	1344	2673	111	66.48	1.67	0
ljus	mörk	84	126	7	0.20	35.82	0
långsam	snabb	55	163	4	0.17	24.11	0
lätt	svår	225	365	5	1.52	3.29	0.020
lätt	tung	225	164	7	0.68	10.25	0
tjock	tunn	53	85	4	0.08	47.98	0

Willners (2001) reports that 17% of the 357 Swedish adjective pairs that co-occurred at a significance level of 10^{-4} in the SUC² were antonyms. The study included all adjectives in the corpus. When the same data were (quite unorthodoxly) sorted according to rising *P-value*, antonyms clustered at the top of the list as in Table 2. Most of the antonym word pairs were classifying adjectives with overlapping semantic range, e.g. *fonologisk-morfologisk* ‘phonological-morphological’ and *humanistisk-samhällsvetenskaplig* ‘humanistic-of Social Sciences’. Among the 83% of the word pairs that were not antonyms were many other lexically related words.

Table 2. The top 10 co-occurring adjective pairs in the SUC, sorted according to rising *P-value*.

Swedish antonyms	Translation
<i>höger-vänster</i>	‘right-left’
<i>kvinnlig-manlig</i>	‘female-male’
<i>svart-vit</i>	‘black-white’
<i>hög-låg</i>	‘high-low’
<i>inre-yttre</i>	‘inner-outer’
<i>svensk-utländsk</i>	‘Swedish-foreign’
<i>central-regional</i>	‘central-regional’
<i>fonologisk-morfologisk</i>	‘phonological-morphological’
<i>horisontell-vertikal</i>	‘horizontal-vertical’

² Stockholm-Umeå Corpus, a one-million-word corpus compiled according to the same principles as the Brown Corpus. See <http://www.ling.su.se/staff/sofia/suc/suc.html>

Furthermore, Willners (2001) compared the co-occurrence patterns of what Princeton calls direct antonyms and indirect antonyms. Both types co-occur significantly more often than chance predicts. However, there is a significant difference between the two groups: while the indirect antonyms co-occur overall 1.45 times more often than would be expected if predicted by chance, the direct antonyms co-occur 3.12 times more often than expected.

The hypothesis we are testing in this study is that there are good and bad antonyms (cf. canonical and non-canonical). *Coco* provides a data-driven method of identifying semantically related word pairs. We used *Coco* to suggest possible candidates for the test set. However, since we wanted a balance between Canonical antonyms, Antonyms, Synonyms and Unrelated word pairs in the test set, human interference was necessary and we picked out the test items manually from the lists produced by *Coco*.

3.2 Result

Using the insights from previous work on antonym co-occurrence as our point of departure, we developed a methodology for selecting data for our experiments. To start with, we agreed on a set of seven dimensions from the output of the corpus searches of sententially co-occurring items that we perceived to be good candidates for a high degree of canonicity and identified the pairs of antonyms that we thought were the best linguistic exponents of these dimensions (see Table 3). For cross-linguistic research we made sure that the word pairs also had well-established correspondences in English. The selected antonym pairs are all scalar adjectives compatible with scalar degree modifiers such as *very*.

Table 3. Seven corresponding canonical antonym pairs in Swedish and English.

Dimension	Swedish antonyms	Translation
SPEED	<i>långsam-snabb</i>	‘slow-fast’
LUMINOSITY	<i>mörk-ljus</i>	‘dark-light’
STRENGTH	<i>svag-stark</i>	‘weak-strong’
SIZE	<i>liten-stor</i>	‘small-large’
WIDTH	<i>smal-bred</i>	‘narrow-wide’
MERIT	<i>dålig-bra</i>	‘good-bad’
THICKNESS	<i>tunn-tjock</i>	‘thick-thin’

Using *Coco*, we ran the words through the SUC. All of them co-occurred in significantly high numbers at sentence level and these pairs were set up as Canonical antonyms. Next, all Synonyms of the 14 adjectives were collected from a Swedish synonym dictionary.³ All the Synonyms of each of the words in each antonym pair were matched and run through the SUC in all possible constellations for sentential co-occurrence. This resulted in a higher than chance co-occurrence for quite a few words for each pair. We extracted the pairs that were significant at a level of $p < 0.01$ for further analysis. Using dictionaries and our own intuition, we then categorised the word pairs according to semantic relations. Finally, we picked two Antonyms, two Synonyms and one pair of Unrelated adjectives from the list of significantly co-occurring word pairs for dimension. Table 4 shows the complete set of pairs retrieved from the SUC: 42 pairs in all.

³ Strömbergs synonymordbok 1995. Alva Strömberg, Angered: Strömbergs bokförlag.

Table 4. The test set retrieved from the SUC. See Appendix A for translations.

Canonical antonyms	Antonyms	Synonyms	Unrelated
långsam-snabb	långsam-flink tråkig-het	långsam-släpig snabb-rask	het-plötslig
ljus- mörk	vit-dunkel melankolisk- munter	ljus-öppen mörk-svart	dyster-präktig
svag-stark	lätt-muskulös senig-kraftig	svag-matt stark-frän	flat-seg
liten-stor	obetydlig-kraftig	stor- inflytelserik	klen-kort
smal-bred	liten-väldig smal-öppen trång-rymlig	liten-oansenlig smal-spinkig bred-kraftig	liten-tjock
dålig-bra	dålig-god ond-bra ⁴	dålig-låg bra-god	fin-tokig
tunn-tjock	genomskinlig- svullen fin-grov	tunn-spinkig tjock-kraftig	knubbig-tät

We also included eleven word pairs from Herrmann et al.’s (1979) study of ‘goodness of antonymy’ (see Table 5). From his ranking of 77 items, we picked every sixth word pair, translated them into Swedish and classified them according to semantic relation: Canonical antonym (C), Antonym (A) and Unrelated (U). None of the pairs from Herrmann et al. (1979) were judged to be synonymous. The word pairs as well as the individual words in Table 4 and Table 5 were used as the test set in the psycholinguistic studies described below.

Table 5. Test items selected from Herrmann et al. (1979).

⁴ Due to sparse data, this item was added despite the fact that it did not meet the general criterion of being over the limit of 0.01. We chose *ond-bra* ‘evil-good’ because we expected interesting results for the English counterpart in the mirror study. *Ond-bra* ‘evil-good’ is included in the test set, but is not included in the result discussions.

Word 1	Word 2	Translated from	Herrmann's score	Semantic relation
ful	vacker	beautiful-ugly	4.90	C
smutsig	fläckfri	immaculate-filthy	4.62	A
trött	pigg	tired-alert	4.14	C
lugn	upprörd	disturbed-calm	3.95	A
hård	böjlig	hard-yielding	3.28	A
irriterad	glad	glad-irritated	3.00	A
sparsmakad	spännande	sober-exciting	2.67	A
overksam	nervös	nervous-idle	2.24	U
förtjusande	förvirrad	delightful-confused	1.90	U
framfusig	Hövlig	bold-civil	1.57	A
vågad	sjuk	daring-sick	1.14	A

4 Elicitation experiment

This section describes the method and the results of the elicitation experiment.

Stimuli and procedure The test set for the elicitation experiment involves the individual adjectives that were extracted as co-occurring pairs from the SUC and translations of selected word pairs from Herrmann et al.'s (1979) list of adjectives perceived by participants as better and less good examples of antonyms (see Table 4 and Table 5). Some of the individual adjectives occur in more than one pair, i.e. they might occur once, twice or three times. For instance, *långsam* 'slow' occurs three times and *snabb* 'fast' occurs twice. All second and third occurrences were removed from the elicitation test set, which means that *långsam* 'slow' and *snabb* 'fast' occur once in the test set used in the elicitation experiment. Once this was done, the adjectives were automatically randomised and printed in the randomised order. All in all, the test contains 85 stimulus words. All participants obtained the adjectives in the same order. The participants were asked to write down the best opposites they could think of for each of the 85 stimuli words in the test set. For instance:

Motsatsen till LITEN är _____

‘The opposite of LITTLE is’ _____

Motsatsen till PRÄKTIG är _____

‘The opposite of SPLENDID is’ _____

The experiment was performed using paper and pencil and the participants were instructed to do the test sequentially, that is, to start from word one and work forwards and to not go back to check or change anything. There was no time limit, but the participants were asked to write the first opposite word that came to mind. Each participant also filled in a cover page with information about name, sex, age, occupation, native language and parents’ native language. All the responses were then coded into a database using the stimulus words as anchor words.

Participants Twenty-five female and 25 male native speakers of Swedish participated in the elicitation test. They were between 20 and 70 years of age and represented a wide range of occupations as well as levels of education. All of them had Swedish as their first language, as did their parents. The data were collected in and around Lund, Sweden.

