IJCLA VOL. 4, NO. 1, JAN-JUN 2013, PP. 29–52
RECEIVED 13/12/12 ACCEPTED 11/01/13 FINAL 21/03/13

Solving Specialization Polysemy in WordNet

ABED ALHAKIM FREIHAT,[1] FAUSTO GIUNCHIGLIA,[1] AND BISWANATH DUTTA[2]

[1] *University of Trento, Italy*
[2] *Indian Statistical Institute (ISI), India*

ABSTRACT

Specialization polysemy refers to the type of polysemy, when a term is used to refer to either a more general meaning or to a more specific meaning. Although specialization polysemy represents a large set of the polysemous terms in WordNet, no comprehensive solution has been introduced yet. In this paper we present a novel approach that discovers all specialization polysemy patterns in WordNet and introduces new operations for solving all the instances of the problem.

KEYWORDS: *WordNet, Polysemy, Specialization Polysemy, regular Polysemy, Polysemy Reduction*

1 Introduction

Solving the polysemy problem in WordNet [1] is very crucial in many research fields including Machine translation, information retrieval and semantic search [13]. Several approaches have been introduced to solve the polysemy problem, but no approach gives a comprehensive solution to the problem. Solving the polysemy problem is very important because the high polysemous nature of WordNet leads to insufficient quality of natural language processing (NLP) and semantic applications.

Current polysemy approaches classify the polysemy problem in *contrastive polysemy* which corresponds to the polysemous terms that have unrelated meanings and *complementary polysemy* which corresponds to the polysemous terms with related meanings. This classification is correct but in general, it is not sufficient to solve the problem. We need further analysis of the different types of related polysemy and introduce a solution that solves the problem according to the specific nature of each of these related polysemy types [12]. For example, the methods for solving the *metonymy polysemy* (described in section 2) cannot be applied for solving the *specialization polysemy*, although both polysemy types belong to the complementary polysemy.

Specialization polysemy is a type of complementary polysemy that refers to the cases, when a term is used to refer to either a more general meaning or a more specific meaning [5]. The more general/ more specific meaning relation between the senses of specialization polysemy terms reflects a hierarchical relation between the senses that is encoded implicitly at lexical level rather than the semantic level. For instance, in the following example, sense 2 is a more general meaning than sense 1:

1. **`correctness`**`, rightness: conformity to fact or truth.`
2. **`correctness`**`: the quality of conformity to social expectations.`

Although Specialization polysemy represents a large set of the polysemous terms in WordNet, no comprehensive solution has been introduced yet. Systematic polysemy approaches such as CORELEX [4] did not provide a solution for specialization polysemy. Regular polysemy approaches such as the work presented in [5] discovered some patterns of specialization polysemy cases without offering a solution. On the other hand, polysemy reduction approaches tried to solve a subset of the specialization polysemy cases through merging the similar meanings of polysemous terms [3].

In this paper, we present a novel approach to solve the specialization polysemy in WordNet. The presented solution solves the specialization polysemy problem by providing a semi automatic method for discovering the specialization polysemy cases by means of regular structural patterns. It also provides criteria for determining the nature of the hierarchical relation between the senses of a specialization polysemy cases and new operations that solve the specialization polysemy problem by transforming the implicit relations between the synsets at lexical level into explicit relations at the semantic level. The advantages of our approach are that it improves the ontological structure of specialization polysemy cases and increases the knowledge in WordNet by adding new missing senses and relations rather than merely decreasing knowledge as it is suggested in polysemy reduction [3] and sense clustering approaches [10, 11].

This paper is organized as follows: in Section two, we give an overview of polysemy types in WordNet and make a comparison between specialization polysemy and other polysemy types. In Section three, we present the structural patterns of specialization polysemy and an algorithm for discovering these patterns. In Section four, we introduce the synset patterns in the case of specialization polysemy and show how we use these patterns to solve the specialization polysemy problem. In Section five, we discuss the results and evaluation of our approach. In Section six, we conclude the paper and describe our future research work.

2 Specialization Polysemy

WordNet [1, 2] is a lexical database that organizes synonyms of English words into sets called synsets where each synset is described through a gloss. WordNet organizes the relations between synsets through semantic relations where each grammatical category has a number of relations that are used to organize the relations between the synsets of that grammatical category. For example, the hyponymy relation (X is a type of Y) is used to organize the ontological structure of nouns. WordNet 2.1 contains 147,257 words, 117,597 synsets and 207,019 word-sense pairs. Among these words there are 27,006 polysemous words, where 15776 of them are nouns. The number of senses of polysemous nouns may range from 2 senses to 33 senses. Nevertheless, 90% of these nouns have less than 5 senses. WordNet uses sense ranking to order the synsets of the polysemous words. This order reflects the familiarity of the senses. The sense number 1 is the most familiar or *the common sense* of the synset. Another important ranking is the synset synonyms ranking. This ranking reflects which term is usually used to express a synset, where the first synonym is the most used term and so on. The first synonym of a synset is also called the *preferred term* of the synset. Note that, in this paper, we use the notion term to refer for a word and its part of speech. For example, the word *love* has two terms: love as a noun, and love as a verb. We use the notion sense(s) to refer to synset(s) of a term. Notice that, in this paper, we are concerned with polysemous nouns only.

