All antonyms are opposite, but some are more opposite than others
An empirical attempt to study the adjectival semantic opposition

Greta Gandolfi
CIMeC - Center for Mind/Brain Sciences - University of Trento
greta.gandolfi@studenti.unitn.it

12 January 2018

Abstract

In daily conversations, the negation of a predicate implies the truth of its nearest alternatives. This hypothesis is the core of the *conversational negation*, an exciting proposal by Kruszewski et al. (2016) that manages to combine this semantic theory of negation based on alternativehood with data from distributional semantics. The proposal works amazingly - both in a theoretical and a computational sense - with nouns.

In this work, I tried to apply this proposal to pairs of adjectives with opposite meaning. More precisely, I tried to use this concept of negation and alternativehood in order to test if the purely theoretical difference between *binary* and *non-binary antonyms* is intuitive and if it can be derived from distributional data. Due to this fact, I worked on two different levels: the level of intuitions and the level of distributional semantics.

First, I tested the intuitions of 26 non-native English speakers, and I tried to see if every adjective, when negated, elicit in speakers' mind the same pattern of alternatives. My project relies on the idea that the number of alternatives suggested by the negation of an adjective can be a factor to use in order to classify the relation between an adjective and its antonym. The simple intuition at the core of this reasoning is: fewer alternatives a negated adjective suggests, more its relation with its antonym is binary. In this context, I tried to test if other speakers share my intuition.

Secondly, I checked if the semantic difference between *graded* and *non-graded* (non-binary and binary) antonyms can be seen also from a distributional point of view, using the analogy task.

The aim of my project was to find, or - at least - to suggest empirical proves of the theoretic distinction between graded and non-graded antonyms. Unfortunately, both tests happened to be useless to detect this kind of semantic distinction.

1 Introduction

Formal Distributional Semantics is a field of research that analyzes the properties of language through the combination of formal and distributional semantics. From this perspective, the negation is one of the most challenging features of language, since the logical notion of negation hardly combines with distributional data, as Kruszewski et al. (2016) highlighted. They proposed a new account for linguistic negation that differs from the common notion of negation in logic[1].
Their hypothesis is based on the concept of alternativehood, and it is called *conversational negation*. Briefly, the negation of a predicate implies[2] the truth or the possibility of its nearest alternatives.

Suppose you are driving your car in the middle of the night and you see something in the street. You don't know what it is, but you think that it could be an unlucky dog. It is plausible that you will say something like *Oh my god, it was a dog!* and it is also plausible that someone in your car replies: *Oh, dear, it was not a dog! It was a cat!*. But you can be sure that it is not plausible at all for the person next to you to say *It was not a dog, it was a boat!*.
In a natural conversation like that one, when your partner applies the negation to the word *dog*, he or she is not saying that in the street there could be all the things that do not share the property *to be a dog*. When someone think or pronounce this type of sentence, probably he or she is trying to highlight the fact that there is something *similar* to a dog in the street, but definitely not a dog.

Distributional semantics can help to find out which are (or could be) the possible alternatives of the negated word *dog* searching for its nearest words in a semantic space. From a distributional point of view, the word *dog* is more similar to *cat*, to *squirrel*, to *mouse* (and to every mammal) than *boat* or *moon* or *screwdriver*. And *cat*, *squirrel* and *mouse* happen to be more plausible alternatives than *boat*, *moon* or *screwdriver* in this context.

Kruszewski et al. (2016) showed how cosine similarity turned out to be a very good indicator of alternativehood. They successfully tested this hypothesis on nouns and they suggested that this approach can work with nearly every word.

2 Motivation

Nouns, like some other grammatical categories, in many cases, do not seem to have intuitive opposites[3] ; so, it is clear that, when we apply negation to the noun *dog*, we are not meaning its opposite. A question like: *What is the opposite of a dog?* does not even make sense, while the question *What is the opposite of false?*, does.

For this reason, at an intuitional level, it could be plausible that adjectives do not behave exactly like nouns. It is possible to assume that the negation of an adjective does not imply the possibility of its nearest alternatives, but that it implies just an an-

[1] In logic, if the meaning of the word *dog* is the set of all the dogs in a domain, the meaning of *no dog* is its complementary set, i.e. the set of all the things in a domain, except for the dogs.