Predictions Our predictions are as follows:

- The test items that we deem to have Canonical antonyms will elicit only one another.
- The test items that we do not deem to be canonical will elicit varying numbers of antonyms - the better the antonym pairing, the fewer the number of elicited antonyms.

- The elicitation experiment will produce a curve from high participant agreement (few suggested antonyms) to low participant agreement (many suggested antonyms).

4.1 Results

We will start by reporting the general results in Section 5.1.1 and then go on to discuss the results concerning bidirectionality in Section 5.1.2. We performed a cluster analysis, the results of which are presented in Section 5.1.3.

4.1.1 General results

The main outcome of the elicitation experiment is that there is a continuum of lexical association of antonym pairs. In line with our predictions, there was a number of test words for which all the participants suggested the same antonym: *bra* ‘good’ (*dålig* ‘bad’), *liten* ‘small’ (*stor* ‘large’), *ljus* ‘light’ (*mörk* ‘dark’), *låg* ‘low’ (*hög* ‘high’), *mörk* ‘dark’ (*ljus* ‘light’), *sjuk* ‘ill’ (*frisk* ‘healthy’), *smutsig* ‘dirty’ (*ren* ‘clean’), *stor* ‘large’ (*liten* ‘small’), and *vacker* ‘beautiful’ (*ful* ‘ugly’). All the elicited antonyms across the test items are listed in Appendix A. Appendix A also shows that there is a gradual increase of responses from the top of the list to the bottom of the list. The very last item is *sparsmakad* ‘fastidious’, for which 33 different antonyms were suggested by the participants (including a non-answer). The shape of the list of elicited antonyms across the test items in Appendix A strongly suggests a scale of canonicity from very good matches to test items with no clear partners.

While Appendix A gives all the elicited antonyms across the test items, it does not provide information about the scores for the various individual elicited responses. The three-dimensional diagram in Figure 1 is a visual representation of how some stimulus words elicited the same word from all participants. Those are the maximally high bars found to the very

left of the diagram (e.g. *bra* ‘good’, *liten* ‘small’, *ljus* ‘light’, etc). Then four words follow for which 49 of the participants suggested the same antonym while another opposite was suggested in the 50th case. These four stimulus words were *dålig* ‘bad’, *svag* ‘weak’, *stark* ‘strong’, and *ond* ‘evil’. Forty-nine of the participants suggested *bra* ‘good’ as an antonym of *dålig* ‘bad’, *stark* ‘strong’ for *svag* ‘weak’, *svag* ‘weak’ for *stark* ‘strong’ and *god* ‘good’ for *ond* ‘evil’. The ‘odd’ suggestions were *frisk* ‘healthy’ for *dålig* ‘bad’, *klar* ‘clear’ for *svag* ‘weak’, *klen* ‘feeble’ for *stark* ‘strong’ and *snäll* ‘kind’ for *ond* ‘evil’. Since there are two response words for each of the four stimuli in these cases, there are two bars, one 49 units high at the back representing the most commonly suggested antonym and one small bar, only one unit high, in front of the big one, representing the single suggestions *frisk* ‘healthy’, *klar* ‘clear’, *klen* ‘feeble’ and *snäll* ‘kind’. The further we move towards the right in Figure 1, the more diverse the responses. In fact, the single suggestions spread out like a rug covering the bottom of the diagram as we move towards the right. However, there is usually a preferred response word which most of the participants suggested.

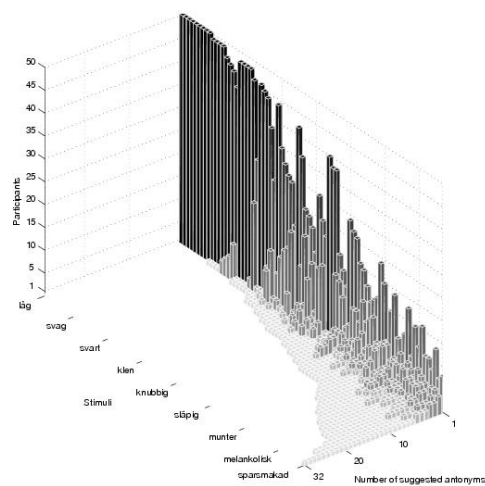


Figure 1. The distribution of Swedish antonyms in the elicitation experiment. The Y-axis gives the test items, with every tenth test item written in full. The X-axis gives the number of suggested antonyms across the participants given on the Z-axis.

Comment [CW3]: Please maximise the figure so that the readers can see properly.

There are some stimuli for which two response words were equally popular choices or which at least were both suggested by a considerable number of participants. For example, for *lätt* ‘light/easy’, 29 participants suggested *tung* ‘heavy’ and 20 *svår* ‘difficult’; *het* ‘hot’ elicited the responses *kall* ‘cold’ (24) and *sval* ‘chilly’ (20); and for *god* ‘good’, participants suggested *ond* ‘evil’ (20) and *äcklig* ‘disgusting’ (19). A common feature of these stimulus words is that they are associated with different strongly competing meaning dimensions or salient readings. Some other examples are *framfusig* ‘bold’: *tillbakadragen* ‘unobtrusive’ (20) and *blyg* ‘shy’ (16); *trång* ‘narrow’: *rymlig* ‘spacious’ (17) and *vid* ‘wide’ (15); *fläckfri* ‘spotless’: *fläckig* ‘spotted’ (17) and *smutsig* ‘dirty’ (15); *grov* ‘coarse’: *fin* ‘fine’ (17) and *tunn* ‘thin’ (14).

Like Appendix A, Figure 1 indicates that there is a scale of canonicity with a group of highly canonical antonyms to the left and a gradual decrease of canonicity as we move towards the right in the diagram. The stimulus words on the left-hand side of Figure 1 cannot be said to have any good antonyms at all.

4.1.2 Bidirectionality

In addition to the distribution of the responses for all the test items across all the participants, we also investigated to what extent the test items elicit one another in both directions. For instance, 50 participants gave *dålig* ‘bad’ as an antonym of *bra* ‘good’ and *ful* ‘ugly’ for *vacker* ‘beautiful’, but the pattern was not the same in the other direction. This is part of the information in Appendix A and Figure 1, but it is not obvious from the way

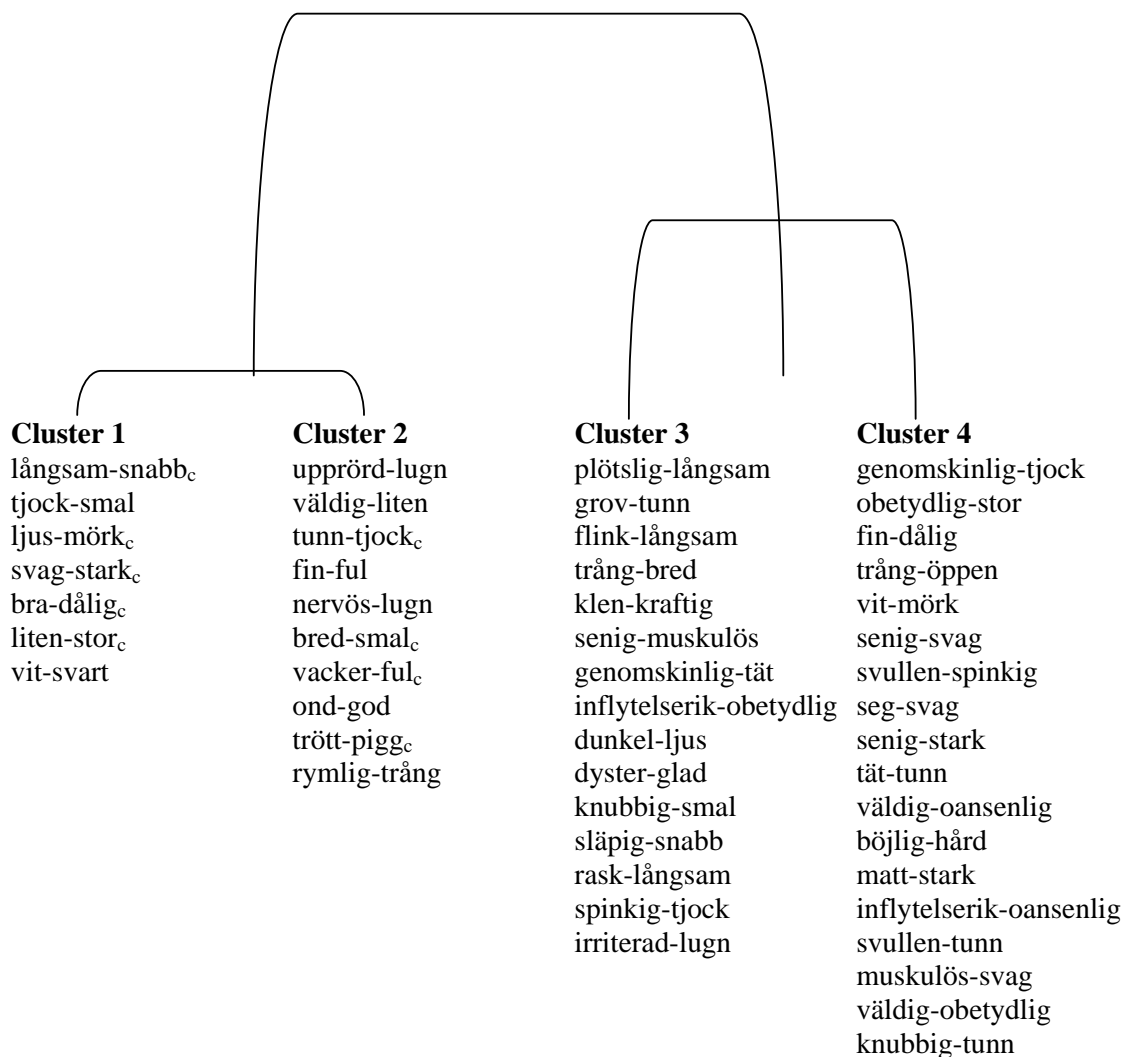
the information is presented. For the test items that speakers of Swedish intuitively deem to be good pairs of antonyms, this strong agreement held true in both directions, although not at the level of a one-to-one match, but one-to-two or one-to-three. While 50 participants responded with *dålig* ‘bad’ as the best opposite of *bra* ‘good’, two antonyms were suggested for *dålig* ‘bad’: *bra* ‘good’ by 49 participants and *frisk* ‘healthy’ by one participant. This points to the possibility that there is a stronger relationship between *bra* ‘good’ and *dålig* ‘bad’ than between *frisk* ‘healthy’ and *dålig* ‘bad’. In other words, Figure 1 shows that the more canonical pairs elicit only one or two antonyms, while there is a steady increase in numbers of ‘best’ antonyms the further we move to the right-hand side of the figure.