Polysemy approaches differentiate between *contrastive polysemy*, i.e. terms with completely different and unrelated meanings—also called homonyms; and *complementary polysemy*, i.e. terms with different but related meanings. Complementary polysemy is classified in three sub types: *Metonymy*, *specialization polysemy*, and *metaphoric polysemy*. Polysemy approaches did not offer a solution for the polysemy problem that takes into account the different nature of each of these types. For example, regular polysemy approaches dealt with metonymy and metaphoric cases only. Classifying polysemy types and providing a solution for each type is a very important improvement towards making WordNet a suitable resource for NLP applications. In the following we explain the different polysemy types and discuss the difference between specialization polysemy and metonymy and metaphors.

2.1 *Specialization Polysemy*

Specialization polysemy is a type of complementary polysemy which denotes a hierarchical relation between the senses of a polysemous term. In case of abstract senses, we say that a sense A is a more general meaning of a sense B. In this case we say also that the sense B is a more specific meaning of the sense A. In the cases, where the senses denote physical entities, we may also use the taxonomic notations type and subtype instead of more general meaning and more specific meaning respectively. In the following examples, sense 2 denotes a subtype of the type denoted by sense 1 for the term turtledove:

1. Australian turtledove, **turtledove**, Stictopelia cuneata: small Australian dove
2. **turtledove**: any of several Old World wild doves.

A very important characteristic of specialization polysemy terms that differentiate it from contrastive polysemy and metonymy terms is the *type compatibility* of the term senses. By type compatibility, we mean that the term senses belong to the same type. For example both types of turtledove belong to the type dove. Some metaphoric cases as we shall see later, have the type compatibility property also.

2.1.1 Metonymy Polysemy

Metonymy polysemy happens when we substitute the name of an attribute or a feature for the name of the thing itself such as the second sense in the following example.

1. `fox: alert carnivorous mammal with pointed muzzle and ears and a bushy tail.`
2. `fox: the grey or reddish-brown fur of a fox.`

In metonymy, there is always a *base meaning* of the term and other *derived meanings* that express different aspects of the base meaning [8]. Sense 1 of the term fox in the previous example is the base meaning and sense 2 is a derived meaning of the term. Metonymy is different from specialization polysemy in the following way: The senses of metonymy terms belong to different types and thus the relation more general meaning/ more specific meaning is not applicable for metonymy. For example, the base meaning of the term fox belongs to the **animal** category while derived meaning belongs to **artifact**. This means, the relation between the derived meanings and the base meaning of a metonymy term cannot be hierarchical as it is the case in specialization polysemy. It is possible to find type compatible metonymy cases. The point here is that in such cases it is very difficult to distinguish between metonymy and specialization polysemy. We think that treating such cases as specialization polysemy is better since the hierarchical relation is stronger than the metonymic relation.

2.1.2 Metaphoric Polysemy

Metaphoric polysemous terms are the terms that have literal and figurative meanings. In the following example, the first sense of the term **honey** is the literal meaning and the second sense is the figurative:

1. **honey**: a sweet yellow liquid produced by bees.
2. beloved, dear, dearest, loved one, **honey**, love: a beloved person.

The metaphoric relation between the literal sense and the metaphoric sense may disappear or it may become difficult to understand the metaphoric link between the metaphoric and literal sense of the term. We call such cases dead metaphors. For example, the senses of animator indicate a dead metaphor:

1. `energizer, vitalizer, `**`animator`**`: someone who imparts energy and vitality to others.`
2. **`animator`**`: the technician who produces animated cartoons.`

From a hierarchical point of view metaphors can be divided into two groups:

a. *Type compatible metaphors*: the cases, where the literal meaning and the figurative meaning belong to the same type. Consider the term role player for example:

   1. `pretender, `**`role player`**`: a person who makes deceitful pretenses.`
   2. `actor, `**`role player`**`: a theatrical performer.`

b. *Type incompatible metaphors*: the cases, where the literal meaning and the figurative meaning belong to the different types. The literal meaning of the term **honey** for example belongs to the **food** category, while the figurative meaning belongs to **person**.

The metaphoric relation is not hierarchical. The metaphoric link between the senses is raised usually through inconsistency between the literal and the metaphoric sense. Although both senses of the term **role player** belong to the type person, these senses are inconsistent and cannot be generalized to a common type. In the case of dead metaphors and/or the cases, where it is difficult to grasp the metaphoric link between the senses, compatible metaphors can be treated as specialization polysemy while incompatible metaphors can be categorized as homonyms.

2.1.3 Contrastive Polysemy

The senses of a contrastive polysemous term have different etymological origins and they are not related. These senses are also said to be *homographs*. For example, the origin of sense 1 of the term bank is Italian, while the second sense is Norwegian.

1. `depository financial institution, `**`bank`**`,: a financial institution.`

2. **bank**: sloping land (especially the slope beside a body of water).

Although, there is no relation between the senses of contrastive terms, it is possible to find cases with related senses. For example, the two senses of the term animator can be considered homonyms. The link between the senses can be ignored in some cases as in the following example:

1. **Pascal**, Pa: a unit of pressure equal to one newton per square meter.
2. **Pasca**l: a programing language designed to teach programming.

