Regular Polysemy in WordNet

Lucie Barque, François-Régis Chaumartin

Lucie Barque, François Régis Chaumartin

Regular polysemy in WordNet

1 Introduction

The importance of describing regular polysemy in a lexicon has often been outlined, especially in the field of natural language processing (for a good overview of this issue, see (Ravin and Leacock, 2000)). Unfortunately, no existing broad-coverage semantic lexicon has been built following this relatively recent advice. And since producing a broad coverage semantic lexicon is a very time-consuming task, one has to put this idea into practice on existing lexicons. WordNet is an appropriate lexical semantic ressource for running this experiment as it is machine readable and has a wide coverage (Fellbaum, 1998). In this paper, we introduce a method to create regular polysemy patterns from WordNet data and to automatically detect their occurrences in the lexicon.

WordNet is structured as a hierarchy of synsets (sets of lexical units that are synonyms), representing lexical concepts. These concepts are linked by several kinds of semantic relations (hyperonymy, meronymy, antonymy, etc.). Each synset is also described by a lexicographic definition, as illustrated below (we denote by L#i the i^{th} lexical unit of a polysemic word in WordNet) :

- {treachery#2, betrayal#1, treason#3, perfidy#2} = *an act of deliberate betrayal*

As one can see in this example, the definition includes a word (*betrayal*) of the same form as one of the lexical units of the synset (betrayal#1). From now on, we will systematically distinguish the notion of **lexical unit** (one form associated to one meaning) from the notion of **word** (one form associated to possibly several meanings, in other terms, a polysemic unit). So we will say that the lexical unit `betrayal#1` is defined by another lexical unit of the word BETRAYAL but we don't know which one until the words of the definition are disambiguated[1]. In fact, WordNet contains implicit relations between pairs of lexical units that share the same form. Our goal is to make these relations explicit, and, when it is relevant, to generalize them in order to account for regular polysemy relations.

This paper consists of three sections. The first section defines the notion of regular polysemy and gives a quick overview of studies devoted to the description of regular polysemy in WordNet. The second one exposes our goals and explains our method to reach them, which consists in the "computer-assisted" creation of polysemy patterns.

[1] One has to mention the "eXtended WordNet" project, developed in 2003 at the University of Dallas (http://xwn.hlt.utdallas.edu/index.html). This project enhances WordNet 2.0 with a logical representation and a syntactic analysis of the definitions associated to the synsets. After the syntactic analysis, words from the definitions are disambiguated, with a *gold*, *silver* and *normal* quality (*gold* quality representing a manual evaluation). Since only 14 715 words out of 631 684 open-class words are gold quality (2,33%), we decided not to use this resource in our study.

The last section presents our results, which take the form of a classification of these patterns and a measure of their regularity in the most recent version of WordNet.

2 Regular polysemy in WordNet : an overview

In this section, we will first define the important notions our study is based on, mainly the notion of regular polysemy. We will then explain why it is important to describe this phenomenon in the lexicon. Finally, we will give a quick overview of studies that have been devoted to the description of polysemy in WordNet.

2.1 A definition of regular polysemy

We follow J. Apresjan's definition of regular polysemy :

> " Polysemy of a word A with the meaning a_i and a_j is called regular if, in the given language, there exists at least one other word B with the meaning b_i and b_j, which are semantically distinguished from each other in exactly the same way as a_i and a_j and if a_i and b_i, a_j and b_j are non-synonymous." (Apresjan, 1974):16

For instance, polysemy of the word CHERRY, with the meaning **`fruit`** and **`color`**, is regular since there exists at least one other word, CHESTNUT, that also has the meaning **`fruit`**[2] and **`color`**, as illustrated below with data from WordNet:

- {cerise#1, **cherry**#4} = *the red color of **cherries***
- {**chestnut**#4} = *the brown color of **chestnuts***

J. Apresjan gives us another important definition of a particular type of regular polysemy, that he calls **productive** and that we will call **systematic**, following the authors who have addressed this issue (Pustejovsky, 1995), (Buitelaar, 1998), among others.