4.1.3 Cluster analysis

In order to shed light on the strength of the lexicalised oppositeness, a cluster analysis of strength of antonymic affinity between the lexical items that co-occurred in both directions was performed. It is important to note that only items that were also test items were eligible as candidates for participation in bidirectional relations. This means that some of the pairings suggested by the participants were not included in the cluster analysis. For instance, *tung* ‘heavy’ was considered the best antonym of *lätt* ‘light’ by 29 of the participants (as compared to 20 for *svår* ‘difficult’), but since neither *tung* nor *svår* were included among the test items, the pairings were not measured in the cluster analysis. The results of the cluster analysis are, however, comparable to the results of sentential co-occurrence of antonyms in the corpus data and the results of the judgement experiment, since the same word pairs are included.

To this end, a hierarchical agglomerative cluster analysis using Ward amalgamation strategy (Oakes 1998:119) was performed on the subset of the data that were bidirectional. Agglomerative cluster analysis is a bottom-up method that takes each entity (i.e. antonym pairing) as a single cluster to

start with and then builds larger and larger clusters by grouping together entities on the basis of similarity. It merges the closest clusters in an iterative fashion by satisfying a number of similarity criteria until the whole dataset forms one cluster. The advantage of cluster analysis is that it highlights associations between features as well as the hierarchical relations between these associations (Glynn et al. 2007, Gries & Divjak 2009). Cluster analysis is not a confirmatory analysis but a useful tool for exploratory purposes.



god-dålig
 senig-kraftig
 melankolisk-munter
 släpig-rask
 rymlig-liten
 irriterad-glad
 spinkig-kraftig
 knubbig-spinkig

Figure 2. Dendrogram of the bidirectional data

Figure 2 shows the dendrogram produced on the basis of the cluster analysis. The number of clusters was set to four to match the four conditions on the basis of which we retrieved our data from the sententially co-occurring pairs in the first place (Canonical antonyms, Antonyms, Synonyms and Unrelated). Figure 2 shows the hierarchical structure of the clusters. There are two branches. The left-most branch hosts Cluster 1 and Cluster 2 and the right-most branch Cluster 3 and Cluster 4. The closeness of the fork to the clusters indicates a closer relationship. The tree structure reveals that there is a closer relation between Cluster 3 and Cluster 4 than between Cluster 1 and Cluster 2.

Comment [CW4]: When we refer to the experimental categories, we use initial cap, otherwise lower case.

Figure 2 gives the actual pairings in the boxes at the end of the branches. There are fewer pairs at the end of the left-most branches than at the end of the branches on the right-hand side. Five of the word pairs in Cluster 1 were included in the test set as canonical antonyms: *långsam-snabb* ‘slow-quick’, *ljus-mörk* ‘light-dark’, *svag-stark* ‘weak-strong’, *bra-dålig* ‘good-bad’ and *liten-stor* ‘small-large’ (subscripted with c in Figure 2). The other two word pairs in Cluster 1 were *vit-svart* ‘white-black’ from the LUMINOSITY dimension and *tjock-smal* ‘fat-thin’ from THICKNESS. The rest of the word pairs in Cluster 1 were not included as pairs in the experiment.

In Cluster 2, there are four word pairs featured in the test set as canonical: *tunn-tjock* ‘thin-thick’, *bred-smal* ‘wide-narrow’, *vacker-ful*



‘beautiful-ugly’ and *trött-pigg* ‘tired-alert’. The rest of the word pairs in Cluster 2 are intuitively good pairings. They were, however, not among the pairings that we deemed canonical in the design of the test set, e.g. *upprörd-lugn* ‘upset-calm’, *väldig-liten* ‘enormous-small’, *fin-ful* ‘pretty-ugly’, *nervös-lugn* ‘nervous-calm’, *ond-god* ‘evil-good’ and *rymlig-trång* ‘spacious-narrow’.

It is not obvious what the systematic differences are between the degrees of oppositeness in Clusters 3 and 4. As the dendrogram above shows, they are in fact associated. However, they do not correspond to the Synonyms and Unrelated word pairs in the test set, since the cluster analysis is based on the results of the elicitation experiment where the participants were asked to provide the best antonym.

5 Judgement experiment

This section describes the methodology of the judgement experiment in which the participants were asked to evaluate word pairings in terms of how good they thought each pair was as a pair of antonyms. The experiment was carried out online. The design of the screen is shown in Figure 3.

How good is strong-weak as a
pair of opposites?

very bad
excellent

Figure 3. An example of a judgement task in the online experiment (translated into English).

As Figure 3 shows, the participants were presented with questions of the form: *Hur bra motsatser är X-Y?* ‘How good is X-Y as a pair of opposites?’ and *Hur bra motsatser är Y-X?* ‘How good is Y-X as a pair of opposites?’ The question was formulated using *bra* ‘good’ (not *dålig* ‘bad’) in order for the participants to understand the question as an impartial how-question, since *Hur dåliga motsater är fet-smal?* ‘How bad is fat-lean as a pair of opposites?’ presupposes ‘badness’. The end-points of the scale were designated with both icons and text. On the left-hand side there is a sad face (*very bad* antonyms), while there is a happy face on the right-hand side (*excellent* antonyms). The task of the participants was to tick a box on a scale consisting of eleven boxes. We were also interested in whether the ordering of the pairs had any effect. Our predictions were as follows.

- The nine test pairings that we deem to be canonical will receive 11 on the scale of ‘goodness’ of pairing of opposites.
- The order of presentation of the Canonical antonyms as well as the Antonyms will give rise to significantly different results. *Word1-Word2* will be considered better pairings than *Word2-Word1*.
- There will be significant differences between the judgements about Canonical antonyms, Antonyms, Synonyms and Unrelated pairings.

Stimuli The same test set as in the elicitation experiment was used (see Table 4 and 5), but while the pairing of the antonyms was not an issue in the first experiment, it was essential to the judgement test. The stimuli were presented as pairs and the test items were automatically randomised for each participant. Half of the participants were given the test items in the order *Word1-Word2*, while the other half were presented with the words in reverse order, i.e. *Word2-Word1*.