Although both senses share the same term that refers to the famous French mathematician **Pascal**, they are in fact homonyms since they belong to two different categories: unit of measurement and programming language, respectively.

3 Structural Patterns

In defining regular structural patterns, our approach relies on Apresjan's definition of regular polysemy: *"A polysemous Term T is considered to be regular if there exists at least another polysemous T' that is semantically distinguished in the same way as T"* [8].

In the following, we describe type compatible structural patterns, and how we use these patterns to discover specialization polysemy terms.

3.1 Types of Structures

Structural patterns in WordNet are found at three levels of the ontological structure of WordNet. In general, the patterns at the upper level ontology correspond to metonymy and incompatible metaphoric cases. The patterns at the middle level and lower level correspond to specialization polysemy and compatible metaphoric cases. Homonyms do not follow any pattern and can be found at any level of the ontological structure of WordNet. Accordingly, we consider homonyms found in specialization polysemy patterns as false positives. In the following, we define a subset of the of the structural patterns in wordNet, namely the type compatible patterns, where we consider the type within its subtypes as a pattern to capture type compatible patterns.

Definition 1: Type Compatible Pattern Let T be a polysemous term that has n meanings, $n > 1$. Let S be the set of the synsets of T. Let R be a subset of S. Let Q an ordered sequence of R, where $|R| = m$, $2 \leq m \leq n$, and $Q =< s_1,..,s_m >, s_i \in R, s_i \neq s_j$, for $i \neq j$. A pattern *ptrn* of T is defined as $p\#< p_1,.., p_m >$, such that each p_i is a direct hyponym of p and subsumes s_i, $1 \leq i \leq m$. We call p the type (the category) of the pattern and p_i the subtypes of the pattern. For example, **vascular plant** is the type of the pattern *vascular plant#<herbaceous plan, bulbous plant>* that has the subtypes **herbaceous plant** and **bulbous plant**.

The previous definition is suitable for capturing type compatible patterns in the upper and middle level ontology of WordNet. However, this definition is not suitable to capture patterns at the lower level ontology since polysemous terms at the lower level ontology correspond usually to the cases, where the senses of each polysemous term share a common parent. To be able to capture the structural regularity at the lower level ontology, we define the common parent class:

Definition 2: Common parent class Let T be a term that has n meanings, $n > 1$. Let S be the set of the senses of T. T belongs to the common parent class if the following occurs:

$$\exists R(R \subseteq S \wedge |R| > 1 \wedge \forall s(s \in R \Rightarrow \exists p\,(hypernym(s, p) \wedge \neg\exists s'(s' \in R \wedge \neg hypernym(s', p))))$$

In Figure 1, the sense of croaker is a hypernym of the two senses of white croaker and is therefore an example of common parent class.

Not all polysemous terms at the lower level ontology share the same parent. There are cases, where the direct parent of one synset is an indirect parent of the other. In some cases, the distance between the indirect synset and the common root is two. We consider these terms as members of the common parent class.

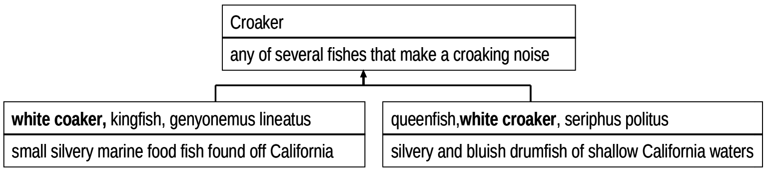

Fig. 1: An example for common parent class

Definition 3: Regular Type Compatible Pattern Let T be a polysemous term that has n meanings. Let S be the set of the synsets of T, Let *ptrn* be a pattern of *T*. *T* is considered to belong to a regular patterns if the following occurs: There exists at least another Term T' such that T and T' are not synonyms and T' belongs to *ptrn* or $\exists Q(Q \subseteq S \wedge Q \in common\ parent)$.

The pattern *vascular plant#<herbaceous plan, bulbous plant>* is regular since there are 6 terms that belong to it. In addition to regular patterns we are also interested in sub patterns. Our hypothesis is that the sub patterns of a specialization polysemy pattern belong also to specialization polysemy.

Definition 4: Sub pattern For a regular pattern $ptrn = p\#< p_1,..,p_m >$, A $ptrn'$ is considered to be a sub pattern of *ptrn* if $ptrn' = p\#< p_1',..,p'_k >$ and $\exists p_i, p_j'(p_i = p_j'), 1 \le i \le m, 1 \le j \le k$.

For example, the regular pattern *vascular plant#<herbaceous plant, bulbous plant>* has the following sub pattern: *vascular plant#<bulbous plant, hydrophytic plant>*.

