

Towards semi-automatic methods for improving WordNet

Nervo Verdezoto
University of Trento &
LOA-ISTC-CNR, Trento
nverdezoto@gmail.com

Laure Vieu
IRIT-CNRS, Toulouse &
LOA-ISTC-CNR, Trento
vieu@irit.fr

Abstract

WordNet is extensively used as a major lexical resource in NLP. However, its quality is far from perfect, and this alters the results of applications using it. We propose here to complement previous efforts for “cleaning up” the top-level of its taxonomy with semi-automatic methods based on the detection of errors at the lower levels. The methods we propose test the coherence of two sources of knowledge, exploiting ontological principles and semantic constraints.

1 Introduction

WordNet (Princeton WordNet (Fellbaum, 1998), henceforth WN) is a lexical resource widely used in a host of applications in which language or linguistic concepts play a role. For instance, it is a central resource for the quantification of semantic relatedness (Budanitsky and Hirst, 2006), in turn often exploited in applications. The quality of this resource therefore is very important for NLP as a whole, and beyond, in several AI applications. Neel and Garzon (2010) show that the quality of a knowledge resource like WN affects the performance in recognizing textual entailment (RTE) and word-sense disambiguation (WSD) tasks. They observe that the new version of WN induced improvements in recent RTE challenges, but conclude that WN currently is not rich enough to resolve such a task. What is more, its quality may be too low to even be useful at all. Bentivogli et al. (2009) discuss the results¹ of 20 “ablation tests” on systems submitted to the main RTE-5 task in which WN (alone) was ablated: 11 of these tests demonstrated that the use of this resource has a positive impact (up to 4%) on the performance of the systems but 9 showed a negative (up to 2% improvement when ablated) or null impact.

In the area of automatic recognition of part-whole relations, Girju and Badulescu (2006) proposed a learning method relying on WN’s taxonomy. Analyzing the classification rules obtained, we could see that WN taxonomical errors lead to absurd rules, which can explain wrong recognition results. For instance, the authors obtain pairs such as $\langle \textit{shape}, \textit{physical phenomenon} \rangle$ and $\langle \textit{atmospheric phenomenon}, \textit{communication} \rangle$ as positive constraints for part-whole recognition, while sentences like *a curved shape is part of the electromagnetic radiation* or *rain is part of this document* would make no sense.

Some semantic problems of WN are well-known: confusion between concepts and individuals (in principle solved since WN 2.1), heterogeneous levels of generality, inappropriate use of multiple inheritance, confounding and missing senses, and unclear glosses (Kaplan and Schubert, 2001; Gangemi et al., 2003; Clark et al., 2006). Nevertheless, the number of applications where WN is used as an ontology has been increasing. In fact, apart from the synonymy relation on which synsets are defined, the hyponymy/hypernymy relation is WN’s semantic relation most exploited in applications; it generates WN’s taxonomy, which can be seen as a lightweight ontology, something it was never designed for, though. Several works tried to address these shortcomings. Gangemi et al. (2003) proposed a manual restructuring through the alignment of WN’s taxonomy and the foundational ontology DOLCE², but this restructuring just focused on the upper levels of the taxonomy. Applying formal ontology principles

¹http://www.aclweb.org/aclwiki/index.php?title=RTE5_-_Ablation_Tests

²See (Masolo et al., 2003) and <http://www.loa-cnr.it/DOLCE.html>

(Guarino, 1998) and the OntoClean methodology (Guarino and Welty, 2004) have also been suggested for manually “cleaning up” the whole resource. This however is extremely demanding, because the philosophical principles involved require a deep analysis of each concept, and as a result, is unlikely to be achieved in a near future. Clark et al. (2006) also gave some general suggestions as design criteria for a new WN-like knowledge base and recommended that WN should be cleaned up to make it logically correct, but did not provide any practical method for doing so. Two other more extensive works rely on manual interventions, either the mapping of each synset in WN to a particular concept in the SUMO ontology (Pease and Fellbaum, 2009), or the tagging of each synset in WN with “features” from the Top Concept Ontology (Alvez et al., 2008) to substitute or contrast the original WN taxonomy. Such approaches are clearly very costly, as each synset needs to be examined. In addition, the ontological value of these additional resources themselves remains to be proven. The method used in (Alvez et al., 2008) has though helped pointing out a large number of errors in WN 1.6.