Procedure The judgement experiment was performed online using E-prime as experimental software. E-prime is a commercially available Windows-based presentation program with a graphical interface, a scripting language similar to Visual Basic and response collection. E-prime conveniently logged the ratings as well as the response times in separate files for each of the participants. The participants were presented with a new screen for each word pair (see Figure 3). The task of the participants was to tick a box on a scale consisting of eleven boxes. The screen immediately disappeared upon clicking which prevented the participants from going back and changing their responses. Between each judgement task there was a blank screen with an asterisk, and when the participants were ready for the next task they signalled that with a mouse-click. Before the actual test started, the participants were asked to give some personal data (name, age, sex,

occupation, native language and parents' native language). There then followed some instructions such as how to do the mouse-clicks and information about the fact that the test was self-paced. Each participant had two test trials before the actual judgement test of the 53 test items. The purpose of the study was revealed to the participants in the instructions.

As has already been mentioned, the judgement experiment was divided into two parts: 25 participants were given the test set as Non-Reverse (*Word1-Word2*, e.g. *långsam-snabb* 'slow-fast') and 25 participants were given the test set in the reverse order: Reverse (*Word2-Word1*, e.g. *snabb-långsam* 'fast-slow'). This was done to measure whether the order of the sequence influenced the results in any way.

Participants Fifty native speakers of Swedish participated in the judgement test. None of them had previously participated in the elicitation test. Twenty-nine of the participants were women and 21 were men between 20 and 62 years of age. All of them had Swedish as their first language.

5.1 Results

This section reports on the results of the judgement experiment. We start reporting on the results concerning sequencing in Section 6.1.1 since they affect the treatment of the data reported in the section on strength of canonicity (Section 6.1.2).

5.1.1 Sequencing

As has already been pointed out, the test was performed in such a way that half of the participants were presented with the test items in the order: *Word1-Word2*, and the other half in reverse order, *Word2-Word1*. We assumed that the order would have an impact on the results, at least for the canonical antonyms. A subject analysis and an item analysis were performed. The factors involved were directionality, category (Canonical

antonyms, Antonyms, Synonyms and Unrelated) and the interaction between directionality and category.

In the subject analysis, each participant was the basic element for analysis. All judgements for the individual participants were averaged within each of the four conditions, yielding four numbers per participant. Then a repeated measures ANOVA analysis of variance (Woods et al 1986: 194-223) was performed on both data sets. In the item analysis, each item (i.e. word pair) was the basic element for analysis. The judgements given by each participant on each condition were averaged, resulting in four numbers for each item, and a Univariate General Linear Model analysis was performed. Finally, Bonferroni's post hoc test (Field 2005: 339) was used to compare the differences between the categories. The same procedure was used for the response times.

The statistical analysis shows that the order of sequence does not have any effect on the results: $F1[1,48]=1.056$, $p=0.309$, $F2[1,98]=0.206$ $p=0.651$. The interaction between the sequence and category does not have an effect either: $F1[3,144]=0.811$, $p>0.05$, $F2[3,98]=0.069$, $p=0.976$. Category, on the other hand, does have an effect: $F1[3,144]=1777.991$, $p<0.001$, $F2[3,98]=138.987$, $p < 0.001$. Figure 4 shows that the two test batches (marked with REV=0 and REV=1) follow the same pattern. Since the order of the sequence did not have an impact on the results, the data for the two directions will be treated as one batch and will not be separated in the analyses that follow.

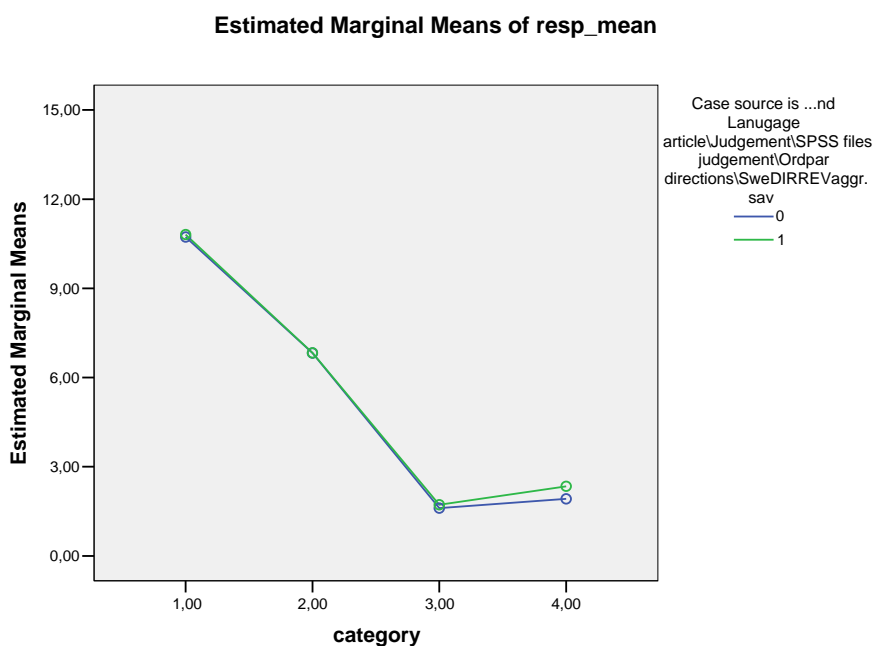


Figure 4. Sequential ordering: there is no significant difference between the mean answers of the two test batches.

5.1.2. Strength of canonicity

The mean response for each word pair in the test set is presented in Table 6. The mean responses for the Canonical antonyms vary between 10.40 and 10.92. None of the word pairs have a response mean of 11, which we expected for the Canonical antonyms. They do, however, top the list. The means for the Antonyms vary greatly, from 10.32 for *fin-grov* ‘fine-course’ to 1.68 for *genomskinlig-svullen* ‘transparent-swollen’. Below 2.52, a mix of unrelated and synonymous word pairs are found and the word pair that was judged to be the ‘worst’ antonym pair was *tunn-spinkig* ‘thin-skinny’ (1.24).

Table 6. Mean responses for each of the word pairs in the test set, both directions included.

Word pair	Mean response	Semantic category
ljus-mörk	10.92	C
långsam-snabb	10.88	C
liten-stor	10.84	C
svag-stark	10.8	C
trött-pigg	10.76	C
dålig-bra	10.68	C
ful-vacker	10.64	C
smal-bred	10.6	C
tunn-tjock	10.4	C
fin-grov	10.32	A
trång-rymlig	10.2	A
dålig-god	9.84	A
smutsig-fläckfri	9.36	A
lugn-upprörd	9.28	A
melankolisk-munter	9.04	A
liten-väldig	8.52	A
framfusig-hövlig	8.4	A
hård-böjlig	7.84	A
långsam-flink	7.8	A
ond-bra	6.84	A
irriterad-glad	6.56	A
senig-kraftig	5.88	A
obetydlig-kraftig	5.44	A
vit-dunkel	5.4	A
lätt-muskulös	4.44	A
liten-tjock	4.08	U
tråkig-het	3.68	A
smal-öppen	3.2	A
sparsmakad-spännande	2.76	A
bra-god	2.52	S
dyster-präktig	2	U
overksam-nervös	1.92	U
fin-tokig	1.88	U
förtjusande-förvirrad	1.84	U

långsam-släpig	1.8	S
svag-matt	1.76	S
stark-frän	1.76	S
stor-inflytelserik	1.68	S
knubbig-tät	1.68	U
genomskinlig- svullen	1.68	A
snabb-rask	1.6	S
bred-kraftig	1.6	S
flat-seg	1.56	U
dålig-låg	1.56	S
ljus-öppen	1.48	S
klen-kort	1.48	U
liten-oansenlig	1.44	S
het-plötslig	1.44	U
mörk-svart	1.4	S
tjock-kraftig	1.4	S
vågad-sjuk	1.32	A
smal-spinkig	1.28	S
tunn-spinkig	1.24	S

The overall mean responses for the four categories are presented in Table 7. The Canonical antonyms have a mean response of 10.72, close to the maximum, 11. The standard deviation is also small for this category, 0.6, which reflects high consensus among the participants. The Antonyms have a significantly lower mean of 6.82, but with a large standard deviation, 3.37. This indicates a lower degree of consensus among the participants. The response for the Synonyms is 1.61, with a standard deviation of 1.33, and for the Unrelated it is 1.92, with a standard deviation of 1.55. There is no significant difference between the last two categories.