3.2 Discovering Specialization Polysemous Terms via Structural Patterns

The basic idea of our solution is to find all terms in WordNet, where the senses of these terms fulfill the type compatibility criterion since this criterion is the main characteristic of all specialization polysemy terms. At the lower level ontology, the terms that belong to the common parent class automatically fulfill this criterion. The patterns at the top level ontology including CORELEX patterns do not fulfill this criterion. In the middle level ontology, we have patterns that correspond to specialization polysemy and other patterns that correspond to compatible metaphoric terms. Both polysemy types fulfill the type compatibility criterion. Thus our task is to classify those patterns into specialization polysemy patterns and metaphoric patterns. Notice here that there are patterns that include both polysemy types. These patterns require further step to identify the specialization polysemy terms. Our approach works in four phases as follows:

A. Patterns Identification
B. Patterns Classification
C. Polysemy Type Assignment
D. Validation

The first and the third phases are automatic, while the second and fourth are manual. In the following we discuss the four phases of our approach that we applied on the nouns that have exactly two senses.

A. Patterns Identification

In this phase, we used the following algorithm to identify the regular type compatible patterns.

Algorithm: Regular Type Compatible Patterns Extraction
Input:
PNOUNS = Polysemous nouns in WordNet
UNIQUEBEGINNERS = list of the unique beginners in WordNet
SENSENUMBER = the number of the term synsets,
Output:
N = an associative array to store the regular patterns.
M = an associative array to store the sub patterns
P = a list to store the elements of the common parent class
O = a list of singleton patterns
1. *poly_nouns* = retrieve_polysemous_nouns(*SENSENUMBER*)
2. For each *noun* in *poly_nouns*
3. *S* = retrieve_synsets(*noun*)
4. *ptrns* = get_patterns(S)
5. For each $Q \subseteq S$
6. If $Q \in$ *Common Parent*
7. add <*noun*, *Q*> to P.
8. For each pattern *ptrn* = $p\#<p_1,..,p_m>$ in *ptrns*
9. If $p \notin$ *UNIQUEBEGINNERS*

```
10.             Add noun to the list under ptrn in N.
11. For each  ptrn in N
12.      If |N[ptrn]| > 1
13.            M[ptrn] =  sub_patterns(ptrn)
14.            Remove sub_patterns(ptrn) from N
15. For each ptrn in N
16.      If |N[ptrn]| < 2
17.            Add ptrn to O
18.            Remove ptrn from N
19. return <N,M,P,O>
```

The presented algorithm works in three phases:

1. *Patterns and common parent terms identification* (lines 1 to 10): We retrieve the list of all nouns that have the sense number given in the algorithm input. We check, whether the term belongs to the common parent class and also whether it has regular patterns. We exclude the top level ontology patterns such as *physical entity<physical object, physical process>*. Such patterns correspond usually to CORELEX patterns and they are not specialization polysemy patterns. Notice also that it is possible for terms that have more than 2 senses to have more than one pattern.
2. *Sub patterns identification* (lines 11 to 14): If more than one term belong to a pattern, thus it is a regular pattern, then we search all singleton patterns to identify possible sub patterns of that pattern. Identified sub patterns are removed from the patterns list and added to the sub patterns list.
3. *Singleton patterns identification* (lines 15 to 18): After identifying the sub patterns, the remaining singleton patterns are removed from the patterns list and added to the list of the singleton patterns.

The results of applying the algorithm on the terms that have two synsets are as follows: the total number of the nouns in WordNet that have two senses is 9328 nouns. 2899 nouns of them were identified by the algorithm to belong to type compatible patterns. The algorithm returned four lists: a pattern list that contains 333 patterns, a sub patterns list that contains 344 sub patterns, the list of the common parents that contains1002 terms, and a list that contains 358 singleton patterns.

B. Patterns Classification

Our task in this phase is to classify the patterns in specialization polysemy and metaphoric polysemy. First of all, the terms that belong to the common parent are considered as specialization polysemy candidates. We consider also the polysemy type of the sub patterns as the polysemy type of the pattern, they belong to. To classify the patterns, we have arranged them into hierarchies. The roots of the hierarchies are shown in the following table. The numbers right to the types correspond to the number of patterns that belong to that type.

Table 1. The roots of type compatible patterns in WordNet

Patterns under physical entity		Patterns under abstract entity	
Type	#patterns	Type	#patterns
substance	6	psychological feature	2
organism	4		
person	106	cognition	12
animal	20	attribute	26
plant	18	communication	18
artifact	73	measure	14
process	9	group	9
location	4	time period	4
thing	5	relation	3

Analyzing the patterns under these types shows that these patterns can be classified into four groups:

1. Specialization polysemy patterns
2. Metaphoric patterns
3. Homonymy patterns
4. Mixed patterns

In the following, we explain our criteria by classifying the patterns.

1. *Specialization Polysemy patterns*: the type of some specialization polysemy patterns can be determined directly by considering the type of the pattern only. For example, it is clear that the patterns whose type belongs to *animal*, and the types under *animal* are specialization polysemy or at least it is not common at all to find a metaphoric link between the types under *animal*. The criteria for determining other

specialization polysemy patterns is the *consistency* of the pattern subtypes.

