1. Definitional criteria of antonymy: a corpus-based approach

The term antonymy, in a restricted sense, is used to designate a binary and oppositional relation between logically contraries or semantically gradable terms (Lyons 1977). Lexical units (X and Y) are contraries if they cannot be predicate for the same object A at the same time and if this object may be neither X or Y, but the mid-term between them. For example, A cannot be called both hot and cold however it can be called neither hot or cold, but lukewarm or tepid. Lexical units are gradable if they "denote degrees of somevariable property", such as temperature (Cruse 1986: 204, see also Sapir 1944, A. & K. Lehrer 1982).

Linguistic definitions of antonymy, like these above, provide criteria to determine which lexical units are antonyms:
- it must exist a paradigmatical relation between these units (Murphy 2003);
- an opposition between all the senses of the units or some of them (for a distinction between "perfect antonyms" and "partial antonyms", see, for example, Duchêne 1965);
- the senses are poles on a gradable scale (see, for example, Ljung 1974);
- to be opposed, the poles must be symmetrical to a mid-term or a middle-area (Kat 1972, Chaffin & Herrmann 1984).

The aim of my PhD project is to confront these criteria with the daily use of co-occurring antonyms in written modern French. My hypothesis is that this confrontation will show that these criteria are not totally able to define antonymy in context.

For my purpose, I need a large and recent corpus. So I chose all the articles published in the French newspaper, Le Monde, from 2002 to 2011. This choice was motivated by the fact that newspaper corpora are less form-oriented than literary corpora. The risk of finding antonyms used to serve a stylistic effect is minimized. However it’s possible to use Frantext to compare the observation from a newspaper corpus to a literary corpus.

3. Case study: the chaud/froid scales

- Let’s examine the last criterion with the help of the corpus: to be opposed, the poles must be symmetrical to a mid-term or a middle-area. The common French example to illustrate this is the opposition between chaud and froid. They belong to two opposite scales.

- On the scale of froid, glacial (and its synonyms glacé, gelé) is the upper degree ("plus froid"), frais is the lower degree ("peu froid"). On the scale of chaud, brûlant (and its synonym bouillant) is the upper degree ("plus chaud") and tiède the lower ("peu chaud") (Berbinschi 2003: 44).

Chaud and froid are antonyms, so are brûlant and glaciel, but chaud and glacial like brûlant and froid are not, because they are not symmetrical with respect to the middle-area (frais and tiède) between the two adjoining scales.

4. Asymmetrical antonymy on the scales chaud/ froid: analysis of the corpus data

Consider the pairs brûlant / glacial and brûlant / froid, the first is symmetrical and supposed to be antonymic, the second is not symmetrical and is not antonymic. However my first observations show that the second pairs co-occur well, but very often in non-antonymic contexts. See for example the following sentence.


Is this really a lexicalized sense of brûlant or just a restricted and conventional meaning in discourse? First observations seem to give credence to the lexicalization hypothesis. Brûlent, when co-occurring with glacial tend to have the meaning of "très chaud" but when co-occurring with froid tend to have the meaning of "qui brûle". Further research on co-occurrence in corpus Le Monde will be able to address that question more precisely.

If brûlant is not the good candidate opposite for glacial, maybe bouillant is. This pair co-occurs, in antonymic context, very often. Because they are not symmetrical, froid and bouillant are not expected to co-occur in antonymic contexts. However, they sometimes co-occur as can be seen in the example on the right-hand side.

2. Extracting data from the corpus

To extract data from the corpus, I shall use the new software developed by Cyril Briquet (McMaster University, Hamilton, Canada) under the direction of Stéfan Sinclair. It was designed to enable search of co-occurrence between keywords taken from a predefined list.

This list was compiled from the antonymic pairs proposed by the Grand Robert de la langue française (2001). It could be enriched by references in other dictionaries. I chose to define the context of appearance of these antonyms as a number of sentences rather than as a number of words. It seems necessary for two reasons.

The first reason is that an antonymic pair, in a newspaper corpus, is often spread over two sentences. The distance calculated by the number of words between two antonyms could only find such pairs by chance. The second reason is that it seems less difficult to determine a priori the number of sentences beyond which two words are no longer co-occurring.

From my initial observations, a satisfactory context could be defined as: considering pair A/B, the context of a co-occurrence is defined as, from the sentence that contains A, the three sentences before and the three sentences after, with B being an element among one of these 7 sentences. It seems satisfactory for three reasons: 1) it tends to ensure a real semantic co-occurrence between two antonyms, 2) it minimizes as much as possible the risk of not detecting antonymous pairs co-occurring, 3) it can replace the pair A/B in a sufficiently large to understand what it is about.

Below, you can see a print screen of the software interface. It shows the different sub-corps I chose, the context width panel and one of the results it provides. The type list (list of lemmatized keywords to search in the corpus) and the stop list (list of words which are not considered as keywords) are defined by default following the lists I constructed but any other type list or stop list can be added. I am currently the only user of the software but in the future it may be implemented on the web platform Voyeur Tools (http://hermeneutica.ca/voyeur).

5. References