Table 7. Mean responses for Canonical antonyms, Antonyms, Synonyms and Unrelated word pairs

Category	Mean	Std. deviation
Canonical antonyms	10.724	0.6084
Antonyms	6.824	3.3728
Synonyms	1.609	1.3279
Unrelated	1.920	1.5502

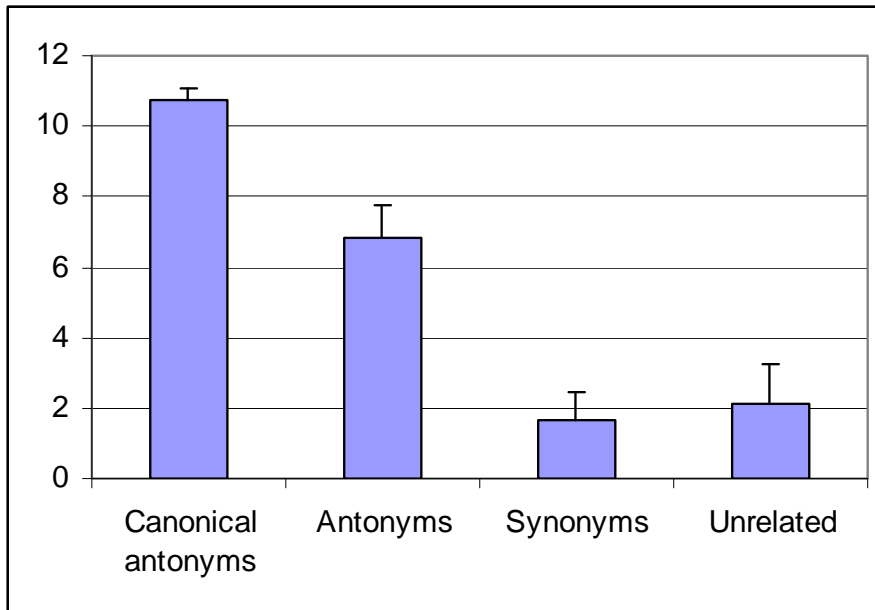


Figure 5. Mean responses for Canonical antonyms, Antonyms, Synonyms and Unrelated word pairs.

The results in Table 7 are also illustrated in Figure 5. We performed a repeated measures ANOVA and the differences between the Canonical antonyms and Antonyms as well as between Antonyms and the two other categories (Synonyms and Unrelated) were significant both in the subject analysis ($F(3,147)=1784.874$, $p<0.001$) and in the item analysis ($F(3,49)=70.361$, $p<0.001$).

Post hoc comparisons using Tukey's HSD procedure (Fields 2005: 340) suggested that the four conditions form three subgroups: (1) Canonical antonyms, (2) Antonyms and (3) Synonyms and Unrelated.

6 Discussion

The main goal of this paper was to investigate and report on three different methods of studying antonym canonicity and to increase our knowledge about Swedish antonyms. We used a corpus-driven method to suggest possible candidates, categorise the semantic relations between the suggested word pairs and pick six items from each semantic dimension manually. We then used two different psycholinguistic techniques to investigate the strength of oppositeness between the word pairs in the test set. Summaries of the results of the three parts of the study will be given in Sections 7.1, 7.2 and 7.3. Then a discussion of the advantages and disadvantages of using various types of research technique for the same topic will follow in section 7.4.

6.1 Data extraction

Under the assumption that semantically related words co-occur significantly more often than chance predicts, we used a corpus-driven method to suggest possible candidates for the test set. We collected Synonyms of the Canonical antonyms from seven predefined semantic dimensions (SPEED, LUMINOSITY, STRENGTH, SIZE, WIDTH, MERIT and THICKNESS) and the figures for Expected and Observed sentential co-occurrence as well as *P-value* was calculated for all possible permutations of word pairs within each dimension. The word pairs that co-occurred significantly at a p-level of 0.05 qualified as candidates for the test set. From these pairs we selected one antonymous pair, two pairs of Synonyms and two pairs of Unrelated words for each dimension. Together with the Canonical antonyms of each dimension as well as 11 word pairs that were previously studied for antonym canonicity by Herrmann et al. (1979), they made up the test set used in the psycholinguistic experiments (see Tables 4 and 5).

Due to a shortage of publicly/generally available large corpora for Swedish, the present corpus study is performed on a fairly small corpus, the SUC, which comprises one million words. It would be a significant improvement to do all the calculations for the word pairs on a larger corpus, as we have done for English data in Paradis et al. (submitted), where we used the 100-million-word corpus BNC.

6.2 Elicitation experiment

Fifty participants, evenly distributed over gender, were asked to provide the best opposite they could think of for 85 stimulus words. In accordance with what we predicted, the participants' responses consisted of a varying number of unique response words for the different test items as shown in Figure 1. There were nine stimulus words for which all participants gave the same answer, eight for which all 50 participants but one gave the same answer and then the number of participants giving the same answer decreases as the number of unique answers increases. The results generally confirm our predictions: (1) the test items which were suggested by the co-occurrence data and which we deemed to have canonical status elicited one another strongly; (2) the test items that we did not deem to be canonical elicited varying numbers of antonyms - the better the antonym pairing, the fewer the number of elicited antonyms; and (3) the elicitation experiment produced a curve from high participant agreement (few suggested antonyms) to low participant agreement (many suggested antonyms).

The predictions imply that both words in a canonically antonymous pair would elicit only each other, but this was not the case. Only for the semantic dimensions LUMINOSITY (*mörk-ljus* 'dark-light') and SIZE (*stor-liten* 'large-small') did the participants' responses agree 100% in both directions. This might be interpreted as canonicity somehow being linked to directionality, or it may be the case that direction is a result of polysemy rather than inherent to canonical antonyms.

Figure 6 illustrates elicitations from the field of MERIT in which we find one word pair with strong bidirectional evidence and three word pairs with strong unidirectional evidence. *Bra-dålig* ‘good-bad’ is one of the word pairs in the study with 100% agreement in both directions, i.e. all 50 participants offered *dålig* ‘bad’ to the stimulus *bra* ‘good’ and vice versa. The three word pairs with strong unidirectional evidence are *fin-ful* ‘pretty-ugly’, *ful-vacker* ‘ugly-beautiful’ and *ond-god* ‘evil-good’.⁵ *Dålig* ‘bad’ was also suggested as the best opposite of *god* ‘good’ by six participants and *fin* ‘pretty’ or perhaps ‘fine’ in this context, by one participant, as the fields of BEAUTY and GOODNESS can become entangled in the field of MERIT. It is always possible to construe opposition with the help of context, even for words that do not seem to be in semantic opposition at all, and some word pairs are good antonyms in certain contexts, but not in all, e.g. *fin* ‘pretty’ and *dålig* ‘bad’ are very good antonyms in the context of fruit and vegetables, whereas *dålig* ‘bad’ and *god* ‘good’ are often used about books. It is not possible to develop this further since we did not control for context in this study.

⁵ To keep the figure simple, we only included words from the test set that were also found among the responses. That is the reason why the numbers by the arrows in Figure 6 do not always add up to 50.

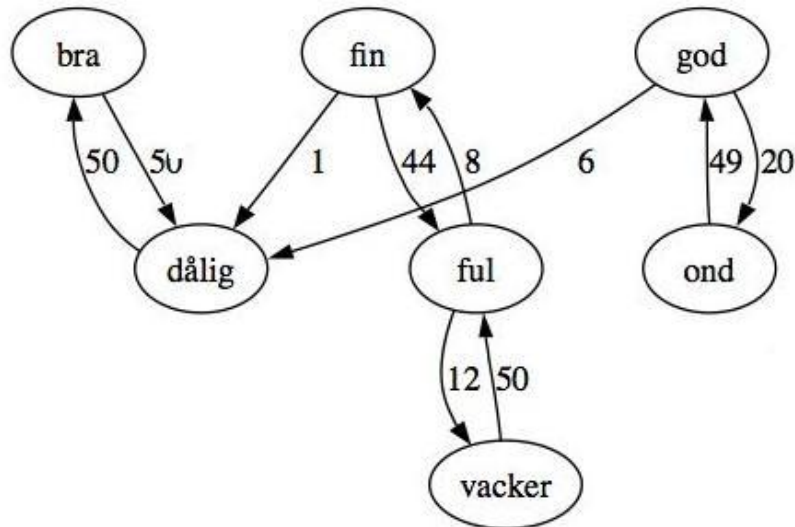


Figure 6. Relations between *bra* ‘good’, *dålig* ‘bad’, *fin* ‘pretty’, *ful* ‘ugly’, *vacker* ‘beautiful’, *god* ‘good’ and *ond* ‘evil’, based on the elicitation experiment. The number of responses is marked by each arrow.

In the cluster analysis, four clusters were predefined: all Canonical antonyms from the test set appear in Clusters 1 and 2, which are also closely related (see the dendrogram in Figure 2). We also find some other canonical word pairs that were not part of the test set in these clusters, such as *vit-svart* ‘white-black’ and *ond-god* ‘evil-good’, since we included all data for which we had bidirectional results in the cluster analysis, not just the word pairs in the test set. Clusters 3 and 4 were less closely related than the two previous clusters, and the two pairs of clusters were in turn related to each other. A drawback of the elicitation method is that even though all words in the test set were included as stimuli, most of the suggested pairs were not part of the test set. Since we asked the participants for the best opposite, we do not find test items from the Synonyms and Unrelated in the result.