[2] From now on, I will use the verbs *to imply*, *to suggest*, *to elicit (in speakers' minds)* interchangeably, with the awareness that they must have different meanings in other contexts.

[3] Of course there are exceptions: *entry-exit, truth-falsity, love-hate*, etc.

other adjective, with the opposite meaning. At least it can be true for *some* adjectives. Couple of adjectives with opposite meaning are called antonyms. In linguistic theories, antonyms can be distinguished in contrary and contradictory pairs. A pair of words with opposite meanings where the two meanings lie on a continuous spectrum (e.g. *hot-cold, young-old*) is defined as a pair of contrary antonyms; while pairs of words with opposite meanings, where the two meanings do not lie on a continuous spectrum (e.g. *true-false, occupied-vacant*) are defined contradictory antonyms[4]. If a cup tea is not cold, it could be hot, boiling, freezing or mild; otherwise a plant, if not dead, could be only alive. Now, it is plausible to conclude that contrary adjectives, when negated, imply more alternatives than contradictory ones. Following this kind of reasoning it might be thought that - from a distributional point of view - the negation of an adjective that share a contradictory relation with its antonym could be closer to it (its opposite) than the negation of an adjective that share a contrary relation with its antonym.
Aina et al. (2018) worked on the negation of antonyms in a semantic space and they proved that the an adjective, when negated, is more similar to itself than to its antonym. They distinguished between binary and non-binary ones, but the conclusion is the same.

Even if the negation is not useful to detect the different semantic relation between pairs of graded and non-graded antonyms, it can be still plausible that they would behave in a different way in a semantic space.
With these thoughts in mind, I tried to check if the theoretical distinction between binary and non-binary antonyms could have empirical proves[5], therefore:

- I tested the intuitions of a small group of English speakers;
- I tried to see if this different semantic relation (between graded and non-graded antonyms) can be computed in a semantic space.

3 Methodology

As I stated before, I tried to analyze the relation between antonyms at the level of intuitions and at a distributional level. For this reason, my project can be divided in two different phases: I will call the *Intuition Test* the first phase and the *Distributional Test* the second one.

3.1 Intuition Test

The hypothesis of conversational negation states that a word, when negated, implies the truth or the plausibility of its nearest alternatives. My idea is to use these concept of negation and alternativehood to test whether the theoretical distinction between contrary and contradictory antonyms can be reflected by *human intuitions* and *distributional data*, or not.

[4]These pairs of antonyms are also called binary and non-binary, graded and non-graded.

[5]Notice that I did not work properly on negation in distributional semantics. I found the topic very interesting and compelling, but maybe too intricate for my level of expertise. However, the topics of negation in distributional semantics has been my starting point and the reason I started to think about antonyms, so, in this section, I only explained the direction of my thoughts and the skeleton of my reasoning.

I had a strong intuition on the fact that, in daily conversation, binary antonyms (in this context, adjectives) when negated, imply the truth of less alternatives than the non-binary ones.
I needed to check if my intuition was shared by other people, so I asked 26 non-native English speakers to participate at a survey[6] in which I asked them to count how many alternatives were elicited in their mind[7] by the negation of some contrary and contradictory antonyms.
More precisely:

- I used a simple dataset: I chose randomly 50 antonyms from the WordNet set of antonyms;
- I created a list of simple incomplete sentences of this type: *If it is not x, it is...*, where x correspond to one antonym for every pair of antonyms;
- I asked the speakers to think about how many different possibilities would come up to their mind if the task was to complete the sentences. If they could think of only one possibility, they were asked to write down that possibility;
- Looking at the results I checked wether the adjectives that, when negated, elicit only one alternative could be defined as binary, comparing them to an already built list of antonyms[8].

[6]The survey can be found here: goo.gl/forms/WWcG7V3eTDxXJyAV2.

[7]I though that, in this specific context, the number of implicatures could correspond with the number of alternatives that come to the mind of the speakers.

[8]Unfortunately, it was difficult to find proper lists of only binary antonyms, so I built a list of them using the dataset created by Aina (2018): `https://lauraina.github.io/data/notadj.pdf` and the wikipedia article Opposite (Semantics): `https://en.wikipedia.org/wiki/Opposite_(semantics)`, that both contain two small different lists of binary antonyms.

