A Bottom-up Comparative Study of EuroWordNet and WordNet 3.0
Lexical and Semantic Relations

Maria Teresa Pazienza\textsuperscript{a}, Armando Stellato\textsuperscript{a}, Alexandra Tudorache\textsuperscript{ab}

\textsuperscript{a} AI Research Group, Dept. of Computer Science, Systems and Production University of Rome, Tor Vergata Via del Politecnico 1, 00133 Rome, Italy \{pazienza,stellato,tudorache\}@info.uniroma2.it

\textsuperscript{b} Dept. of Cybernetics, Statistics and Economic Informatics, Academy of Economic Studies Bucharest Calea Doroban\c tilor 15-17, 010552, Bucharest, Romania alexandra.tudorache@gmail.com

Abstract

The paper presents a comparative study of semantic and lexical relations defined and adopted in WordNet and EuroWordNet. This document describes the experimental observations achieved through the analysis of data from different WordNet versions and EuroWordNet distributions for different languages, during the development of JMWNL (Java Multilingual WordNet Library), an extensible multilingual library for accessing WordNet-like resources in different languages and formats. The aim was to realize an operative mapping between the relations defined in the two lexical resources and to unify library access and content navigation methods for both WordNet and EuroWordNet. The analysis focused on similarities, differences, semantic overlaps or inclusions, factual misinterpretations and inconsistencies between the intended and practical use of each single relation defined in these two linguistic resources. The paper details with examples the produced mapping, discussing required operations which implied merging, extending or simply keeping separate the examined relations.

1. Introduction

We introduce a comparative study of semantic and lexical relations defined and adopted in two renowned lexical databases: WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1993; Fellbaum, 1998) and EuroWordNet (Vossen, 1998). The study was conducted during the development of the Java Multi WordNet Library (JMWNL), an extensible multilingual library for accessing WordNet-like resources in different languages and formats, based on John Didion’s JWNL library (http://sourceforge.net/projects/jwordnet). The analysis and comparison between the two resources was carried out in the pre-design stage of development of the above library. The aim was to realize an operative mapping between the relations defined in the two lexical resources and to unify library access and content navigation methods for both WordNet and EuroWordNet. The work was conducted bottom-up by analyzing the raw data and several examples either from English WordNet and different language EuroWordNet (EWN from now on) resources. The analysis focused on similarities, differences, semantic overlaps or inclusions, factual misinterpretations and inconsistencies between the intended and practical use of each single relation defined in these two linguistic resources.

The mapping between the relations defined in the two resources required a two layered investigation. In most cases it sufficed to establish a template level mapping like telling that has\_hyperonym is equivalent – at least in the intentions of the lexicographers – to hypernym. This way, in theory, all instantiated relationships based on these two relations could be interchangeable. In other cases, (even partial) mappings between apparently different relations emerged by looking at a vast quantity of sample data and studying cross-linguistic similarities.

This document describes the experimental observations achieved through the analysis of data from different WordNet versions and EuroWordNet distributions for different languages, during the development of JMWNL. The paper focuses on relations mapping, cross part-of-speech relations and the partial mapping of Fuzzynym relations like Derivationally Related Form.

2. WordNet and EuroWordNet

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. WordNet “Synsets are interlinked by means of conceptual-semantic and lexical relations” (About WordNet, 2006). Its European counterpart, EuroWordNet, is a multilingual database with wordnets for several European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian). The wordnets are structured in the same way as the American wordnet for English in terms of synsets with basic semantic relations between them (Vossen, EuroWordNet Web Abstract, 2001). EuroWordNet is based on the 1998 WordNet version 1.5 (Fellbaum, 1998) and contains more (and different) relations than current English WordNet (version 3.0 - 2007), so a one- to-one relations mapping is not achievable.

Although the structure of various wordnets is similar and consistent, the relations defined in each version are not identical and moreover some wordnets are richer in relations than others.

3. Overall Mapping Statistics

Our analysis revealed that only some EuroWordNet relations could be mapped directly with WordNet relations. Other EuroWordNet relations had to be added.
English WordNet has in total 46 relations defined on all parts of speech while EuroWordNet has more than one hundred relations. During our integration for building the JMWNL library we could map directly 32 EuroWordNet relations and we had to add 17 new relations that were present in EuroWordNet from which 10 are defined across multiple parts of speech (XPOS). We created these 10 new relations using the defined WordNet relation pointer symbols adding an “x” character as prefix to express the cross relationship.