The experiment was self-paced and the test items were presented out of context. The possible effects of this is that the participants may have had time to construct their own scenarios for each word and may not have

always written down the first opposite that came to mind. The lack of control for context is also an issue for the polysemous items in the test set such as *lätt* ‘light/easy’ which is a member of two meaning dimensions and consequently forms two pairs: *lätt-tung* ‘light-heavy’ and *lätt-svår* ‘easy-difficult’. This also applies to *god* ‘good/tasty’, which forms the pairs *god-ond* ‘good-evil’ and *god-äcklig* ‘tasty-disgusting’. There was a more or less equal number of responses connected to each meaning. This experiment was not designed to determine whether the participants made conscious choices, or whether half of them had shorter access time to one meaning or the other, which would have helped in the analysis of the polysemous items.

6.3 Judgement experiment

The judgement experiment was performed online and involved 50 participants who were asked to judge how good they thought each of the pairs in the test set were as a pair of antonyms. They made their judgements on an 11-unit scale and since we expected the ordering of the pairs to have an impact, half of the participants were given the stimulus word pairs in one order and the other in reversed order, i.e. *Word1-Word2* and *Word2-Word1*.

Our expectations concerning order of sequence were built on markedness theory (e.g. Lehrer 1985 and Haspelmath 2006), in which results show that one member of an antonym pair is more natural than the other, i.e. the unmarked one. Unexpectedly, the order of the words did not have a significant impact on the result. Even though this result has interesting implications for markedness theory, that track is beyond the scope of this study and we put all the data together in one batch, disregarding direction of presentation.

The general result for the judgement study, using all the data in the same batch, was that the responses to the four predefined categories formed three significantly different groups: Canonical antonyms (M=10.72), Antonyms (M=6.82), and Synonyms (M=1.61) and Unrelated (M=1.92),

which formed one group. We predicted significant differences in the judgements of all four groups, and this was confirmed by Canonical antonyms and Antonyms which are significantly different both from each other and from the Synonyms and Unrelated. Our prediction concerning the difference between Synonyms and Unrelated word pairs was disconfirmed: they were not judged to be significantly different with respect to degree of oppositeness.

The results support the general hypothesis of this paper in that there is a group of Canonical antonyms significantly different from non-canonical antonyms.

Herrmann et al. (1979) performed a judgement test using pen and paper and found that the word pairs in the test set were ranked on a scale of ‘goodness of antonymy’. The result for the 11 word pairs translated from Herrmann et al.’s (1979) study included in our study is consistent with his ranking (see Table 5). As in his study, *ful-vacker* ‘ugly-beautiful’ top the list as the best opposite word pair. *Smutsig-fläckfri* ‘filthy-immaculate’ and *trött-pigg* ‘tired-alert’ have traded places in the ranking. The main diverging result is *framfusig-hövlig* ‘bold-civil’ which has a ranking of 1.74 in Herrmann’s 5-unit scale but 8.4 in our 11-unit scale. Our intuitions agree with the participants’ judgements that the Swedish pair *framfusig-hövlig* ‘bold-civil’ actually are good opposites. The reason for this discrepancy may be that the translation into Swedish does not match the English original in terms of the semantic dimension of the antonymy relation.

While our data seem to converge with Herrmann et al.’s (1979) results (see Table 5), they used it to support a non-dichotomous view of canonicity. In contrast, our results, using similar methods, support a dichotomous view, since we do find a significant difference between Canonical antonyms and Antonyms.

Table 8. Test items selected from Herrmann et al. (1979).

Word 1	Word 2	Translated from	Herrmann's score	Score in present study	Semantic relation
ful	vacker	beautiful-ugly	4.90	10.64	C
smutsig	fläckfri	immaculate-filthy	4.62	9.36	A
trött	pigg	tired-alert	4.14	10.76	C
lugn	upprörd	disturbed-calm	3.95	9.28	A
hård	böjlig	hard-yielding	3.28	7.84	A
irriterad	glad	glad-irritated	3.00	6.56	A
sparsmakad	spännande	sober-exciting	2.67	2.76	A
överksam	nervös	nervous-idle	2.24	1.92	U
förtjusande	förvirrad	delightful-confused	1.90	1.84	U
framfusig	hövlig	bold-civil	1.57	8.4	A
vågad	sjuk	daring-sick	1.14	1.32	A

6.4 Dichotomy vs. continuum

Both psycholinguistic experiments point in the same direction. In the elicitation experiment, the Canonical antonyms elicit one another to a larger extent than the non-canonical antonyms and are all found in Clusters 1 and 2. The cluster analysis is not confirmatory, but the result is in favour of the dichotomy approach to ‘goodness of antonymy’. In the judgement experiment, they were judged significantly different from the Antonyms as a group. This confirms that there seems to exist a small group of opposite word pairs that are ‘better’ antonyms than others.

Focusing on the results of the non-canonical antonyms, we find clear indications of a continuum, namely: the varying number of unique responses in the elicitation which is reflected both in the ‘staircase’ form of Appendix B and in the slope of the bars at the back and the gradually growing ‘carpet’ in Figure 1; and the large dispersion of means for the non-canonical

antonyms in the judgement test, varying between 1.68 and 10.32, also reflected in the large standard deviation (3.37). Our results for non-canonical antonyms also validate Herrmann et al.'s (1979) study.

To conclude, there seems to be both a dichotomy and a continuum involved in the categorisation of 'goodness of antonymy'. The non-canonical antonyms vary greatly in the degree of oppositeness they exhibit, while there is a small group of extremely good antonyms that are not dispersed on a continuum of oppositeness.

6.4. Methodological remarks

Three different methods have been used in the studies reported in this paper. The research process can be described as following cycles involving the *researcher's intuitions, knowledge from the literature, corpus-based research* and *intuitive data* (Mönnink 2000), as discussed in Section 3. To this, we can add *encyclopaedic knowledge*, since dictionaries and encyclopaedias were important sources when we constructed the test set, although this can also be viewed as a special case of *knowledge from the literature*.

The choice of the test set starts out with a corpus study of the co-occurrence patterns of all possible combinations of adjective pairs in the SUC. In combination with our intuition, we picked out seven well-established semantic dimensions designated at the end poles by adjective pairs that co-occurred significantly more often than chance predicts. We used encyclopaedic knowledge to pick out the Synonyms of each of the words from the previous step. Corpus-driven methods were used to suggest possible word pairs for the test set, and these were categorised manually by the researchers who then picked two non-canonical antonyms, two Synonyms and one Unrelated word pair for each dimension. The test set was randomised and used as a stimulus in the elicitation experiment, in which we collected intuitive metalexical data from the participants. The data were

analysed using statistical methods and the results were interpreted in relation to the literature and to the researchers' intuitions.

The cycle takes yet another turn in the judgement experiment, for which the test set was randomised for each participant, who judged the stimulus word pairs on 'goodness of oppositeness' according to their intuitions. The researchers' intuitions, grounded in the literature, were used in the analysis of the statistical results of the judgement experiment as well as in bringing the different studies together.

Table 9. Research cycles of the reported studies in relation to Mönnink (2000).

Cycles	
Researchers' intuitions & Knowledge from the literature Corpus-driven methods	The research idea itself Running <i>Coco</i> on all permutations of adjectives in the SUC shows that the canonical antonyms co-occur more often than non-canonical antonyms and other semantically related word pairs.
Researchers' intuitions & Encyclopaedic knowledge	Selection of dimensions and canonical antonyms from the results of the previous step
Encyclopaedic knowledge	Collecting all synonyms of each of the words among the canonical antonyms
Corpus-driven methods	<i>Coco</i> suggests other significantly co-occurring word pairs as candidates for the test set.
Researchers' intuitions & Encyclopaedic knowledge & Knowledge from the literature	Manual categorisation of semantic categories
Researchers' intuitions & Encyclopaedic knowledge & Knowledge from the literature	Selecting six word pairs from each semantic dimension
Intuitive data from participants	Elicitation experiment
Researchers' intuitions & Knowledge from the literature	Analysis and interpretation of the results

Intuitive data from participants	Judgement experiment
Researchers' intuitions & Knowledge from the literature	Analysis and interpretation of the results
Researchers' intuitions & Knowledge from the literature	Bringing the results of the different studies together

The main reason for using cycles in the research process in this way is that it gives us a diversified picture of the issue. In this study, we have made several turns which have provided us with a number of perspectives on the issue of 'goodness of antonymy'.