Our purpose in this paper is to show that automatic methods to spot errors, especially in the lower levels of WN’s taxonomy, can be developed. Spotting errors can then efficiently direct the manual correction task. Such methods could be used to complement a manual top-level restructuring and could be seen as an alternative to fully manual approaches, which are very demanding and in principle require validation between experts. Here, we explore methods based on internal coherence checks within WN, or on checking the coherence between WN and annotated corpora such as those of Semeval-2007 Task 4 (Girju et al., 2007).

The paper is structured as follows: Section 2 presents the data used and the methodology; Section 3 discusses the results; Section 4 concludes, exploring how the method could be extended and applied.

2 Methodology

To spot errors in WN, our basic idea is to contrast two sources of knowledge and automatically check their coherence. Here, we contrast part-whole data with WN taxonomy structure, on the basis of constraints stemming from the semantics of the part-whole relations and ontological principles. The part-whole data used is taken either from the meronymy/holonymy relations of WN or from available annotated corpora.

An incoherence between two sources of knowledge may be caused by an error in either one (or both). Contrasting part-whole data with the taxonomy will indeed help detecting errors in the taxonomy—the most numerous—but errors are also found in the part-whole data itself (see Section 3.3).

2.1 Extracting the Dataset

We started extracting WN taxonomy from the hypernym relations in the current version of WN (3.0), a network of 117,798 nouns grouped in 82,155 synsets. We also extracted WN meronymy relations, i.e., 22,187 synset pairs, split into 12,293 “member”, 9,097 “part” and 797 “substance”, to constitute the first part-whole dataset. In order to replicate our methodology, we also extracted 89 part-whole relation word pairs annotated with WN senses from the SemEval-2007 Task 4 datasets (Girju et al., 2007). We kept the positive examples from the training and test datasets,³ excluding redundant pairs, and correcting a couple of errors. This data is also annotated with the meronymy sub-relations inspired from the classification of Winston et al. (1987), but five subtypes instead of WN’s three, although “member-collection” can safely be assumed to correspond to WN’s “member” meronymy. We will call this sub-relation *Member*, be it from WN or from SemEval.

We also tried to get similar datasets from the SemEval-2010 Task 8 but, not being annotated with WN senses, they are useless for our purposes. Figure 1 illustrates a WN-extracted meronymy pair from our corpus⁴, encoded in our own xml format. Synsets are presented with the standard WN sense keys for each word, the recommended reference for stability from one WN release to another.⁵

³<http://nlp.cs.swarthmore.edu/semeval/tasks/task04/data.shtml>

⁴Available at <http://www.loa-cnr.it/corpus/corpus.tar.gz>

⁵A sense key combines a lemma field and several codes like the synset type and the lexicographer id. See <http://>

```

<pair relationOrder="(e1, e2)" comment="meronymy_part" source="WordNet-3.0">
  <e1 synset="head%1:06:04" isInstance="No">
    <hypernym>
      {obverse%1:06:00}...{surface%1:06:00}...{artifact%1:03:00}...{physical_object%1:03:00}{entity%1:03:00}
    </hypernym>
  </e1>
  <e2 synset="coin%1:21:02" isInstance="No">
    <hypernym>
      ...{metal_money%1:21:00}{currency%1:21:00}...{quantity%1:03:00}{abstract_entity%1:03:00}{entity%1:03:00}
    </hypernym>
  </e2>
</pair>

```

Figure 1: Example pair from the annotated dataset

2.2 The Tests

2.2.1 Ontological constraints

The semantics of the part-whole relation on which the meronymy/holonymy relations are founded involves ontological constraints: in short, the part and the whole should be of a similar nature. Studies in Mereology show that part-whole relations occur on all sub-domains of reality, concrete or abstract (Simons, 1987; Casati and Varzi, 1999). As a few cognitively oriented works explicitly state, the part and the whole should nevertheless belong to the same subdomain (Masolo et al., 2003; Vieu and Aurnague, 2007). Other work, e.g., the influential (Winston et al., 1987), more or less implicitly exploit this homogeneity constraint. Our tests examine and compare the nature of the part and the whole in attested examples of meronymy, looking for incoherences. Here we use only a few basic ontological distinctions, namely, the distinction between:

- endurants (ED) or physical entities (like a dog, a table, a cave, smoke),
- perdurants (PD) or eventualities (like a lecture, a sleep, a downpour), and
- abstract entities (AB — like a number, the content of a text, or a time).