2. *Metaphoric patterns*: to determine metaphoric patterns, we followed the idea that metaphors are human centric in the sense that we use metaphors to express our feelings, judgments, situations, irony and so on. For example, when we use sponger to refer to some one, we are making a judgment upon that person. This gives us a hint, where to search for metaphoric patterns, namely under the person type or the types whose subtypes indicate meaning transfer from their literal meaning to a (metaphoric) human centric meaning as discussed below. Here, the type attribute is an example of such cases.

a. *Metaphoric patterns under person*: we found under the type person 106 patterns. Some of these patterns are specialization polysemy patterns and others are metaphoric. To determine metaphoric patterns under the type person, we searched for inconsistency between the subtypes of the patterns. We find such inconsistency for example in the pattern person#<bad person, worker>, the sub type bad person is not consistent with the type worker and therefore a specialization polysemy is totally excluded in this pattern. The term iceman is an example of terms that belong to this pattern:

1. iceman: someone who cuts and delivers ice.
2. hatchet man, iceman: a professional killer.

On the other hand the subtypes of the pattern person#<expert, worker> are consistent and is considered as a specialization polysemy pattern. The term technician is an example for this pattern:

1. **technician**: someone whose occupation involves training in a technical process.
2. **technician**: someone known for high skill in some intellectual or artistic technique.

b. Metaphoric patterns under attribute: Our criteria here was to find meaning transfer between the sub types. Attribute has the following four patterns: attribute#<property, trait>, attribute#<property, state>, attribute#<property, quality>, and attribute#<quality, trait>, with the following meanings:

```
Property: a basic or essential attribute shared by all
members of a class.
State: a state of depression or agitation.
Quality: an essential and distinguishing attribute of
something or someone.
Trait: a distinguishing feature of your personal nature.
```

The meaning transfer from property to human centric meaning is clear in the first three patterns. For example, in the term *chilliness*:

```
1. chilliness, coolness, nip:  the property of being moder-
   ately cold.
2. coldness, frigidness, iciness, chilliness: a lack of af-
   fection or enthusiasm.
```

In the fourth pattern, the relation between quality and trait depends on whether the term under the quality subtype refers to *an attribute of something* or *an attribute of someone*. The first case corresponds to metaphoric polysemy while the second corresponds to specialization polysemy.

3. *Homonymy Patterns:* In general, homonymy can not be considered as a type of regular polysemy. Nevertheless, we cannot exclude the existence of homonymy patterns. WordNet contains few homonymy patterns such as the following pattern: *organism#<animal, plant>*, where we find type mismatch between the subtypes. Specialization or metaphoric polysemy in such patterns is totally excluded.

4. *Mixed patterns*: This group contains the patterns that were identified to have more than one polysemy type. For example, the pattern *attribute#<quality, trait>* belongs to this group.

In summary: there are some patterns whose sub types indicate type inconsistency. After excluding these patterns, all patterns under the *physical entity* are candidates for specialization polysemy except the patterns under *person* which contains both polysemy types. In the case of *abstract entity*, most of the patterns under attribute are candidates for metaphoric polysemy. The patterns under *cognition* and *communication* contain both polysemy types, and the rest types are candidates for specialization polysemy.

C. Polysemy Type Assignment

In this phase, each of the nouns, that were determined to belong to type compatible polysemy patterns, is assigned to specialization polysemy or metaphoric according to the pattern of the term. The terms that belong to both polysemy types and the terms that belong to the singleton patterns are not assigned and they are subject to manual treatment in the validation phase.

D. Validation

In this phase, we manually validate the assigned polysemy type. Our criterion is to determine the relation between the senses of a term and thus the polysemy type, is the synset gloss. In difficult cases, we also consider the hierarchical properties of the term synsets. We have three tasks in this phase:

1. *Validation of the assigned polysemy types*: we check whether each of the nouns belong to its assigned polysemy type.
2. *Assigning the polysemy type*: for the terms that belong to the mixed patterns and singleton patterns.
3. *Excluding of false positives*: we exclude the false positives from the terms of the 4 groups.

Our judgments during the validation process are based on knowledge organization in such a way that word etymology and linguistic relatedness have secondary role in our judgments. The primary criterion is:

1. In case of specialization polysemy: Is it possible for both senses to be generalized to a common type? If the answer is no or we don’t know, then we consider the term to be a homonymy case. The term cardholder is an example for such cases:

    1. **cardholder**: a person who holds a credit card or debit card.
    2. **cardholder**: a player who holds a card or cards in a card game.

2. In case of metaphoric polysemy: Is it easy to discover the metaphoric link between the senses? If the answer is no or we don’t know, then we consider the term to be specialization polysemy candidate. The term agreeableness that belongs to the metaphoric pattern attribute#<quality, trait> is an example for such cases:

1. **agreeableness**, amenity: pleasantness resulting from agreeable conditions.
2. **agreeableness**, agreeability: a temperamental disposition to be agreeable.

4 Synset patterns

The structural patterns served as a criterion for identifying specialization polysemy candidates. The next step is how to solve the polysemy problem for the identified candidates. The *more general meaning/more specific meaning* relation between the senses of the specialization polysemy terms reflects a hierarchical relation between the senses. Thus, the solution should reflect this relatedness. In the following, we explain how the synonyms of the specialization polysemy synsets are used to organize the hierarchical relation between the senses.