3.2 Distributional Test

In this second phase, my aim was to check if this wooly semantic distinction between classes of opposites can have a distributional counterpart. Vylomova et al. (2016) showed how analogies in semantic spaces can account for several semantic relations (they can detect hypernyms, meronyms, cause-effect relations, etc.).

I wondered if the analogy task could detect the different semantic relation between graded and non-graded antonyms[9]. Trying to verify this hypothesis:

- I built a semantic space using Gensim model Word2vec (Rehurek, 2010) [10], trained on Text8 corpus[11];
- I tried to compute some analogies between:
  - pairs of binary antonyms (pairs of antonyms of the same type that share the same semantic relation);
  - pairs of binary and non-binary antonyms (pairs of antonyms of different types that do not share the same semantic relation).

I mainly expected only the first kind of analogies to work (i.e. to give good and intuitive

[9]This is possible only if the analogies are capable of account for antonyms, in general. Later on, I will explain how this implicit premise fails.

[10]I used the Python implementation. These are the model's parameters: skip-gram model, window = 5, size = 200.

[11]All the information about the corpus are here: `http://mattmahoney.net/dc/textdata.html`

results).
Analogies seem to work if they combine words that share the same kind of semantic relation. So, the idea at the core of this test is: if the analogies between pairs of the same kind of antonyms[12] work, and the mixed ones do not, maybe it can be a reason to think that they share a different semantic relation.

Notice that, in this context, the semantic difference could be extracted from the data and not derived a priori.

4 Results

4.1 Intuition Test

In the Appendix section, I reported the *Table 1* that summarizes the *Intuition Test*'s results. In the first column, it is possible to find the negated adjective[13].
The other columns report the number of alternatives elicited in the speakers' mind for every negated adjective.
As I expected, some adjectives suggested more alternatives than others, but only a few of them suggest only one alternative:

- *true* (and its alternative is *false*)
- *out* (and its alternative is *in*)
- *off* (and its alternative is *on*)
- *open* (and its alternative is *closed*)

Is it possible to state that these pairs of adjectives share a binary relation, and the rest of them do not? Maybe it is not.

These results are not compelling nor exciting, since the only binary pair of adjectives seems to be the first, *true* - *false*[14].
It is also true that, in those cases in which *out* and *off* are used as adjectives[15] in copula position (e.g. in common sentences like *The dog is out!* or *The lights are off.*), they seem to have only one possible alternative (e.g. the dog cannot be somethin-in-between *in* and *out*). For the couple *open-closed* the situation slightly differs; a door, if not *open*, can be *closed* but also *ajar*, while a store, if not *open*, can be only *closed*.

Moreover, other pairs of adjectives generally classified as binary (*alive-dead, male-female*) seem to suggest more than one alternative to the speakers.
Due to this fact, the relation I stated between antonyms and alternativehood seem not to be justified by other speakers' intuitions.

Nevertheless, some problems are probably not related with to that general idea and can depend on other factors.
First, the list of 50 adjectives resulted to be too small for the aim of the test and only a few of the cases proposed to the speakers are generally classified - a priori - as adjectives with a precise binary antonym (*true-false, alive-dead, male-female*).
Secondly, the task was not so easy to understand, so maybe some speakers were confused on what they were expected to do; but the main problem, as I mentioned before, seems to be the fact that the distinction between binary and non-binary antonyms is quite slippery and wooly.

[12] Notice that I didn't try to compute analogies between graded antonyms both at the left and right side of the equal sign of the analogies since I expected that they will not work because every pair has a different level of semantic gradation.

[13] The *x* in *if it is not x, it is...* sentence form.

[14] It is the only pair from the list above labeled as binary in Aina's and Wikipedia classifications.

[15] If it is possible to think about *out* and *off* as adjectives.

There is also a fundamental factor that can invalidate the test: the linguistic knowledge of the speakers. That's the reason why I told them to feel free to translate the examples in the test in their language and to search for the translation of words they did not know. However, it is plausibly impossible to delete the errors form human intuitions, since errors are purely human. Moreover, their intuition would be not pure and spontaneous if shaped a posteriori.