> We will call a given type 'A' – 'B' of regular polysemy *productive*, if for any word which has the meaning 'A' it is true that it can be used also in the meaning 'B' (if 'A', then 'B') [...] Consequently, productivity is determined only by totality of scope of the units with the given combination of properties; the class itself of such units may be very small. (Apresjan (1992): 214)

The regular polysemy illustrated above between a lexical unit that denotes a `color` and a lexical unit that denotes a `fruit` is not systematic. Indeed, the example given under (1c) shows that it is difficult to generate a lexical unit that denotes a `color` from every lexical unit that denotes a `fruit`.

[2] Note that this data doesn't indicate that the word used in the definition denotes a `fruit`, since "*cherries*" and "*chestnuts*" are not disambiguated.

(1) a. *I like your **cherry** shirt.*

b. *I like your **banana** shirt.*

c. ? *I like your **pear** shirt.*

On the other hand, the regular polysemy between a lexical unit that denotes a `dish` and a lexical unit that denotes the `quantity` of food contained in this dish is a systematic one since you can derive the `quantity` (of food) sense from every word that denotes a `dish` (*a **bowl** of rice, a **spoon** of flour*, etc).

Between the minimal condition of regularity (at least two words that present the same sense alternation) and the systematicity (every word that denotes an `X` can also denote an `Y`), there is a large scale of regularities that a lexicon should be able to take into account.

It is one thing to consider the type of regularity of a given polysemy relation. Another thing is to consider the nature of this polysemy relation. In our study, we will consider the three following well-known categories of regular polysemy :

1. **Specialization** : a lexical unit L2 is a specialization of a lexical unit L1 if its meaning is more specific than that of L1. In the example given below, the lexical unit `pressure#7` denotes a particular kind of the pressure denoted in the definition.
   - {**pressure**#7} = *the **pressure** exerted by the atmosphere*
2. **Metaphor** : two lexical units L1 and L2 are in a metaphoric relation if L1 and L2's referents are in a relation of analogy, that is, if they are similar in almost one aspect (Johansen and Larsen, 2002). For example, the laugh denoted by `cackle#3` resembles a *hen's cackle*, as stated in the definition.
   - {**cackle**#3} = *a loud laugh suggestive of a hen's **cackle***
3. **Metonymy** : two lexical units L1 and L2 are in a metonymic relation if L1 and L2's referents are in a relation of contiguity, in other words if the two referents are "in contact", in the concrete or in the abstract sense of the word "contact". For example the relation between the two meanings of chestnut (`color` and `fruit`) is a metonymic relation since the color denoted by `chestnut#4` is the color of the fruit denoted by CHESTNUT in the definition.
   - {**chestnut**#4} = *the brown color of **chestnut***

2.2 Description of regular polysemy

The description of regular polysemy, by means of lexical rules (Ostler and Atkins, 1991), (Copestake and Briscoe, 1995) or by means of generative mechanisms during composition (Pustejovsky, 1995), presents at least two benefits.

From a theoretical point of view, the description of regular polysemy accounts for lexical creativity (Pustejovsky, 1995). For example, a lexicon that contains the description of the metaphoric regular polysemy relation between a lexical unit that denotes the fact of `tying an animal/person's part of the body` and a lexical unit that denotes the fact of `preventing someone from doing something` (regular polysemy that concerns TO MUZZLE, TO HOBBLE, TO GAG, etc) potentially contains the description of a new meaning. For example, the meaning of HANDCUFF, illustrated in (2)[3] below, is not listed in the consulted dictionaries (WordNet, Longman and Cobuild) and can thus be considered as a lexical creation. Having the description of the rule that has generated this meaning gives us information about this new use of HANDCUFF.

(2) *How can we build a 'Knowledge economy' if research* ***is handcuffed*** *?*

The second benefit, a more practical one, concerns a better methodology to develop the lexicon on which is based the description of regular polysemy (in our case, WordNet). Indeed, the description of regular polysemy by means of lexical rules allows for a more systematic encoding of the lexical data since they provide an underspecified definition to the lexicographer (Ostler and Atkins, 1991). For example, the underspecified definition, illustrated below in bold letters, can be used to define other lexical units of type `quantity` (*of food contained in a dish*), as illustrated below :

- L2 of X = **quantity of X contained in** L1
- plate#2 of X = **quantity of X contained in** plate#1
- bowl#2 of X = **quantity of X contained in** bowl#1

2.3 Description of regular polysemy in WordNet

Several studies have already been devoted to the description of regular polysemy in WordNet. We present some of them here, explaining how they converge and how they differ from our study.