We found in total 21 WordNet relations that are not present in EuroWordNet and that couldn’t be mapped even partially.

4. Comparative results and observations

As previously mentioned, it is not possible to build a complete 1-to-1 mapping between WordNet and EuroWordNet lexical and semantic relations. However a correct and complete record of all lexical and semantic relations is indispensable for building multilingual applications that use available wordnets as their lexical databases. Moreover, it is necessary to establish relationships between their models to create the grounds for working consistently across different languages. This section will show the differences between WordNet and EuroWordNet identifications of lexical and semantic relations focusing on the most important EuroWordNet relations and their correspondent WordNet ones.

4.1. Synonymy and Antonymy

Unlike in WordNet, EuroWordNet distinguishes between tight and loose synonymy and antonymy relationships, introducing two relations respectively called **NEAR SYNONYM** and **NEAR ANTONYM**; for example, in Italian EuroWordNet the word "nemico" ("enemy", in English) is **NEAR ANTONYM** of "allievo" (Eng. "allay"), while in original WordNet “enemy” is only **ANTONYM** of “friend”.

These two relations can be mapped directly as **SIMILAR TO** and **ANTONYM** in WordNet. The tight version of synonymy is implicit in the WordNet definition of synset (words appearing in the same synset are, by definition, synonyms), while, in the case of antonymy, EWN tight and loose antonyms both collapse in the **ANTONYM** definition in WordNet (see section 4.4).

4.2. Meronymy and Holonymy

In EuroWordNet **HAS_MERO/HAS_HOLO** with all their variants express respectively **HOLONYM/MERONYM** relations.

More specifically, in EuroWordNet, **HAS MERO MADEOF** and **HAS HOLO MADEOF** relations have a partial overlap with both **SUBSTANCE MERONYM/HOLONYM** and **PART MERONYM/HOLONYM**.

In Figure 1 are shown two examples of **HAS MERO MADEOF** and **HAS HOLO MADEOF** relations. The first example is the overlap with **SUBSTANCE HOLONYM**: “abito” (suit) is made of “seta” (silk). The second example shows the overlap of **HAS MERO MADEOF** with **PART MERONYM**: “palo” (pole) is part meronym of “steccato” (wooden fence).

In Table 1 are presented both **HAS MERO PART** and **HAS HOLO MEMBER** relations in parallel starting from the same base concept: “tree”. Looking at this example we could conclude that there are not big differences between definitions of the relations for such base concepts.

| Base Word | Relation | Italian | French | English Translation |
|-----------|----------|---------|--------|--------------------|
| Albero (It) | Has mero part | frutto | N/A | fruit |
| Arbre (Fr) | ramo | corteccia | souche | bark |
| Tree (En) | cima | foglia | N/A | leaf |
| Has holo member | cima | tronco | d’arbre | branch |
| | radice | cime | tronc | trunk |
| | bosco | N/A | cime | flower |

Table 1: Multilingual Holonymy and Meronymy Relations
For less specific concepts instead meronymy and holonymy relations are loosely defined. E.g. "football americain" (american football) "has_mero_part" "match de football" (football game); "fasciatura" "has_mero_member" "fascia"; "bowling" "has_mero_part" "roll" (the act of rolling something (as the ball in bowling)). These differences in the definition meronymy and holonymy relations are mostly attributable to human interpretation and language particularities but also to the grade of the maturity of concepts. Concepts present in modern vocabulary tend to have more loosely defined grade of the maturity of concepts. Concepts present in EuroWordNet are collapsed in one relation. E.g. "Football américain" becomes either PERTAINYM (A\N), PARTICIPLE OF VERB (A\V). At the same time IS DERIVED FROM in WordNet becomes either NEAR ANTONYM (A\N), PARTICIPLE OF VERB (A\V). In Figure 3 is presented the mapping of EuroWordNet NEAR ANTONYM relation with WordNet ANTONYM Relation. E.g. "piccolo" (little, small) is near antonym of "grande" (big). In English WordNet the same relation is defined as antonym.