4.2 Distributional Test

To test the behaviour of antonyms in a semantic space, I created two lists of antonyms, separating binary adjectives from non-binary ones[16].

My idea was to compute two kinds of analogies (with only binary antonyms and with mixed - binary and non-binary - antonyms), expecting only the first kind to work[17].

First, I executed the analogies between pairs of only binary antonyms[18].

The fifteen analogies of the form *true: false = binary antonym : x* that I computed, do not seem to work. Only the combination of the two pairs *religious-secular* and *true-false* works, but it is the only exception. I tried other combinations - just to see if the problem can be the related to the *true-false* pair - and some of them worked, but it is a very small number of cases[19].

These results dropped the implicit premise of my reasoning, so, the computation of analogies between binary and non-binary antonyms lost all its meaning.

However, I tried to compute 28 mixed analogies[20], and nine of them worked, maybe because of the fact that antonyms - in semantic spaces - result to be near to the word with the opposite meaning. This happens because pairs of antonyms often occurs in same contexts. For this reason, it is possible that these analogies worked only by chance. In fact, they probably returned the expected word only because off its distributional proximity (similarity) to its opposite one.

This consideration leads to the main problem about antonyms in distributional semantics. They are extremely difficult to detect because they cannot be distinguished, relying of their distribution, from synonyms of the word with their opposite meaning. This can be the reason of the failure of the analogy task, since it is based on cosine similarity (a measure of distance of vectorial representation of words in a semantic space).

[16] The first list is a combination of adjectives labeled as binary from Aina's dataset and Wikipedia article: *Opposite (Semantics)*; the second one is composed by all the adjectives used in the intuition test (except from true-false, dead-alive, in-out since they are already present in the other list).

[17] An analogy work if it returns what I expected to be the *x*, in an analogy of the form $a : x = c : d$. In this case, what I expected to be the x is the antonym of a, according to WordNet list of antonyms. In this context, I decided to consider valid the analogies that return the expected antonym in the first 5 results.

[18] I decided to use as a benchmark the couple *true-false*, and to combine it with the other pairs. This couple of antonyms is often used as the main example of antonyms with a binary relation and, as it can be clearly seen in the *Table 1*, seems to be the most intuitive one.

[19] These results are reported in *Table 2* in Appendix.

[20] These results are reported *Table 3*, Appendix. The full code can be found here: `https://drive.google.com/open?id=1UUmgD_eJLzw_kij4awwvf0YCYX5O7103`

5 Conclusion

Non-graded antonyms can be distinguished from graded ones by the fact that the truth of one antonym implies the falsity of the other in a pair. This definition is intuitive and strongly theoretical; but is it possible to distinguish this different types of opposites from an empirical point of view?

I worked on two levels on analysis; first, I tried to test the intuition of some speakers starting from a general idea of alternativehood suggested by the *conversational negation*, then I checked if the different semantic relation between binary and graded antonyms can be seen from a distributional perspective (using the analogy task).

At the level of intuitions, I found out that some antonyms, when negated, imply (or maybe only elicit in the mind of the speakers) more possible alternatives than others. Some of them seem to lead to only one possible alternative, showing a kind of binary relation (e.g. *true-false, out-in, on-off*), but some others - theoretically defined as binary - seems to elicit more than one alternative (*dead-alive*). This analysis does not seem to demonstrate that all the intuitions correspond to the theoretical distinction, maybe because of the size of word tested, the number of speakers, and the idea to link the concept of alternativehood to this kind of semantic relation.

At the distributional level, the analogies seem to be incapable to detect any relation between antonyms. This can be a consequence of the type of data I choose and the number of analogies I computed but it can be mainly related to the fact that the distributional behaviour of antonyms seems to contradict by the distributional hypothesis.

References

[1] Aina, L., Bernardi R. and Fernandez, R. *A distributional study of negated adjectives and antonyms*. 2018.

[2] Kruszewski, D., Paperno, G., Bernardi. R. and Baroni, M. *There Is No Logical Negation Here, But There Are Alternatives: Modeling Conversational Negation with Distributional Semantics.* Computational Linguistics, Vol. 42, No. 4. 2016.