4.1 Types of Synsets

In our approach, we have analyzed the relation between the synset synonyms and the possible relation between the synsets of specialization polysemy cases. The idea here is that the nature of the relation between the synsets of specialization polysemy terms can be determined based on the synonyms of such terms. Based on the synset synonyms , we divided the specialization polysemy terms in three groups:

1. Twin synsets
2. Type – sub type synsets
3. General meaning – example meaning synsets

1. *Twin synsets*: both synsets of such terms contain other synonyms beside the polysemous terms. Analyzing these cases shows that the *is a* relation does not hold between the synsets themselves. In fact both synsets are more specific in meanings of some (non existing) third synset as in the following example:

1. **white croaker**, queenfish, Seriphus politus: silvery and bluish fish of California.
2. **white croaker**, kingfish, Genyonemus lineatus: silvery fish of California.

2. *Type - sub type synsets*: One synset contains the polysemous terms only, the other contains the polysemous terms and other synonyms. Analyzing these cases shows that the gloss of the synset that contains the polysemous terms only usually begins with the following phrase: "*any of several* " which reflects that this synset encodes a more general meaning while the synset with additional synonyms describes a specific type that belong to the type of both synsets. For example, sense 1 describes a specific type, while sense 2 is a general description of turtledove.

1. `Australian turtledove,` **`turtledove`**`, Stictopelia cuneata: small Australian dove.`
2. **`turtledove`**`: any of several Old World wild doves.`

3. *General meaning - example meaning synsets*: both synsets contain the polysemous terms only. Analyzing these cases shows that there is a synset which denotes the meaning of the term in general while the other synset denotes an example of that general meaning. According to our analysis, the synset with the general meaning has usually sense rank 1. For example sense 1 denotes the general meaning of the term timetable while sense 2 is an example of the term. Notice that, there are many other examples of timetables such as *schedule of lessons in the school*. We think that sense 2 is an example for unnecessary sense enumeration in WordNet and we consider the senses as candidates to be merged.

1. **`timetable`**`: a schedule listing events and the times at which they will take place.`
2. **`timetable`**`: a schedule of times of arrivals and departures.`

4.2 Organizing Specialization Polysemous Terms via Synset Patterns

According to the above analysis, we suggest to solve the specialization polysemy by reorganizing the ontological structure of the synsets, where the implicit hierarchical relation between the synsets at lexical level is transformed into explicit hierarchical relation at semantic level. This requires adding missing synsets, is a relations and removing redundant is a relations.

1. *Solution for Twins synsets*: We add a new (missing) parent in cases, where the polysemous meanings of a term T can be seen more specific

meanings of an absent more general meaning: Let s_1, s_2 be two synsets of a term belonging to the missing parent cases. Let $\{T_1,..,T_n\}$ be the set intersection of s_1 and s_2. Let T' be the preferred term in s_1 and s_2 or the term with the highest rank in both synsets. Let T be the preferred term of the type of s_1 and s_2. We create a common parent S_p of s_1 and s_2 as follows:

i) Create a new synset S_p such that:

   The lemmas are the intersection of the lemmas of s_1 and s_2;
   The gloss of S_p = T' is a T.

ii) Remove the common lemmas from s_1 and s_2
iii) Connect S_p to T via the *is-a* relation
iv) Connect the senses s1 and s2 to S via the *is-a* relation
v) Remove redundant relations

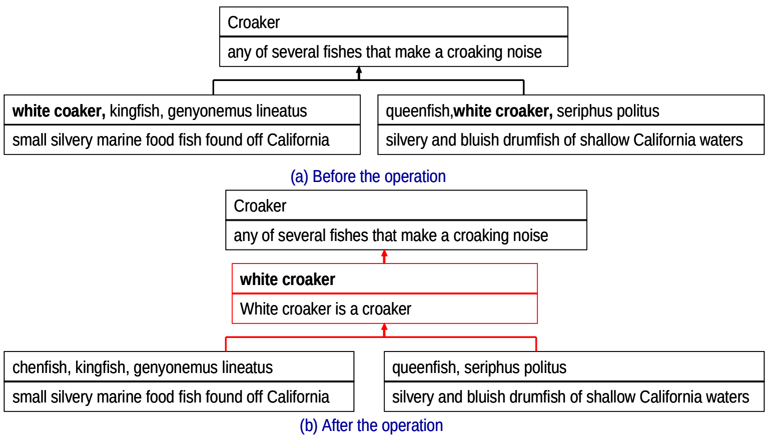

Fig. 2. Example for adding a new missing parent

2. *Solution for type – sub type synsets*: In such cases we establish a missing *is_a* relation to denote that a sense of a polysemous term T is more specific than another more general meaning of T: Let s_1, s_2 be two synsets of a term belonging to the missing relation cases. Let s_2 be the synset that has the polysemous terms and additional terms. Let s_1 be the synset that contains the polysemous terms only.

i) Connect s1 to s2 via the *is-a* relation: s_2 *is-a* s_1.
ii) Remove redundant relations.