### 4.3. Fuzzynym

Two interesting EuroWordNet relations that we explored are FUZZYNYM (X has some strong relation to Y, same POS) and XPOS FUZZYNYM (X has some strong relation to Y, different POS). FUZZYNYM and XPOS FUZZYNYM are mostly semantic relations that are not belonging to other categories. As underlined by Morris & Hirst (2004), NLP applications need to explore such not perfectly structured and context dependent relations. These relations can’t be mapped directly to any WordNet relation but a part of their instantiations may be considered as members of the DERIVATIONALLY RELATED FORM found in WordNet. During the integration process we found that using an algorithmic multilingual stemmer (e.g. Snowball - http://snowball.tartarus.org/) it is possible to extract most of standard WordNet DERIVATIONALLY RELATED FORM relations from EuroWordNet FUZZYNYM relation. This process would need only a fast human validation to properly import the correct relation instances. The initial check of this process was done using a small portion of the English EuroWordNet resource, automatically comparing the results with original WordNet.

More tests were performed on other languages present in EuroWordNet (e.g. Italian and French) with a manual validation (since the original WordNet resource is only in English). With proper stemmer settings and word similarity measure (Cohen, Ravikumar, & Fienberg, 2003), we got high precision ratings ranging from 80% to 90%, thus requiring a light, but careful, filtering work by a human supervisor. In Figure 2 is shown an example of transformation of a FUZZYNYM relation instance into an original WordNet DERIVATIONALLY RELATED FORM.

### 4.4. Collapsed Relations

Some relations that belong either to WordNet or EuroWordNet are collapsed in one relation. E.g. WordNet ANTONYM relation groups EuroWordNet NEAR ANTONYM and ANTONYM while EuroWordNet IS DERIVED FROM in WordNet becomes either FUZZYNYM (X\N), PARTICIPLE OF VERB (X\V). At the same time IS DERIVED FROM (as relation is not mappable for nouns. E.g. "generalmente" (generally) is derived from "generale" (general)). In Figure 3 is presented the mapping of EuroWordNet NEAR ANTONYM relation with WordNet ANTONYM Relation. E.g. "piccolo" (little, small) is near antonym of "grande" (big). In English WordNet the same relation is defined as antonym.

### 4.5. Extended relations and cross part of speech relations

In EuroWordNet are defined some relations between different parts of speech that are not present in WordNet. We introduced them in order to preserve all EuroWordNet relations. These relations are marked as XPOS (cross POS), like HAS XPOS HYPERONYM, XPOS NEAR ANTONYM, XPOS NEAR SYNONYM, HAS XPOS HYPONYM and XPOS FUZZYNYM. In our mapping, to express cross relations we have chosen to maintain WordNet pointer symbols whenever possible, adding only one “x” as prefix.
4.6. WordNet relations non present in EuroWordNet
A number of English current WordNet relations are not defined in the first Fellbaum version. These relations are: INSTANCE HYPERNYM, ATTRIBUTE, ALSO SEE, VERB GROUP, TOPIC, DOMAIN and USAGE relations with all their versions.

4.7. Notes on equivalent relations
In WordNet are present some equivalent relations (EQ) linked to the ILI (Inter-Lingual-Index). Although the ILI does not cover all language internal relations, it can be used to aid in cross language mapping. Such equivalent relations are: EQ SYNONYM, EQ NEAR SYNONYM, HAS EQ HYPERONYM, HAS EQ HYPONYM, EQ HAS HOLONYM, EQ IN MANNER, EQ BE IN STATE, EQ HAS MERONYM, EQ CAUSES, EQ IS STATE OF, EQ INVOLVED, EQ IS CAUSED BY, EQ RULE, EQ HAS SUBEVENT, EQ CO ROLE, EQ IS SUBEVENT OF.

The most important relation is EQ SYNONYM that expresses a one to one mapping between synsets in different languages. If one synset in one language matches more synsets in the other language, than the EQ NEAR SYNONYM relation is preferred.

The HAS EQ HYPERONYM and HAS EQ HYPONYM relations are typically used if a meaning is more specific than any available ILI-record.

5. Relations mapping
Table 2 presents a complete list of JMWNL relations including full and partial mapping, also non mapped relations and pointer symbols. Based on this table were generated the resource files for English WordNet 3.0 and EuroWordNet. For other lexical resources is just necessary to generate the adequate resource file using one of the provided templates.