These are only three of the four topmost distinctions in DOLCE (Masolo et al., 2003), that is, we actually group qualities (\mathcal{Q} , the fourth top-level category) into abstract entities here.

Tests 1–3 are directly aimed at detecting ontological heterogeneity in meronymy pairs that mix the three categories ED, PD and AB, as just explained. The tests are queries on our corpus to extract and count meronymy pairs (pairs of synsets of the form $\langle e1, e2 \rangle$ where $e1$ is the part and $e2$ is the whole) that involve an ontological heterogeneity. **Test 1** focuses on pairs mixing endurants and abstract entities (pairs of type $\langle ED, AB \rangle$ or $\langle AB, ED \rangle$), **Test 2** on endurants and perdurants ($\langle ED, PD \rangle$ or $\langle PD, ED \rangle$) and **Test 3** on perdurants and abstract entities ($\langle PD, AB \rangle$ or $\langle AB, PD \rangle$).

However, WN 3.0’s top-level is not as simple as DOLCE’s, so to recover the three basic categories we had to group several classes from different WN branches. In particular perdurants are found both under *physical_entity%1:03:00* (*process%1:03:00*) and under *abstraction%1:03:00* (*event%1:03:00* and *state%1:03:00*). The map we first established was then as follows:

- ED = *physical_entity%1:03:00* \ *process%1:03:00*;
- PD = *process%1:03:00* \cup *event%1:03:00* \cup *state%1:03:00*;
- AB = *abstraction%1:03:00* \ (*event%1:03:00* \cup *state%1:03:00*).

Since all groups in WordNet are under *abstraction%1:03:00* irrespective of the nature of the members, it was obvious from the start that most “member” meronymy pairs would be caught by Tests 1 or 3. This is the reason why groups were actually removed from AB so the final map posited:

- AB = *abstraction%1:03:00* \ (*event%1:03:00* \cup *state%1:03:00* \cup *group%1:03:00*).

2.2.2 Semantic constraints

Two more tests were designed to check basic semantic constraints involved in meronymy relations.

Test 0 is related to the problem of confusion between classes and individuals evoked above and checks for meronymy pairs between an individual and a class. Meronymy in WN applies to pairs of classes and to pairs of individuals, but mixed pairs are also found, either between a class and an individual or between an individual and a class. The semantics of WN meronymy is not precisely described in Fellbaum (1998), but observing the data, the following appears to fit the semantics of “is a meronym of” between two classes A and B : the disjunction of the formulas “for all/most instances a of A , there is an instance b of B such that $P(a, b)$ ” and “for all/most instances b of B , there is an instance a of A such that $P(a, b)$ ”, where P is the individual-level part-whole relation. On this basis, a meronymy between a class A and an individual b would simply mean: “for all/most instances a of A , $P(a, b)$ ”, while a meronymy between an individual a and a class B would mean: “for all/most instances b of B , $P(a, b)$ ”. The former can make sense, cf. $\langle \text{surat}\%1:10:00, \text{koran}\%1:10:00 \rangle$ (all suras are part of the Koran). However, the latter would imply that all (most) instances of the class would share a same part, i.e., they would overlap. That the instances of a given class all overlap is of course not logically impossible, but it is highly unlikely for lexical classes. The purpose of Test 0 is to check for such cases, expected to reveal confusion between individuals and classes, that is, errors remaining after the introduction of the distinction in WN 2.1.⁶

Test 4 is dedicated to the large number of `Member` pairs in WN and SemEval data, somehow disregarded by the removal of groups from `AB` above. The semantics of this special case of meronymy clearly indicates that the whole denotes some kind of group, e.g., a collection or an organization, and that the part is a member of this group (Winston et al., 1987; Vieu and Aurnague, 2007). Group concepts in WN are hyponyms of $\text{group}\%1:03:00$. A last coherence check, done by Test 4, thus extracts the `Member` pairs in which the whole is not considered a group because it is not an hyponym (or instance) of $\text{group}\%1:03:00$.