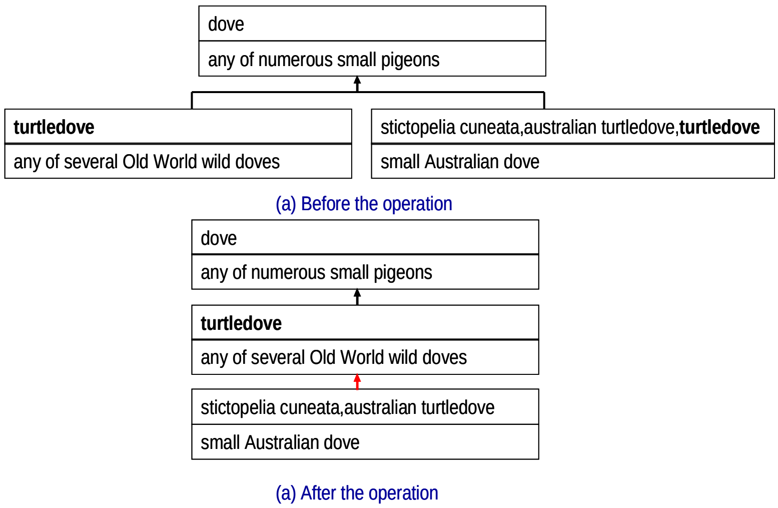

Fig. 3. Example for adding missing relation

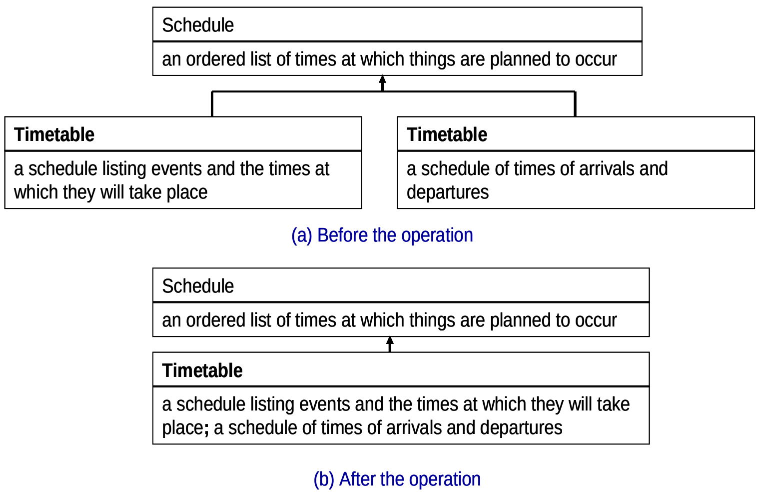

Fig. 4. Example for merge operation

3. *Solution for general - example synsets*: In such cases, we merge the senses of the terms as follows**.**

Let s_1, s_2 be two synsets of a term belonging to the merge cases. We keep the synset with sense rank 1 and remove the other one as follows:

i) The lemmas of s_1 are the same as before since both synsets share the same lemmas.
ii) The gloss of s_1 = the gloss of s_1; the gloss of s_2
iii) The relations of s_1 are the union of the relations of both synsets
iv) Remove redundant relations

5 Results and Evaluation

In the following, we describe the results and the evaluation of our approach. Table 2 presents the results of the four pattern groups and common parent group after the validation.

Table 2. Validated results of the approach

Patterns group	#total cases	#Specialization Polysemy	#Metaphors	#Homonyms
Spec. Polysemy Patterns	807	673	26	108
Metaphoric Patterns	221	28	170	23
Homonyms Patterns	56	0	0	56
Mixed Patterns	111	41	39	31
Common Parent	1002	927	40	35
Sub patterns and singleton patterns	702	455	90	157
Total nouns	2899	2124	365	410

In Table 3, we present the pattern groups that have been identified.

Table 3. Distribution of type compatible patterns

#Patterns	#Spec. Polysemy Patterns	#Metaphoric Patterns	#Homonym Patterns	#Mixed Patterns
333	225	79	15	14

As we can see in Table 2, 73% of the identified terms belong to specialization polysemy. In table 3, we find that 67.5% of the identified patterns are specialization polysemy patterns. In Table 2, we can also see that not all terms that belong to the common parent group are specialization polysemy terms. About 4% of these terms are in fact homographs. Consider for instance, term *apprehender* that belongs to the common parent group.

1. knower, **apprehender**: a person who knows or apprehends.
2. **apprehender**: a person who seizes or arrests.

Although both senses belong to the type person, they are in fact homographs. Also, about 3.5% of the common parent group were identified as metaphors. Consider for example, the following senses of the term *moment of truth*.

1. **moment of truth**: the moment in a bullfight when the matador kills the bull.
2. **moment of truth**: a crucial moment on which much depends.

We have examined CORELEX patterns to find overlap between CORELEX patterns and the patterns identified in our approach. We did not find any overlap between them. This was expected, since CORELEX patterns belong to the top level ontology, where as the specialization polysemy patterns were found at the middle and lower level ontology. An important thing to note here is that none of the terms that belong to CORELEX were identified as specialization polysemy terms. They belong mainly to metonymy.

In Table 4, we list the distribution of specialization polysemy operations.

Table 4. Specialization polysemy operations

Operation	Adding missing parent	Adding missing relation	Merge	Total
#cases	1045	685	409	2124

The total number of reduced polysemous words is 2124 words. The total number of merged synsets represents about 14% of the total processed cases. We have added 1045 new synsets and 1730 new relations, while deleted 409 synsets and 409 relations. Compared to polysemy reduction approaches, 86% of the cases were not merged. Instead of merging, we have reorganized the ontological structure of the terms. It is important here to notice that our approach improves the ontological structure of WordNet by increasing knowledge rather than decreasing knowledge as it is suggested by other approaches.