6. Conclusion and Future Work
This paper presented an empirical study on mapping lexical and semantic relations between WordNet 1.6/EuroWordNet and the Princeton English WordNet version 3.0. This paper described our study on relations mapping, cross part of speech relations and the partial mapping of Fuzzynym relations like Derivationally Related From. We also showed the evolution of relations from EuroWordNet (mostly similar to WordNet 1.5) to WordNet 3.0. The modern WordNet has the tendency to have fewer relations for a better computability but looses a little linguistic expressivity.

Some relations could be mapped directly but others could not. A large number of EuroWordNet relations can be grouped to define one modern relation. This study was a necessary step during the development of JMWNL, to properly include EuroWordNet data in a coherent way, and to help build multilingual applications based on WordNet/EuroWordNet.

7. References
About WordNet. (2006). Retrieved November 6, 2007, from WordNet site: http://wordnet.princeton.edu/
Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. IJCAI-2003.

2 RELATION "state_of"
3 TARGET_CONCEPT
4 PART_OF_SPEECH "n"
4 LITERAL "grande"
5 SENSE 2
3 FEATURES
4 REVERSED
1 EQ_LINKS
2 EQ_RELATION "eq_synonym"
3 TARGET_ILI
4 PART_OF_SPEECH "a"
4 WORDNET_OFFSET 1052939
8. Appendix: Mapping Table

| WordNet 1.6 / EuroWordNet Relations | WordNet 3.0 Relations |
|-----------------------------------|-----------------------|
| **1. Nouns relations** | **Relation Symbol** |
| ANTONYM | ! | ANTONYM |
| NEAR_ANTONYM | ! | ANTONYM* |
| NEAR_SYNONYM | & | SIMILAR TO* |
| HAS_HYPERONYM | @ | HYPERNYM |
| HAS_HYPONYM | → | INSTANCE HYPONYM |
| HAS_INSTANCE | → | MEMBER MERONYM** |
| HAS_HOLO_MEMBER | %m | SUBSTANCE MERONYM** |
| HAS_HOLO_PART | %p | PART MERONYM** |
| HAS_HOLO_PORTION | %a | MEMBER HOLONYM** |
| HAS_MERO_MEMBER | #m | MEMBER HOLONYM** |
| HAS_MERO_PART | #p | PART HOLONYM** |
| HAS_MERO_PORTION | #s | SUBSTANCE HOLONYM** |
| XPOS_HYPERONYM | x@ | HYPERNYM (X POS)** |
| XPOS_NEAR_ANTONYM | x! | ANTONYM (X POS)** |
| XPOS_NEAR_SYNONYM | x& | SIMILAR TO (X POS)** |
| FUZZYNYM | + | DERIVATIONALLY RELATED FORM |
| XPOS_FUZZYNYM | x+ | DERIVATIONALLY RELATED FORM (X POS) |
| CAUSES | > | CAUSE |
| HAS_HOLONYM | % | |
| HAS_HOLO_MADEOF | %mo | |
| HAS_HOLO_LOCATION | %ml | |
| HAS_MERONYM | % | |
| HAS_MERO_MADEOF | #mo | |
| HAS_MERO_LOCATION | #m | |
| INVOLVED | i | |
| INVOLVED_AGENT | ia | |
| INVOLVED_INSTRUMENT | ii | |
| INVOLVED_LOCATION | il | |
| INVOLVED_RESULT | ir | |
| INVOLVED_SOURCE_DIRECTION | isd | |
| ROLE | r | |
| ROLE_AGENT | ra | |
| ROLE_DIRECTION | rd | |
| ROLE_INSTRUMENT | ri | |
| ROLE_LOCATION | rl | |
| ROLE_patient | rp | |
| ROLE_RESULT | rr | |
| ROLE_SOURCE_DIRECTION | rsd | |
| ROLE_TARGET_DIRECTION | rtd | |
| CO_AGENT_INSTRUMENT | cai | |
| CO_INSTRUMENT_AGENT | cia | |
| CO_ROLE | cr | |
| DERIVATION | d | |
| IS_DERIVED_FROM | <- | |
| STATE_OF | st | |
| BE_IN_STATE | ist | |
|                  |                                                                                     |
|-----------------|-------------------------------------------------------------------------------------|
| **IS CAUSED BY**| <                                                                                   |
| **HAS SUBEVENT**| has                                                                                  |
| **IS SUBEVENT OF** | is(e)                                                                               |
|                  | Instance Hypernym                                                                  |
| **IS** | Attribute                                                                           |
| **OF**          |                                                                                     |
|                  | Domain of synset - TOPIC                                                            |
|                  | Member of this domain - TOPIC                                                       |
|                  | Domain of synset - REGION                                                           |
|                  | Member of this domain - REGION                                                       |
|                  | Domain of synset - USAGE                                                            |
|                  | Member of this domain - USAGE                                                       |