Corelex is an ontology of 126 semantic types, each of them representing a systematic polysemous class of nouns (Buitelaar, 1998). This ontology aims to account for underspecification in discourse analysis, the author assuming that underspecification is often due to systematic polysemy. The method that has been employed to build this ontology is the following : the first step consists in delimitating a set of 39 basic types from the top level synsets of the noun hierarchy (for example, `entity`, `human`, `animal`, `food`, `act`, `event`, `location`, etc.). In the second step, each lexical unit from a polysemic word is assigned one of these basic types. For example, the seven senses of BOOK can be reduced to two basic types `artifact` and `communication`. The third step consists in grouping all polysemic words that share the same

[3] This sentence has been found on the web.

alternation of basic types. For example, CATALOGUE, HOROSCOPE, PRESCRIPTION, etc. share with BOOK the '`artifact-communication`' alternation. This step gives rise to 1648 classes, reduced to 529 polysemous classes once classes that contain only one member are removed. This 529 polysemous classes are then grouped into 126 semantic types. For example, the semantic type `cae` puts together the following polysemous classes : "`act∼artifact∼communication`" (CHOREOGRAPHY, DECORATION, DEVICE), "`act∼artifact∼communication∼psychological_feature`" (CALL, CONSTRUCTION, PORTRAYAL), "`act∼artifact∼psychological_feature`" (ARCHITECTURE, DESIGN, HABIT, ...), "`artifact∼communication`" (BOOK, CATALOGUE, HOROSCOPE), etc.

The method is interesting for at least two reasons. Firstly the Corelex approach gives rise to more consistency among the assignments of lexical semantic structure, reducing the degree of polysemy that is known to be too fine-grained in WordNet (Fellbaum and Grabowsky, 2002). Secondly, the method has a broad coverage since the 126 semantic types cover around 40 000 nouns. On the other hand, the results are not evaluated so we have no idea wether these 40 000 polysemic words are really instances of polysemy classes described by the 126 semantic types. Another criticism that could be addressed regarding the semantic types is that they are very general and thus not very intuitive, in that they give little information on the type of polysemy relation.

Our study is closer to the one presented in (Peters, 2006) since the author exploits both the synset hierarchy **and** glosses associated to the synsets in order to select candidates for instantiations of regular polysemy. The first step consists in an automatic selection of candidates on the basis of systematic sense distribution of nouns (pairs of hypernyms that subsume the sense combinations of the words involved). For example, in two of their senses, the nouns LAW, ARCHITECTURE, LITERATURE, POLITICS and THEOLOGY fall under the pattern `profession∼discipline`. The second stage consists in characterizing semantic relations between the related word senses by analyzing the glosses that are associated with these pairs of word senses and their hypernyms. If a verb occurs between each pair of word senses, it is taken as the semantic relation that holds between them. For example, the predicate `is mastered by` describes the semantic relation that holds between pairs of word senses that fall under the pattern `profession∼discipline`. Such relations have been extracted for around 5000 candidates. The author gives several examples of relations but doesn't say how many there are. Nor does he give an evaluation of the data that have been extracted.

3 Goal and Method

Our main goal is to describe regular polysemy relations of WordNet and to automatically detect their occurrences in the database. We then propose to enhance the lexicon by providing new lexical relations, in this case metonymy and metaphor relations.

Our method has four steps. The first one consists in extracting "auto-referent" synsets, that is, synsets whose definition includes one word that shares the form of one of the defined lexical units. The second step consists in manually describing polysemy patterns from the observation of the 1984 synsets extracted in the first step. In the following step, we disambiguate, with the help of the patterns, the meaning of the polysemic word of the definition that is implied. The last step consists in generalizing the method to synsets that are not "auto-referent".