To evaluate our approach, 834 cases have been evaluated by two evaluators. In Table 5, we report the evaluation statistics, where the column polysemy type refers to homonymy, metaphoric, or specialization polysemy and polysemy operation refers to creating missing parent, adding missing relation, or merging operation. Note that, polysemy operation is applicable in case of specialization polysemy. The table presents the agreement between the evaluators and our approach. The third row represents the number of cases, where at least one evaluator agrees with our approach.

Table 5. Evaluation results

	Polysemy type agreement	Polysemy operation agreement
Evaluator 1	803 ≈ 96.2%	750 ≈ 89.9%
Evaluator 2	775 ≈ 92.9%	686 ≈ 82.2%
Partial agreement	824 ≈ 98.8%	796 ≈ 95.4%

6 Conclusion

In this paper, we have introduced an approach for solving the specialization polysemy problem based on type compatible regular patterns. This approach decreases polysemy, but at the same time knowledge is increased. It improves the ontological structure of WordNet, where the implicit relations between the synsets of polysemous terms which is encoded at lexical level are transformed into explicit semantic relations.

In the current paper, we presented the result of our approach applied on nouns that have two senses. Our future plan is to apply the approach on all other nouns in WordNet as a first step towards solving the other polysemy types.

References

1. Miller, G.A.: WordNet: a lexical database for English. In: Communications of the ACM, 38 (11), pp. 39–41 (1995)
2. Miller, G. A., Beckwith, R., Fellbaum, Ch., Gross, D. and Miller, K.: Introduction to WordNet: An on-line lexical database. In: International Journal of Lexicography (1990)

3. Mihalcea, R., Moldovan, D.I.: EZ.WordNet: Principles for Automatic Generation of a Coarse Grained WordNet. In: FLAIRS, pp. 454–458 (2001)
4. Buitelaar, P. P.: CORELEX: Systematic Polysemy and Underspecification. In: PhD thesis, Brandeis University, Department of Computer Science (1998)
5. Bouquet, P., Giunchiglia, F.: Reasoning About Theory Adequacy. A New Solution to the Qualification Problem. In: Don Perlis (editor): Fundamenta Informaticae, vol. 2/3/4, pp. 247-262 (1995)
6. Barque, L., Chaumartin, F.R.: Regular Polysemy in WordNet. In: JLCL 24(2), pp. 5–18 (2009)
7. Giunchiglia, F., Yatskevich, M., Avesani, P., Shvaiko, P.: A Large Scale Dataset for the Evaluation of Ontology Matching Systems. In: The Knowledge Engineering Review Journal (KER), Cambridge University Press, 24 (2) (2009)
8. Apresjan, J.: Regular Polysemy. In: Linguistics, 142, pp. 5–32 (1974).
9. Peters, W., Peters, I.: Lexicalized systematic polysemy in WordNet, Language Resources and Evaluation (2000)
10. Snow, R., Prakash, S., Jurafsky, D., Ng, A.Y.: Learning to Merge Word Senses. EMNLP-CoNLL 2007: 1005–1014 (2007)
11. Navigli, R.: Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance. ACL (2006)
12. Freihat, A.A., Giunchiglia, F., Dutta, B.: Approaching Regular Polysemy in WordNet. In: proceedings of 5th International Conference on Information, Process, and Knowledge Management (eKNOW), Nice, France (2013)
13. Giunchiglia, F., Kharkevich, I., Zaihrayeu, I.: Concept search. In Processing of ESWC'09, Lecture Notes in Computer Science. Springer (2009)
14. Peters, W., Peters, I.: Lexicalised Systematic Polysemy in WordNet, Department of Computer Science,University of Sheffield, U.K.
15. Giunchiglia, F., Dutta, B., Maltese, V.: Faceted lightweight ontologies. In: Conceptual Modeling: Foundations and Applications, Alex Borgida, Vinay Chaudhri, Paolo Giorgini, Eric Yu (Eds.) LNCS 5600 Springer (2009)
16. Pustejovsky, J.: The Generative Lexicon, MIT Press, Cambridge, MA (1995)
17. Navigli, R.: Word sense disambiguation: a survey, ACM Computing Surveys 41(2):1–69 (2009)
18. Nerlich, B., Clarke, D.D.: Polysemy and flexibility: introduction and overview. Berlin, New York: Mouton de Gruyter, 3–29 (2003)
19. Zaihrayeu, I., Su, L., Giunchiglia, F., Pan, W., Ju, Q., Chi, M., Huang, X.: From Web Directories to Ontologies: Natural Language Processing Challenger, 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference, ISWC + ASWC 2007, Busan, Korea. (2007)

Abed Alhakim Freihat
Dept. of Information Engineering and Computer Science,
University of Trento,
Trento, Italy
E-mail: <fraihat@disi.unitn.it>

Fausto Giunchiglia
Dept. of Information Engineering and Computer Science,
University of Trento,
Trento, Italy
E-mail: <fausto@disi.unitn.it>

Biswanath Dutta
Documentation Research and Training Centre (DRTC),
Indian Statistical Institute (ISI),
Bangalore, India
E-mail: <bisu@drtc.isibang.ac.in>