3 Results, Analysis and Discussion

Table 1: Number of pairs extracted by the tests

Error Category	Test	WordNet	SemEval
Semantic	0	349 1.57%	0 0%
	4	550 4.47%	7 7.87%
Ontological	1	163 1.62%	2 2.78%
	2	45 0.45%	2 2.78%
	3	108 1.07%	0 0%

The number of pairs extracted by our queries are summarized on Table 1. The error rates are quite low, ranging from 0 to 7.87% depending on the data set of meronymy pairs (WN or SemEval). The highest error rate is provided by Test 4: 550 (4.47%) of the 12,293 WN `Member` pairs and 7 (7.87%) of 19 `Member` pairs in SemEval dataset were identified as semantic errors because the whole is not a group in WN taxonomy. Test 0 has the lowest rate, just 349 (1.57%) of 22,187 WN meronymy pairs are suspected of confusing classes and individuals. More important than the error rate is that the tests achieved maximal precision. After manual inspection of all the suspect pairs extracted, it turns out all the pairs indeed suffered from some sort of error or another. Of course, the few tests proposed here cannot aim at spotting all the taxonomy errors in WN, i.e., recall surely is low, but their precision is a proof of the effectiveness of the method proposed, which can be extended by further tests to uncover more errors.

For Tests 1–3, since the three categories `ED`, `PD` and `AB` are large and diverse, the analysis of the errors started with looking for regularities among the taxonomic chains of hypernyms of the synsets in

⁶Another, very simple and superficial test could be to check synsets for names with capital letters. This of course doesn’t rely on ontological knowledge.

the pairs. In particular, we looked for taxonomic generalizations of sets of pairs to divide the results in meaningful small sets. These sets were manually examined in order to check the intended meaning of the meronymy relations and determine the possible problems, either in the taxonomy or in the meronymy; for this we used all the information provided by WordNet as synset, synonymy, taxonomy, and glosses. For Tests 0 and 4, similar regularities could be observed. Several regularities denote a few systematic errors relatively easily solved using standard ontological analysis, described in the Sections 3.1–3.5.

3.1 Confusion between class and group

Several individual collections e.g., *new_testament%1:10:00*, organizations e.g., *palestine_liberation_organization%1:14:00*, and genera e.g., *genus_australopithecus%1:05:00* are considered as classes in WN instead of groups (errors extracted with Test 0). The first example, *new_testament%1:10:00*, is glossed as “the collection of books ...”, but is not considered as an instance of group, it is a subclass of *document%1:10:00*.⁷ The latter two are seen as subclasses instead of instances of group; this would mean that all instances of *palestine_liberation_organization%1:14:00* (whatever these could be) and all instances of *genus_australopithecus%1:05:00* (which makes more sense) actually are groups. But if there are instances of the genus *Australopithecus* at all, these are individual hominids, not groups. In fact, the hesitation of the lexicographer is visible here, since *lucy%1:05:00* is both a Member of *genus_australopithecus%1:05:00* and an instance of *australopithecus_afarensis%1:05:00*, a subclass of *hominid%1:05:00* (not of group). To show further the confusion here, *australopithecus_afarensis%1:05:00* itself also is a Member of *genus_australopithecus%1:05:00*, which, with the semantics of Member between classes, would mean that instances of *australopithecus_afarensis%1:05:00* are members of instances of *genus_australopithecus%1:05:00*, which is clearly not adequate.

Despite this confusion, dealing with collections, organizations and groups as individuals poses no real problem. The Member meronymy is adequately used elsewhere in WN to relate individuals (e.g., *balthazar%1:18:00*, an instance of *sage%1:18:00*, is a Member of *magi%1:14:00*, an instance of *collection%1:14:00*). Dealing with biological genera is arguably more complex, as one can see them both as classes whose instances are the individual organisms, and as individuals which are instances of the class *genus%1:14:00*. A first-order solution to this dilemma, which applies more generally to socially defined concepts, proposes to consider concepts (and genera) as individuals, and to introduce another sort of instance relation for them (Masolo et al., 2004). Beyond genera, related problems occur with the classification of biological orders, divisions, phylums, and families, most of which are correctly considered as groups (e.g., *chordata%1:05:00*), except for a few, pointed out by Test 4 (e.g., *amniota%1:05:00*, *arenaviridae%1:05:00*). All these though should be group individuals, not group classes as now in WN.