3.1 Extracting "auto-referent" candidates

We decided to first extract "auto-referent" synsets, that is, synsets whose definition includes one word that shares the form of one of the defined lexical units. This decision relies on the well known fact that a semantic link between two lexical units (L1 and L2) is more obvious if L2 is definied by L1. Here are a few examples of these "auto-referent" synsets extracted during the first step :

- {cerise#1, **cherry**#4} = *the red color of* ***cherries***
- {**driver**#3} = *a golfer who hits the golf ball with a* ***driver***
- {**falsify**#4} = ***falsify*** *knowingly*

Before extracting these auto-referent synsets, we automatically attribute morpho-syntactic tags to the words contained in the definitions[4] in order to make sure that L1 and L2 share the same part of speech. For instance, the two following synsets are not auto-referent since the defined lexical unit is a verb while the lexical unit used in the definition is a noun (in the case of GUN) and an adjective (in the case of SLOW)[5].

- {**gun**} = *shoot with a* ***gun***
- {**slow**#2, slow down, slow up, slack, slacken} = *become* ***slow*** *or slower*

We also eliminate (by identifying the three key words *equal*, *trade name* and *capital of*) synsets in which L1 corresponds to L2, as illustrated below :

- {kopec, **kopeck**, copeck} = *100* ***kopecks*** *equal 1 ruble in Russia*
- {sildenafil, sildenafil citrate, **Viagra**} = *virility drug name* ***Viagra***
- {Bern, Berne, **capital of Switzerland**} = *the* ***capital of Switzerland***

After this first step we obtained 1984 synsets that are likely to be occurrences of regular polysemy relations.

[4] For this task, we used the Antelope NLP framework (www.proxem.com) (Chaumartin, 2009).

[5] We have chosen to adopt this restriction to limit our first investigation of regular polysemy in WordNet. However, our method of description of regular polysemy can be applied to pairs of lexical units that share the same form but that do not belong to the same part of speach. These pairs are indeed numerous in English and are linked by regular semantic links, as illustrated above with the GUN example (manner verbs).

3.2 Describing polysemy patterns

The second step consists of manually describing polysemy patterns from the observation of the 1984 synsets extracted in the first step. The method we use to attribute a category of regular polysemy to an occurrence L1-L2 relies on two main criteria applied to WN definitions (Martin, 1972), (Fass, 1988) : the position of the inclusion of L1 in the definition of L2 and the elements from the definition that are pertinent with respect to the polysemy relation. We will first present these two criteria and explain how they interact in the attribution of a regular polysemy category to a given occurrence. We will then present some examples of polysemy patterns on which our study is based.

3.2.1 Elements of the polysemy patterns

First of all, we consider the position of L1 in the definition of L2. Indeed, L1 can be included in the first part of the definition (and then be the *genus* of the definition), as illustrated by BEHAVE below, or in the rest of the definition (and be one of the *differentiae*), as illustrated by SWEEP :

- {**behave#3**} = ***behave*** *well or properly*
- {**sweep#6**} = *clean by* ***sweeping***

In addition to the position of L1 in the definition of L2, we extract sub-strings that are recurrent in the definitions. Here are three examples of inclusion in the second part of the definition distinguished by sub-strings that introduced them :

- {**mint#5**} = *a candy that is flavored with a* ***mint*** *oil* (→ "that is flavored with L1")
- {**bluefish#2**} = *fatty bluish flesh of* ***bluefish*** (→ "flesh of L1")
- {**fin#5**} = *a stabilizer on a ship that resembles the* ***fin*** *of a fish* (→ "that resembles L1")

Merging these two criteria (position of the inclusion and elements of the definition), one can automatically attribute a category (specialization, metaphor or metonymy) to a given occurrence. If L1 is in the first part of the definition of L2, the relation is a specialization or a metaphor. The two examples given below both show an inclusion of L1 in the first part of the definition but the first one is a specialization and the second one is a metaphor.

- {**arrange#5**} = ***arrange*** *attractively*
- {**grow#9**} = ***grow*** *emotionally*

If L1 appears in the second part of the definition, the relation is a metonymy or a metaphor, according to the element that introduces this inclusion. Among the three examples given above, (TO) SWEEP[6], MINT, and BLUEFISH are examples of metonymy whereas FIN is an example of metaphor. The ambiguity of the category can be resolved by definitional patterns. For instance, if the inclusion is preceded by "*that resembles*", as in the definition of `fin#5`, it means it's a metaphor, not a metonymy.