7 Conclusion

The main goal of this paper was to combine three methods for the study of antonym canonicity and to report the results of experiments using Swedish data. We used corpus-driven methods to extract possible candidates for the test set from seven predefined semantic dimensions (SPEED, LUMINOSITY, STRENGTH, SIZE, WIDTH, MERIT and THICKNESS) and then picked one pair of Canonical antonyms, two pairs of non-canonical antonyms, two pairs of Synonyms and one Unrelated word pair from each dimension. We also included 11 word pairs from Herrmann et al. (1979) translated into Swedish. The test set was used as individual stimulus words in the elicitation experiment and as word pairs in the judgement experiment.

The elicitation experiment produced a curve from high participant agreement, i.e. all participants suggested the same opposite to the stimulus, to low participant agreement, i.e. many suggested antonyms. The cluster analysis shows that there is a group of Canonical antonyms in the test set, while the non-canonical antonyms vary greatly in 'goodness of antonymy', which is reflected in the variation of unique response words. There were many polysemous items and we did not control for context, which is why it is not possible to draw any conclusions about directionality, i.e. whether it is

of importance to ‘goodness of antonymy’ that the two words within a pair elicit one another as best opposite.

The judgement experiment also points to a group of Canonical antonyms significantly different from the non-canonical antonyms. Both were significantly different from the Synonyms and Unrelated word pairs, while, unexpectedly, the two latter categories were not significantly different from each other. Also unexpected, we found that the order of sequence (*Word1-Word2* vs. *Word2-Word1*) did not have any significant effect on the results. While the result for the Canonical antonyms is clear, the means for the non-canonical antonyms are spread out over a majority of the 11-unit scale used in the experiment. We interpret this as evidence that the non-canonical antonyms are sensitive to ‘goodness of antonymy’ in a scalar format.

We use and report on a variety of methods to study ‘goodness of antonymy’: data-driven suggestions for the test set, manual semantic categorisation and final choice of test set items, an elicitation experiment performed with paper and pencil and a judgement experiment performed online. The study as a whole goes through several cycles of researchers’ intuitions – encyclopaedic knowledge – participants’ intuitions – knowledge from the literature, which should vouch for scientific soundness. The variation of method gives a more complex and more complete picture of the issue.

Both the psycholinguistic experiments show that there is a small group of exceptionally good antonyms (Canonical antonyms) which are significantly different from the non-canonical antonyms, which in reality is an indefinite number. While we see a clear dichotomy between the canonical and non-canonical antonyms, the non-canonical vary greatly in degree of ‘goodness of antonymy’ – there is a continuum, as well as a dichotomy. It has previously been postulated that the canonical antonyms are of great importance to the organisation of the vocabularies of languages

(e.g. Fellbaum 1998), but it is equally interesting that virtually any two words can be construed as antonyms with contextual means (Paradis submitted).

Our next step is to move on to neuro-linguistic methods to see if we can validate the existence of a group of canonical antonyms as well as a continuum of non-canonical antonyms using event-related potentials.

References

- Cruse, D. Alan (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Fellbaum, Christiane (Ed.) (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Field, Andy (2005). *Discovering statistics using SPSS*. London: Sage Publications.
- Francis, Gill (1993). A corpus-driven approach to grammar: principles methods and examples. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp.137-156).
- Glynn, Dylan, Dirk Geeraerts, & Dirk Speelman (2007, December). *Entrenchment and social cognition. A corpus-driven study in the methodology of language description*. Paper presented at the first SALC (Swedish Association for Language and Cognition) conference at Lund University 29 December – 1 December.

- Gries, Stephan Th. & Dagmar S. Divjak (2009). Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics*. Amsterdam: John Benjamins.
- Gross, Derek & Katherine J. Miller (1990). Adjectives in WordNet. In WordNet: an on-line lexical database. *International Journal of Lexicography* 3.4 265-277.
- Haspelmath, Martin (2006). Against markedness (and what to replace it with). *Journal of Linguistics* 42.1: 25-70.
- Herrmann, Douglas J., Roger J.S. Chaffin, Gina Conti, Donald Peters & Peter H. Robbins (1979). Comprehension of antonymy and the generality of categorization models. *Journal of experimental psychology: human learning and memory* 5, 585-597.
- Jones, Steven (2002). *Antonymy: a corpus-based perspective*. London: Routledge.
- Jones, Steven, Carita Paradis, M. Lynne Murphy, and Caroline Willners (2007). Googling for opposites: a web-based study of antonym canonicity. *Corpora* 2.2, pages129-154.
- Justeson, John S. & Slava M. Katz (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics* 17, 1-19.

- Justeson, John S. & Slava M. Katz (1992). Redefining antonymy: the textual structure of a semantic relation. *Literary and Linguistic Computing* 7, 176–184.
- Lehrer, Adrienne (1985). Markedness and antonymy. *Journal of Linguistics* 21, 397- 429.
- Lehrer, Adrienne & Keith Lehrer (1982). Antonymy. *Linguistics and Philosophy* 5, 483-501.
- Lundbladh, Carl-Erik (1988). Adjektivets komparation i svenskan. En semantisk beskrivning. *Lundastudier i nordisk språkvetenskap*, A:40. Lund University Press.
- Lyons, John (1977). *Semantics*. Cambridge: Cambridge University Press.
- Muehleisen, Victoria (1997). *Antonymy and Semantic Range in English*. PhD.diss. Evanston, IL, Northwest University.
- Murphy, M. Lynne (2003). *Semantic relations and the lexicon: antonyms, synonyms and other semantic paradigms*. Cambridge: Cambridge University Press.
- Murphy, M. Lynne, Carita Paradis, Caroline Willners & Steven Jones (2009). Discourse functions of antonymy: a cross-linguistic investigation. *Journal of Pragmatics*.
- de Mönnink, Inge (2000). *On the move. The mobility of constituents in the English noun phrase: A multi-method approach*. Amsterdam/Atlanta, GA: Rodopi.

- Oakes, Michael P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
- Paradis, Carita (2003). Is the notion of *linguistic competence* relevant in Cognitive Linguistics? *Annual review of Cognitive Linguistics 1*. Pages 247-271.
- Paradis, Carita (2005). Ontologies and construal in lexical semantics. *Axiomathes 15*, 541-573.
- Paradis, Carita & Caroline Willners (2007). Antonyms in dictionary entries: methodological aspects. *Studia Linguistica 61.3*, 261-277.
- Paradis, Carita, Caroline Willners & Steven Jones (submitted). Good and bad opposites: using textual and experimental techniques to measure antonym canonicity.
- Storjohann, Petra (2005). Corpus-driven vs. corpus-based approach to the study of relational patterns. *Proceedings of the Corpus Linguistics conference 2005 in Birmingham*. Vol. 1, no. 1, <http://www.corpus.bham.ac.uk/conference2005/index.htm>.
- Tognini-Bonelli, Elena (2001). *Corpus Linguistics at work*. Studies in Corpus Linguistics 6. Amsterdam/Philadelphia: John Benjamins.
- Willners, Caroline (2001). *Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjective*. Travaux de Institut de Linguistique de Lund 40. Lund, Sweden: Dept. of Linguistics, Lund University.

Willners, Caroline & Anders Holtsberg (2001). Statistics for sentential co-occurrence. *Working Papers Lund, Sweden: Dept. of Linguistics, Lund University* 48, 135-148.

Woods, Anthony, Paul Fletcher & Arthur Hughes (1986). *Statistics in language studies*. Cambridge: Cambridge University Press

Appendix A

Translations of the Swedish words included in the test set.