2. Private Nouns Relations

| HAS HOLO MEMBER | %m MEMBER MERONYM                                                                 |
| HAS MERO MEMBER | #m MEMBER HOLONYM                                                                |

3. Verb Relations

| CAUSEN          | >                                                                                   |
| HAS HYPERONYM   | @                                                                                   |
| HAS HYPONYM     | ~                                                                                   |
| NEAR_ANTONYM    | !                                                                                   |
| **IS SUBEVENT OF** | *                                                                                   |
|                  | ENTAILMENT                                                                         |
| HAS XPOS HYPONYM| x~                                                                                   |
| NEAR_SYNONYM    | &                                                                                   |
| XPOS NEAR ANTONYM| xl                                                                                   |
| XPOS NEAR SYNONYM| x&                                                                                   |
| XPOS FUZZYNYM   | x+                                                                                   |
| IN MANNER       | im                                                                                   |
| INVOLVED        | i                                                                                   |
| INVOLVED_DIRECTION | id                                                                                   |
| INVOLVED_AGENT  | ia                                                                                   |
| INVOLVED_INSTRUMENT | ii                                                                                   |
| INVOLVED_LOCATION | il                                                                                   |
| INVOLVED_PATIENT | ip                                                                                   |
| INVOLVED_RESULT  | ir                                                                                   |
| INVOLVED_SOURCE_DIRECTION | irsd                                                                                   |
| INVOLVED_TARGET_DIRECTION | irtd                                                                                   |
| BE_IN_STATE     | ist                                                                                   |
| **IS CAUSED BY** | <                                                                                   |
| **HAS SUBEVENT** | has                                                                                  |
|                  | ^                                                                                   |
|                  | Also see                                                                            |
|                  | s                                                                                   |
|                  | Verb Group                                                                         |
|                  | z                                                                                   |
|                  | Domain of synset - TOPIC                                                            |
|                  | Domai of synset - REGION                                                            |
|                  | Domain of synset - USAGE                                                            |

4. Adjective Relations

| IS DERIVED FROM | \                                                                                   |
| IS DERIVED FROM | <                                                                                   |
| NEAR_ANTONYM    | !                                                                                   |
| NEAR_SYNONYM    | &                                                                                   |
| HAS HYPERONYM   | @                                                                                   |
| HAS HYPONYM     | ~                                                                                   |
| XPOS NEAR SYNONYM| x&                                                                                   |
| FUZZYNYM        | +                                                                                   |
| XPOS FUZZYNYM   | x+                                                                                   |
| DERIVATION      | d                                                                                   |
| HAS DERIVED     | <-                                                                                   |
| IS DERIVED FROM | <                                                                                   |
| **IS CAUSED BY** | <                                                                                   |
| **STATE OF**    | st                                                                                   |
|                  | Attribute                                                                          |
|                  | ^                                                                                   |
|                  | Also see                                                                            |
|                  | z                                                                                   |
|                  | Domain of synset - TOPIC                                                            |
|                  | Domai of synset - REGION                                                            |
|                  | Domain of synset - USAGE                                                            |
Table 2: EuroWordNet and WordNet relations correspondence; In black are the relations that could be directly mapped, in blue the new defined relations and in red the relations that didn’t have a correspondent either in EuroWordNet or WordNet.

* In EuroWordNet is preferred a loose synonymy and antonymy relation

** In EuroWordNet HAS_MERO/HOLO express respectively HOLONYM/MERONYM relations

*** XPOS Relations – In WordNet this relations are not present. We introduced them in order to preserve all EuroWordNet relations.