3.2 Confusion between class and individual which is a specific instance of the class

Test 0 also points at a few errors where a class is confused with a specific instance of this class. This error corresponds to a missing sense of the word, used with a specific sense. Examples include the individual-class pairs $\langle \textit{great_divide}\%1:15:00, \textit{continental_divide}\%1:15:00 \rangle$,⁸ $\langle \textit{saturn}\%1:17:00, \textit{solar_system}\%1:17:00 \rangle$, $\langle \textit{renaissance}\%1:28:00, \textit{history}\%1:28:00 \rangle$, in which the continental divide at stake is not any one but that of North America, the solar system, ours, and the history, the history of mankind. Sometimes the gloss itself makes it clear that the lexicographer wanted to do two things at a time; cf. for *continental_divide%1:15:00*: “the watershed of a continent (especially the watershed of North America formed by a series of mountain ridges extending from Alaska to Mexico)”.

⁷This particular error doesn't show again with Test 4 because the meronyms of *new_testament%1:10:00* are “part” meronyms, not Member meronyms.

⁸WN has chosen a restrictive sense for the Great Divide, making it a proper part of the Continental Divide. In other interpretations these two names are synonyms.

3.3 Confusion between meronymy and other relations

The meronymy relation itself can be wrong, that is, it is confused with other relations, especially “is located in” (*balkan_wars%1:04:00, balkan_peninsula%1:15:00*) (Test 2), (*nessie%1:18:00, loch_ness%1:17:00*) (Test 1); “participates in” (*feminist%1:18:00, feminist_movement%1:04:00*), (*air%1:27:00, wind%1:19:00*) (Test 2); “is a quality of” (*personality%1:07:00, person%1:03:00*), (*regulation_time%1:28:00, athletic_game%1:04:00*) (Test 3); or still other dependence relations such as in (*operating_system%1:10:00, platform%1:06:03*) (Test 1). Diseases and other conditions regularly give rise to a confusion with “participates in” or its inverse, as with (*cancer_cell%1:08:00, malignancy%1:26:00*), (*knock-knee%1:26:00, leg%1:08:01*), and (*acardia%1:26:00, monster%1:05:00*) (Test 2).

3.4 Confusion between property (AB) and an entity (ED or PD) having that property

A regular confusion occurs between an entity and a property of that entity, for instance a shape, a quantity or measure, or a location. Similarly, confusions occur between a relation and an ED or PD being an argument of that relation. Examples are extracted mostly with Tests 1 and 3, but a few examples are also found with Tests 2 and 4, when several problems co-occurred. Such confusions lead to wrong taxonomic positions: *coin%1:21:02, haymow%1:23:00* and *tear%1:08:01* are attached under *quantity%1:03:00* (AB), while the intuition as well as the glosses make it clear that a coin is a flat metal piece and a haymow a mass of hay, that is, concrete physical entities under ED; similarly, *corolla%1:20:00* and *mothball%1:06:00* are attached under *shape%1:03:00* (AB), while there are clearly ED.

Regularities group together some cases, e.g., many hyponyms of *helping%1:13:00* (*drumstick, fillet, sangria...*) are spotted because *helping%1:13:00* is under *small_indefinite_quantity%1:23:00* (AB). It turns out that *small_indefinite_quantity%1:23:00* and its direct hypernym *indefinite_quantity%1:23:00* cover more physical entities of a certain quantity rather than quantities themselves. The tests reveal similar errors at higher levels in the hierarchy: *possession%1:03:00* “anything owned or possessed” is attached under *relation%1:03:00* “an abstraction belonging to or characteristic of two entities or parts together” (AB), that is, the object possessed is confused with the relation of possession. Test 1 points at this error 16 times (e.g., *credit_card%1:21:00* and *hacienda%1:21:00*, clearly not abstracts, are spotted this way). Another important mid-level error of this kind is that *part%1:24:00*, while glossed “something determined in relation to something that includes it”, is attached under *relation%1:03:00* (AB) as well. As a result, all its hyponyms, for instance, *news_item%1:10:00*, and notably, *substance%1:03:00* “the real physical matter of which a person or thing consists” and all its hyponyms (e.g., *dust%1:27:00, beverage%1:13:00*) are considered abstract entities.⁹