[3] Mahoney, M. *About the Test Data.* `http://mattmahoney.net/dc/textdata.html`. 2018

[4] Rehurek R. and Sojka, P. *Software Framework for Topic Modelling with Large Corpora.*, Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. 2010

[5] Vylomova, E., Rimell, L., Cohn, T., Baldwin, T. *Take and Took, Gaggle and Goose, Book and Read: Evaluating the Utility of Vector Differences for Lexical Relation Learning.* Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016

[6] Wikipedia contributors. *Opposite (semantics) - Wikipedia, The Free Encyclopedia*, `https://en.wikipedia.org/w/index.php?title=Opposite_(semantics)&oldid=864571930`. 2018

6 Appendix

6.1 Intuitions

adjective negated	no alternatives	more than 1 alternatives	one alternative
true	.4	.27	.69 (true)
out	.4	.27	.69 (in)
off	.12	.20	.64 (on)
able	.4	.68	.16 (unable)
right (position)	.3	.54	.35 (left)
dull	.8	.73	.15 (inconsistent answers)
fast	.4	.58	.38 (slow)
wet	.4	.58	.35 (dry)
late	.4	.62	.31 (early)
wary	.17	.71	.12 (trusting)
old	.0	.73	.11 (young) - .15 (new)
easy	.0	.69	.31 (difficult)
thin	.0	.81	.8 (thick) - .4 (fat)
oral	.4	.69	.27 (written)
alive	.3	.48	.40 (died) - .4 (coma)
foul	.19	.73	.8 (agreeable)
less	.3	.35	.58 (more)
kind	.0	.77	.18 (rude)
shut	.8	.58	.34 (open)
paid	.8	.58	.19 (unpaid) - .11 (free)
loud	.8	.73	.19 (quiet)
deep	.4	.76	.12 (shallow) - .8 (superficial)
male	.4	.42	.54 (female)
down	.0	.50	.46 (up)
good	.0	.81	.19 (bad)
open	.0	.42	.58 (closed)
deaf	.23	.50	.23 (hearing)
soft	.0	.69	.27 (hard) - .4 (rough)
poor	.0	.62	.35 (rich)
tidy	.8	.65	.27 (inconsistent answers)
rich	.0	.62	.38 (poor)
mild	.8	.77	.15 (strong)
sold	.11	.58	.31 (inconsistent answers)
wide	.4	61	.35 (inconsistent answers)
long	.4	.56	.40 (short)

Table 1: Results of the Intuition Test. Bold style for the adjectives that, when negated, elicit only

one precise alternatives for the major part of the speakers.

6.2 Analogies

Binary Analogies

binary pair	binary pair	good outcomes
true - false	present-absent	0
true - false	uniform-multiform	0
true - false	single-compound	0
true - false	native-foreign	0
true - false	optional-obligatory	0
true - false	separate-joint	0
true - false	private-public	0
true - false	religious-secular	1
true - false	dead-alive	0
true - false	blind-sighted	0
true - false	innocent-guilty	0
true - false	other-same	0
true - false	mortal-immortal	0
true - false	occupied-vacant	0
private - public	religious-secular	1
blind-sighted	innocent-guilty	1

Table 2: Results of analogies between binary antonyms (*Distributional Test*).
Good outcomes are represented by 1 in the last column, bad outcomes by 0.

Mixed Analogies

binary pair	binary pair	good outcomes
true - false	out-safe	0
true - false	off-on	0
true - false	able-unable	1
true - false	left-right	1
true - false	dull-sharp	0
true - false	fast-slow	1
true - false	wet-dry	1
true - false	wary-unwary	0
true - false	old-new	0
true - false	easy-uneasy	0
true - false	thin-fat	0
true - false	oral-anal	0
true - false	foul-fair	0
true - false	less-more	1
true - false	kind-unkind	0
true - false	shut-open	0
true - false	paid-unpaid	0
true - false	loud-soft	0
true - false	deep-shallow	0
true - false	male-female	1
true - false	up-down	0
true - false	open-closed	1
true - false	deaf-hearing	0
true - false	poor-rich	0
true - false	tidy-untidy	0
true - false	mild-intense	0
true - false	slow-fast	1
true - false	long-short	0

Table 3: Results of analogies between binary and non-binary antonyms (*Distributional Test*). Good outcomes are represented by 1 in the last column, bad outcomes by 0.