

Regular Polysemy in WordNet

Lucie Barque (lucie.barque@linguist.jussieu.fr)
Lattice – Université Paris 7, UFRL, Case 7003, 75251 Paris cedex 5

François-Régis Chaumartin (fchaumartin@linguist.jussieu.fr)
Alpage – Université Paris 7, UFRL, Case 7003, 75251 Paris cedex 5

Abstract This paper presents an analysis and modelisation of polysemy in the WordNet English lexical database. It exploits the hierarchy of concepts (constituted by synsets), and the gloss defining each of these concepts. The result consists of a set of rules which enabled us to identify in a largely automated way, with a precision close to 91%, more than 2100 synset pairs, connected by a regular polysemy relation. Our method also allows a partial word sense disambiguation of the definition associated with these synsets.

Keywords: regular polysemy, metaphor, metonymy, WordNet, WSD

WordNet is a well-known semantic lexicon for the English language. This project offers a broad-coverage semantic network of the English language. Lemmas are grouped into sets of synonyms (called *synsets*), related by various semantic (hypernymy...) and lexical (morphological derivation...) relations. Each synset is defined by a gloss. For instance, the synset {cerise#1, cherry#4} is defined by “*the red color of cherries*”. This study concerns the description of regular polysemy (as described by Apresjan in *Regular Polysemy, Linguistics*, 1974) and its automatic detection within WordNet. For instance in {cerise#1, **cherry**#4} = “*the red color of cherries*” and {chestnut#4} = “*the brown color of chestnuts*”, we see that there is a semantic relation between a fruit and a color. Such a pattern can also be applied to other words denoting a fruit and a color, like LEMON, ORANGE...

Why might we wish to describe regular polysemy? First, it is an important source of lexical creativity. We could, for instance, produce a new sense of BANANA from the fruit meaning that denotes a variant of the color yellow. Moreover, we can introduce new types of lexical relation in WordNet, such as metonymy and metaphor.

Our method includes four steps:

1) We extract from WordNet “auto-referent synsets”, i.e. synsets whose definition includes one word that shares the form of one of the defined lemmas. For example, the definition of synset {cerise#1, **cherry**#4} is auto-referent since its definition contains the (ambiguous) noun CHERRY. Our hypothesis indeed is that such synsets are likely to be instances of regular polysemy. After this step, we obtain a set of 1984 synsets.

2) We then examine this set and manually produce about 60 regular polysemy patterns. Each pattern is constrained by the form of a definition (“*L₁ is the color of L₂*”...) and/or the type of L₁ and L₂ synsets (color/fruit, color/gem, color/metal...).

3) These patterns are then applied as follows: let L₁ be one of the 1984 synsets obtained during the first step; the system enumerates all synsets pairs (L₁, L₂) such that L₂ is another

sense of lemma L_1 and stops when it finds a synset L_2 with a type compatible with one of the type constraints that have been associated with a pattern (for instance `gem` or `metal`). At the end of this step, L_2 is fully disambiguated; for instance, we obtain `{toppaz#3}="the light brown color of topaz#2[gem]"`.

4) The third step produced 1427 instances of regular polysemy relations. The goal of the last step is to try to find other instances by dropping the “auto-referent” definition constraint; we only rely here on the type of the synsets pair (L_1 , L_2) provided by the patterns defined at step 2. This enables us to find 367 new instances, such as `{amber#1, gold#2}="a deep yellow color"` (with an implicit link to `gold[metal]`).

Many problems are faced in step 4. First, we cannot use synset types that are too general (such as `entity`, `artifact` or `abstraction`) since they could not represent true type constraints. Secondly, we try to handle the lack of “auto-reference” in certain definitions by introducing a new constraint: we require that the definition of L_1 and L_2 must share at least some common words; for this purpose, we use a gloss overlapping similarity measure. This constraint improves precision, but the price to pay is a lower recall; indeed, we do not select synset pairs that are really a metaphor or a metonymy when they do not share any significant word in common.

The result of our study consists of a full hierarchy of metaphor and metonymy, classifying all the patterns found in step 2, and including synset pairs that are instances of a regular polysemy relation. We manually evaluated the 2351 instances produced by the system, and found that 2140 were correct (precision close to 91%).

Recall was somewhat difficult to estimate, since there is no simple way to find exhaustively all regular polysemy instances; we looked at the *Person resembling Animal* metaphor pattern, examining all possible (L_1 =`Person`, L_2 =`Animal`) synset pairs, and also the *Wood produced from Tree* metonymy pattern. In the first case, our system found 36 instances out of a total of 142 pairs (25.3%) and in the second case 70 instances out of 79 (88.6%). As we can see, the method recall depends on the relation type, which can be more or less regular.

We have presented a method for extracting instances of regular polysemy from WordNet, using the noun and verb hierarchy and the synset definitions. We also propose a metaphor and metonymy hierarchy. This new information can be used as an aide for lexicography (to improve the form of definitions) and for word sense disambiguation systems (to infer new senses of words).