3.2.2 Examples of polysemy patterns

Let us now turn to three concrete examples taken from our 60 polysemy patterns[7]. First, we present a few lines of code implementing the `colorOf` pattern :

```
patterns.Add(new Pattern("colorOf")
    .AddType("color","fruit")
    .AddType("color","gem")
    .AddType("color","metal")
    .AddMatchingRule("color of *")
```

The first line of code defines the polysemy pattern. The three following lines indicate that the pair of lexical units that are likely to be applied this pattern are of types `color` for L1 and `fruit`, `gem` or `metal` for L2. The last line says that the definition of L2 must include the string `colorOf` to be detected as an occurrence of this pattern. Another kind of pattern is the one presented below :

```
patterns.Add(new Pattern("causedBy")
    .AddMatchingRule("resulting from *")
    .AddMatchingRule("caused by *")
```

Unlike the previous pattern, this pattern does not constrain the type of L1 and L2 but only imposes that L2 contains in its definition the string `resulting from` or the string `caused by`. Indeed, the "`caused by`" relation may gather instances of too general types. For example, "`disease` is caused by `organism`" (ERGOT) and "`shape` is caused by `happening`" (DISTORTION) could only be generalized by "`something` is caused by `something`", which is too general to be relevant. As one can see, the sole exploitation of the hierarchy of synsets (that provides types for L1 and L2) would lead to missing some regular polysemy relations (detected only by patterns extracted from definitions).

The application of these polysemy patterns to the 1984 candidates allows us to extract 1427 occurrences of polysemy relations. The other 557 synsets correspond to orphan occurrences of polysemy, that is, to pairs of lexical units that are not associated with **regular** polysemy relations. For example, the pair of lexical units given below is considered as an orphan occurrence of polysemy since the metonymy pattern "`document` written on `paper`" has no instance in WordNet except PAPYRUS.

[6] Note that the verb **sweep**#6 means *to clean (in a certain way)*, not *to sweep (in a certain way)* and thus cannot be considered as an instance of specialization, as it is defined above (see section 2.1).

[7] This patterns can be found online ([8]).

- {**papyrus**#1} = *paper made from the papyrus plant by cutting it in strips and pressing it flat; used by ancient Egyptians and Greeks and Romans*

- {**papyrus**#3} = *a document written on* ***papyrus***

3.3 Disambiguating L1 in the definition of L2

At this stage, L1 is not disambiguated in the definition of L2. The three following examples are detected as occurrences of the `colorOf` pattern, but we don't know if L1 is typed `fruit`, `gem` or `metal` :

- {**emerald**#3} = *the green color of an* ***emerald***

- {tan#2, **topaz**#3} = *a light brown, the color of* ***topaz***

- {**copper**#4} = *a reddish-brown color resembling the color of polished* ***copper***

The patterns based on pairs of types are used to disambiguate L1 in the definition of L2. The method is the following : the system enumerates all the possible types for L1 (other than L2) and stops when the couple L1 L2 matches one of the pairs of types expressed by the pattern. When the situation requires it, the system explores the hierarchy of hypernyms for nouns and verbs). L1 is then disambiguated, as illustrated below :

- {emerald#3} = *the green color of an* ***emerald#1***$_{[gem]}$

- {tan#2, topaz#3} = *a light brown, the color of* ***topaz#2***$_{[gem]}$

- {copper#4} = *a reddish-brown color resembling the color of polished* ***copper#1***$_{[metal]}$

3.4 Generalizing the application of patterns

Finally, we look for new candidates by dropping the "autoreferent" definition constraints. We only rely here on the type of the synsets pair (L1, L2) provided by the patterns defined at step 2. This enables us to find 367 new instances, two of which are illustrated below. The word GOLD has five meanings in WordNet : `coins`, `color`, `metal`, `great wealth` and `something that is precious`. Two meanings are typed `color` and `metal` and can thus be associated with the rule `ColorOf`, even if the definition of *gold#2* doesn't include the word GOLD.