Canonical antonyms		Antonyms	
långsam-snabb	<i>slow-fast</i>	långsam-flink	<i>slow-swift</i>
ljus-mörk	<i>light-dark</i>	tråkig-het vit-dunkel melankolisk-munter	<i>dull-hot</i> <i>white-obscure</i> <i>melancholic-merry</i>
svag-stark	<i>weak-strong</i>	lätt-muskulös senig-kraftig	<i>light-muscular</i> <i>sinewy-sturdy</i>
liten-stor	<i>small-large</i>	obetydlig-kraftig	<i>insignificant-sturdy</i>
smal-bred	<i>narrow-wide</i>	liten-väldig smal-öppen	<i>small-enormous</i> <i>narrow-open</i>
dålig-bra	<i>bad-good</i>	trång-rymlig dålig-god	<i>tight-spacious</i> <i>bad-fair</i>
tunn-tjock	<i>thin-thick</i>	ond-bra genomskinlig-svullen fin-grov	<i>evil-good</i> <i>transparent-swollen</i> <i>fine-rough</i>
Synonyms		Unrelated	
långsam-släpig	<i>slow-draggy</i>	het-plötslig	<i>hot-sudden</i>
snabb-rask	<i>fast-rapid</i>		
ljus-öppen	<i>light-open</i>	dyster-präktig	<i>sad-decent</i>
mörk-svart	<i>dark-black</i>		
svag-matt	<i>weak-flat</i>	flat-seg	<i>abashed-tough</i>
stark-frän	<i>strong-sharp</i>		

stor- inflytelserik	<i>big-influential</i>	klen-kort	<i>delicate-short</i>
liten- oansenlig	<i>small- insignificant</i>		
smal-spinkig	<i>narrow-skinny</i>	liten-tjock	<i>small-thick</i>
bred-kraftig	<i>wide-sturdy</i>		
dålig-låg	<i>bad-low</i>	fin-tokig	<i>fine-crazy</i>
bra-god	<i>good-fair</i>		
tunn-spinkig	<i>thin-skinny</i>	knubbig-tät	<i>chubby-dense</i>
tjock-kraftig	<i>thick-sturdy</i>		

Appendix B

Results from the elicitation experiment. The stimulus words are in bold and the responses are listed to the right of each stimulus. Non-answers are marked with “0”.

bra dålig
liten stor
ljus mörk
låg hög
mörk ljus
sjuk frisk
smutsig ren
stor liten
vacker ful
dålig bra frisk
kort lång långsam
ond god snäll
smal tjock bred
stark svag klen
svag stark klar
tunn tjock bred
öppen stängd sluten
bred smal tunn snäv
långsam snabb fort seg
lätt tung svår svårt
svart vit gul vital
tjock smal mager tunn
vit svart gul mörk

trött pigg utvilad vaken energisk
fin ful grov ofin trasig dålig
ful snygg vacker fin söt regelmässig
glad ledsen sur arg trumpen sorgsen
hård mjuk svag god slapp lös
klen stark kraftig muskulös stabil grov rejäl
rymlig trång liten orymlig snäv tråkig kompakt
snabb långsam slö sen sakta sölig senfärdig
dunkel ljus klar upplyst tydlig genomskinlig stark ljusaktig
god ond äcklig dålig elak oaptitlig dum illa
pigg trött slö sömnig långsam lat sjuk seg
böjlig stel oböjlig styv rak hård utjämnad stum
het kall sval mesig kylig mjuk mild svag ljum
hövlig ohövlig oartig orevlig oförskämd läskig klumpig ohyfsad ofin
nervös lugn säker självsäker stabil cool rofylld sansad modig
trång rymlig vid bred lös smal stor öppen volumiös
väldig liten obetydlig minimal oansenlig pyttlig jätteliten pytteliten oväldig
knubbig smal spinkig tunn slank mager långsmal benig trind gänglig
overksam verksam aktiv driftig flitig företagsam rastlös rask igång
hyperaktiv
rask långsam slö sölig trög sen sjuk släpig saktfärdig senfärdig
spännande tråkig ospännande ointressant trist enahanda menlös obetydlig
likgiltigt långtråkig odramatisk
tråkig rolig kul utåtriktad trivsamt spännande skojig underhållande gladlynt
intressant munter
upprörd lugn samlad oberörd sansad glad tillfreds cool vänlig balanserad
behärskad
fläckfri fläckig smutsig ren befläckt obefläckt besmutsad besudlad
sjavig dunkel fläckad kriminell
framfusig tillbakadragen blyg försynt avvaktande återhållsam reserverad
nervös feg blygsam diskret seg
muskulös spinkig klen tanig svag senig tunn otränad smal omuskulös
musklig mager
dyster glad munter uppåt rolig sprallig färgglad upprymd lycklig gladlynt
uppsluppen pigg uppspelt
genomskinlig ogenomskinlig tät synlig färgad matt täckande solid tjock
dunkel mörk opak 0
grov fin tunn klen smal len späd finlemmad slät gles mjuk tanig finkornig
matt blank pigg glansig klar stark kraftfull spänstig svag vital rapp energisk
livlig
obetydlig betydlig betydelsefull viktig inflytelserik betydande ansenlig
avsevärd synlig anmärkningsvärd framträdande värdefull stor
släpig snabb rask kvick rapp rak företagsamt osläpig upprätt smidig rinnande
stillastående pigg

spinkig tjock kraftig knubbig fet rund muskulös grov mullig korpulent
fyllig smal välbyggd

kraftig smal svag tunn klen lätt liten spinkig mager mesig kraftlös trind
slank tanig

flink långsam klumpig trög oflink slö fumlig lat ofärdig klumsig ohändig
sävlig tafatt senfärdig osäker

irriterad lugn glad nöjd oirriterad balanserad samlad tålmod snäll oberörd
tålig tillfreds vänlig tålmodig välvillig behärskad

munter dyster butter sur nedstämd ledsen deppig trumpen uppåt arg glad
missmodig melankolisk sorgsen trist munter pessimistisk

tät gles otät genomskinlig ihålig tunn fattig brett dragig hålig utspridd luftig
grov spretig lös lucker öppen

inflytelserik obetydlig betydelselös maktlös inflytelos oansenlig
obetydelsefull opåverkbar oansvarig försumbar ickeinflytelserik
opåverkad menlös utanför neutral obetydande mjäkig

tokig klok normal sansad frisk lugn smart redig tråkig vanlig glad mentalt
frisk förständig beräknelig vettig balanserad rolig resonlig rätt

förtjusande hemsk motbjudande otrevlig ful avskyvärd förskräcklig äcklig
förfärlig osympatisk odräglig vämjelig fördjävlig skurkaktig
intetsägande gräslig vidrig frånstötande ocharmig trist

förvirrad klar samlad koncentrerad sansad redig säker klarsynt medveten
lugn organiserad närvarande självsäker fokuserad orienterad ha koll
kontrollerad vettig insiktsfull harmonisk

senig muskulös mörkraftig knubbig tjock biffig rultig slapp korpulent tunn
flexibel mjuk svag plöfsig rulltig stark osenig grov mullig fläskig

frän god mild mjuklen söt ofrån behaglig smaklös svag ljuv frisk töntig
blygsam luktfri hövlig normal lågmäld gammalmodig neutral smakfull

oansenlig ansenlig iögonfallande betydande viktig prålig
uppteendeväckande berömd attraktiv färgstark märkbar synlig
dominant anständig extraordinär mycket betydelsefull praktfull vacker
formidabel trist frappant

plötslig långsam väntad långdragen förutsägbar planerad förberedd
förutbestämd förväntad besinnad sävlig oplötslig successiv utdragen
lugnt konstant varsel ihållande väntat långvarig efteråt senare seg

lugn orolig stirrig nervös upprörd stressad hispig aktiv stökig sprallig
upphetsad uppjagad stimmig störig stressig intensiv labil fartfylld
temperamentsfull uppspelt energisk rörig olugn upptrissad

melankolisk glad munter gladlynt lycklig uppspelt omelankolisk
närvarande uppåt gladsint upphetsad framåt euforisk nöjd lättsam
glättig dramatisk vaken skärpt förnöjsam vigorös uppsluppen livfull 0

svullen normal slät osvullen smal tunn jämn platt utmärglad hopdragen
ihopkrympt dämpad spänstig avtärd ihopsjunken klen spinkig slank
tanig insjunken liten flat mager 0

vågad feg ovågad försiktig blyg tillbakadragen städad säker rädd riskfri konservativ rak oansenlig präktig snäll lagom uniform återhållsam trygg sober diskret reserverad pryd

flat rund bucklig djup ojämn bubblig kuperad vågig tjock gropig kurvig bred böjd betydlig omedgörlig framhållande hög yppig oflat rant buktig bullig bergig spetig knagglig 0

seg mör pigg mjukoseg fast porös flytande snabb spröd fartfull tunnflytande lättflytande oböjlig hård rask lätt skör svag aktiv alert stel brytbar vek hyperaktiv rolig slapp lös böjlig

präktig slarvig opräktig ödmjuk dålig lössläppt spännande tunn oförskämnd osedlig lättsam vågad anspråkslös vulgär busig okonventionell opålitlig opretentiös bräcklig ohederlig vild slafsigt slapp bohemisk angenäm lös bondig upprorisk frivol osäker 0

sparsmakad prålig överdådig generös överdriven extravagant vräkig plåttig svulstig plottrig nöjd vidlyftig slösig rik expansiv kravlös osparsmakad slösaktig vågad glupsk tilltagen flådig vulgär utförlig stark vrusigt mycket påkostad provsmakad explosiv överväldigande allätare enkel 0