- {amber#1, gold#2} = *a deep yellow color* (implicit link to *gold#3*[metal])

- {coral#1} = *a variable color averaging a deep pink* (implicit link to *coral#2*[gem])

This generalization needs to be refined. Indeed, results of the generalization are good with some pattern, but bad with patterns constrained by types that are too general, like `entity`, `artefact`, `abstraction`, ...: the lack of constraints on the iclusion of L1 in the definition of L2 leads to the multiplication of candidates that do not belong to regular polysemy classes. For that reason, we require that the definition of L1 and L2 share at least some words by using a gloss overlapping similarity measure, with a TF-IDF weighting[9].

- {footstep#1} = *the sound of a* ***step*** *of someone* ***walking***
- {footstep#2} = *the act of taking a* ***step*** *in* ***walking***

This constraint improves precision, but the price to pay is a lower recall; we miss metaphor or metonymy instances that do not share any significant word. For example, our system correctly identifies the word TIGER as an instance of the metaphorical rule between an `animal`, `person` (based here on the analogy with the ferocity of the animal). But the system doesn't select it since the two definitions do not share any significant words in common, as illustrated bellow:

- {tiger#1} = *a fierce or audacious person*
- {tiger#2} = *large feline of forests in Asia having a tawny coat with black stripes*

4 Results

The results of the research reported here are of two kinds: a descriptive one with the classification of regular polysemy relations, and a methodological one, with the automatic detection of occurrences of regular polysemy relations with a rather good precision.

4.1 Lexical description : classification of regular polysemy relations

In this section, we propose a classification of regular metonymy and regular metaphor relations based on the patterns identified by our study. Note that our results also include a significant proportion of specialization instances (for example, {**pressure**#7} = *the* ***pressure*** *exerted by the atmosphere*). Nevertheless, this kind of relation is not appropriate for a hierarchical organization, since it is difficult to identify regular types for L1 and L2. Moreover, instances of specialization are already described in WN by the hypernymy relation.

For each class of regular polysemy relation presented here, we indicate in brackets the number of real instances compared to the number of detected candidates. For example,

[9] The *Term Frequency-Inverse Document Frequency* is a method of weighting usually used in text mining. This statistical measure allows us to evaluate the importance of a word in a definition. The weight increases proportionally according to the number of occurrences in the definition. The weight also varies according to the number of occurrences in all the definitions.

the regular metonymy relation "`playing card` **represents** `person` or `entity`" has six candidates, but only five of them are real instances of this relation[10] (5/6). Indeed, an `ace#2` is not a card that represents an `ace#3`, as illustrated below.

- {**ace**#2} (playing card) = *one of four playing cards in a deck having a single pip on its face.*
- **ace**#3} (person) = *someone who is dazzlingly skilled in any field*

This proportion is followed by a few examples of the given regular relation.

4.1.1 Classification of regular metonymy relations

L2 **represents** L1
- → `playing card` **represents** `person` or `entity` (5/6 ; QUEEN, KING ; TEN, NINE)

L2 **is caused by** L1
- → `fee` **is caused by** `action` (27/27 ; ADMISSION, ANCHORAGE)
- → `disease` **is caused by** `organism` (13/17 ; ERGOT, HERPES)

L2 **is produced by** L1
- → `sound` **produced by** `instrument` or `movement` or `device` (15/15 ; DRUM, WHISTLE ; SNAP ; BELL)
- → `piece of work` **is written by** `person`
  - → `book` **written by** `writer` (no example in WN; it could be SHAKESPEARE)
    - → `book` **written by** `prophet` (15/15 ; JOB, JEREMIAH)
  - → `music` **written by** `composer` (9/9; MOZART, WAGNER)
- → L2 **is produced by** `plant` or `tree`
  - → `fruit or plant material` **produced by** `tree or plant` (120/128 ; GUM, COTTON ; ORANGE, CITRUS)
  - → `flower` **produced by** `plant` (50/51 ; CHRYSANTHEMUM, COTTONWEED)
  - → `vegetable` **produced by** `plant` (13/13 ; BEAN, RADISH)

L2 **produces** L1
- → `business firm` **produces** `publication` (2/2 ; NEWSPAPER, MAGAZINE)

L2 **is from** L1
- → `beverage` **is from** `region` (4/4 ; CHAMPAGNE, CHABLIS)
- L2 **is derived from** L1
  - → L2 **is derived from** `animal`
    - → `flesh` **from** `animal, fish, bird or crustacean` (303/303 ; RABBIT ; TROUT ; PHEASANT ; LOBSTER)
    - → `animal skin` **from** `animal` (17/17 ; FOX, CHINCHILLA)
    - → `wool` **from** `animal` (2/2 ; ALPACA, VICUNA)
  - → L2 **is derived from** `leaf, plant, tree...`
    - → `drink` **derived from** `leaf` (3/3 ; TEA, MATE)
    - → `fiber` **derived from** `plant` (13/13 ; COTTON, FLAX)

[10] The reader should keep in mind that the candidates are selected both from the patterns and from a generalization of these patterns, generalization that relies on the semantic type of L1 and L2 (see section 3.4 above). In most cases, "wrong" instances are selected during the generalization. Nevertheless, it could be interesting to study this lack of precision more precisely.

→ `wood` **derived from** `tree` (70/70 ; BAMBOO, BALSA)
→ `wine` **derived from** `vine` (2/2 ; TOKAY, VERDICCHIO)

L2 **is about** L1
→ `discipline` **is about** `abstraction` (56/64 ; PHILOSOPHY, PHYSICS)
→ `division` **is responsible for** L1 (4/6 ; EDUCATION, ENERGY, TRANSPORT)
→ `book` **is about** `person` (6/6 ; JONAH, JOSHUA)

L2 **accompanies** L1
→ `music` **that accompanies** `dance` (32/32 ; POLKA, MAZURKA)

L2 **covers** L1
→ `cloth covering` **that covers** `body part` (14/15 ; ELBOW, KNEE)

L2 **is included in** L1
→ `substance` **included in** `medicine` (17/17 ; ARNICA, MENTHOL)
→ `person` **member of** `group` (37/39 ; SAMURAI, NINJA)
→ `person` **that is in** `construction` (6/6 ; BUILDING, FLOOR)
→ `quantity` **contained in** `container` (39/39 ; TEASPOON, BAG)
→ `quantity of food` **contained in** `dish` (5/5 ; PLATE, CASSEROLE)
→ `river` **passes by** `state` (6/6 ; ALABAMA, DELAWARE)
→ `country` **on** `island` (22/22 ; IRELAND, MALTA)

L2 **is typical of** L1
→ `ball` **for** `game` (7/7 ; PAINTBALL, VOLLEYBALL)
→ `wine` **from** `region` (4/4; CHABLIS, BORDEAUX)
→ `color` **typical of** L1 (7/7)
→ `color` **typical of** `gem` (TOPAZ, EMERALD)
→ `color` **typical of** `metal` (GOLD, COPPER)
→ `color` **typical of** `fruit` (CHERRY, CHESTNUT)
→ `food` **tastes** L1 (13/25)
→ `food` **that tastes** `herb` (MINT, RATAFIA)
→ `part of garment` **characterized by** `part of the body` (12/14 ; BACK, SHOULDER)
→ `person` **characterized by** L1
→ `sportsman` **characterized by** `position` (31/31 ; CENTER, WINGBACK)
→ `singer` **characterized by** `voice` (11/11 ; CONTRALTO, SOPRANO)

`language` **spoke by** `person` (199/223 ; KOREAN, PORTUGUESE)

4.1.2 Classification of regular metaphor relations

→ L2 **is similar to** L1
→ `human communication` **similar to** `animal communication` (3/4 ; TO BARK, TO CACKLE)
→ `animal part of the body` **corresponds to** `human part of the body` (3/3 ; LEG, THROAT)
→ `person` **'s behaviour resembles** `animal` (36/54 ; PIRANHA, POPINJAY)
→ `object` **'s form resembles** `natural object` (38/38 ; MOON, SNAKE)
→ `artifact` **resembles** `part of the body` (5/5 ; NOSE, THROAT)

4.2 Evaluation of the automatic process

As far as we know, there is no gold standard for this kind of experiment. We manually evaluated the 2351 occurrences of regular polysemy relations proposed by our system. We estimate that 2140 are correct, that is a precision of 91,03%.

We didn't find any method that allows a precise automatic evaluation of recall. However, we manually evaluated the recall for two of the regular polysemy patterns presented above: the metaphoric relation `person resembles animal` and the metonymic relation `wood derived from tree`. We manually identified 142 occurrences of the metaphors in WN (recall of 36/142=25,3%) and 79 occurrences of the metonymy (recall of 70/79=88,6%). As one can see, the recall also depends on the nature of the relation, that can be more or less regular.

5 Conclusion

In this paper we propose a method to automatically extract, with a good precision, new lexical relations in WordNet : metonymy and metaphor relations. The resource containing these new relations is available online [11]. Our results can be used in a lexical disambiguation task to infer meanings that are not described in WN. For example, in WN, the word BORDEAUX denotes both the location ({bordeaux#1} = *a port city in southwestern France*) and the wine ({bordeaux#2, bordeaux wine#1}) but BOURGOGNE denotes only the location. Our patterns could be dynamically used to create new interpretation, when the context requires it (*They convince you to drink Bourgogne*).

Our method has been tested on English WordNet but could be applied to other semantic data for the English language, for example to FrameNet data (Fillmore et al., 2003). Indeed, FrameNet also provides an hierarchy of concepts that can be expressed by lexical units (the frames) **and** definitions that analyse the meaning of these lexical units. It could be interesting to compare the result of our method applied on WordNet with those of the same method applied to FrameNet, to see which regular polysemy patterns they share and to see which instances we find for each pattern.

Our method can also be applied to WordNets that are devoted to other languages (once the patterns description is adapted to the language) to help the systematic encoding of definitions and to compare regular polysemy relations that are shared by different languages. This however involves that the ressources propose definitions for each lexical units.

References

Apresjan, J. (1974). Regular Polysemy. *Linguistics*, 142:5–32.

Apresjan, J. (1992). *Lexical semantics.* Karoma Publisher, Ann Arbor.

Buitelaar, P. (1998). Corelex : An ontology of systematic polysemous classes. In *Proceedings of FOIS98, International Conference on Formal Ontology in Information Systems*, Amsterdam.

Chaumartin, F.-R. (2009). Antelope : une plateforme industrielle de traitement linguistique. *Traitement Automatique des Langues*, 49.

Copestake, A. and Briscoe, T. (1995). Semi-productive Polysemy and Sense Extension. *Journal of Semantics*, 1:15–67.

Fass, D. (1988). Metonymy and Metaphor : What's the difference? In *Proceedings of Coling-88*, pages 177–181.

[11] http://www.chaumartin.fr/download/wnpolysemy.zip

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.

Fellbaum, C. and Grabowsky, J. (2002). Polysemy and the (mental) lexicon. In Lenci, A. and Tomaso, V. D., editors, *Exploring the lexicon. Theory and Computation*. Edizioni dell'Orso, Alessandria.

Fillmore, C., Johnson, C., and Petruck, M. (2003). Background to Framenet. *International Journal of Lexicography*, 16:235–250.

Johansen, J. and Larsen, S. (2002). *Signs in Use. An introduction to semiotics*. Routledge, London/New-York.

Martin, R. (1972). Esquisse d'une analyse formelle de la polysémie. *Travaux de linguistique et de littérature*, 10:125–136.

Ostler, N. and Atkins, B. (1991). Predictable meaning shift : Some linguistic properties of lexical implication rules. In Pustejovsky, J. and Bergler, S., editors, *Lexical Semantics and Knowledge Representation : First SIGLEX Workshop Proceedings*. Springer-Verlag, Berlin.

Peters, W. (2006). In Search for More Knowledge : Regular Polysemy and Knowledge Acquisition. In *Proceedings of GWC2006*.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press, Cambridge.

Ravin, Y. and Leacock, C. (2000). Polysemy: An overview. In Ravin, Y. and Leacock, C., editors, *Polysemy, Theoretical and Computational Approaches*. Oxford University Press, Oxford.