3.5 Confusion between two senses of a word

All the tests yield errors denoting missing senses of some words in WN. Test 4 shows that `Member` is systematically used between a national of a country and that individual country, e.g. (*ethiopian%1:18:00, ethiopia%1:15:00*), thus referring to the sense of *country* as “people of that nation”. But while the word *country* has both the “location” and the “people” senses (among others) in WN, individual countries do not have multiple senses and are all instances of *country%1:15:00*, the “location” sense.

Similarly, hyponyms of *natural_phenomenon%1:19:00* (PD) are often confused with the object (ED) involved, i.e., the participant to the process, revealing missing senses (examples extracted with Test 2). *Precipitation* has (among others) two senses, *precipitation%1:23:00* “the quantity of water falling to earth” (a quantity, AB), and *precipitation%1:19:00* “the falling to earth of any form of water” (a natural phenomenon, PD). The actual water fallen (ED), is missing, as revealed by the pair (*ice_crystal%1:19:00, precipitation%1:19:00*) (from Test 2).

Other errors of this kind are more sporadic, as with (*golf_hole%1:06:00, golf_course%1:06:00*) (*golf hole* has only a “playing period” sense, its “location” sense is missing, from Test 1), and (*coma%1:17:00,*

⁹*substance%1:03:00* acquires though a physical entity character through multiple inheritance, since it also has matter and physical entity as hypernyms. It is not obvious why multiple inheritance has been used here.

comet%1:17:00 (*coma* has only a “process” sense, its “physical entity” sense is missing, from Test 2).

3.6 Polysemy in WordNet

The last two types of error, 3.4 and 3.5, point at polysemy issues, as well as the few cases of 3.2. There are two strategies to address polysemy in WN. The main one is the distinction of several synsets for the different senses of a word, but there is also the use of multiple inheritance that gives several facets to a single synset. The literature on WN doesn’t make it clear why and when to use multiple inheritance rather than multiple synsets, and it appears that lexicographers have not been methodical in its use. Some cases of “dot objects” (Pustejovsky, 1995) have been accounted this way. For instance, *letter%1:10:00* inherits both its abstract content from its hypernym *text%1:10:00* (AB) and its physical aspect from its hypernym *document%1:06:00* (ED). However, the polysemy of *book*, the classical similar case, is not accounted for in this way: *book%1:10:00* only is ED. And while *document* has two separate senses, *document%1:10:00* (AB) and *document%1:06:00* (ED), there is no separate abstract sense for *book*. Test 1 points at this problem with the pair $\langle \textit{book_of_psalms}\%1:10:01, \textit{book_of_common_prayer}\%1:10:00 \rangle$, where the part is a sub-class (rather than an instance, but this is an additional problem pointed by Test 0) of *book%1:10:00* (ED), while the whole is an instance of *sacred_text%1:10:00*, a *communication%1:03:00* (AB).

As far as polysemy standardly accounted with multiple senses goes, our tests point at a need for a more principled use there as well. In particular, the polysemy accounted for at a given level is often not reproduced at lower levels, as just observed for *document* and *book*. We also have seen above that the polysemy of the word *country* is not “inherited” by individual countries. Similarly the polysemy of *precipitation* has no repercussion on that of *rain*, which has a sense *rain%1:19:00* under *precipitation%1:19:00*, and none under *precipitation%1:23:00* (on the other hand, the material sense of *rain*, *rain%1:27:00* “drops of fresh water that fall”, an ED, lacks for *precipitation*).

A few pairs extracted with Test 4 show the hesitation of the lexicographer between the classification of a collection as a group, and a classification that accounts for the nature of the collection elements. For instance *constellation%1:17:00* and *archipelago%1:17:00* have members but are ED, while *galaxy%1:14:00* is a group. This could be properly addressed by splitting the group category, erroneously situated among abstract entities anyway, into different group categories (e.g., one for each of ED, PD and AB), or exploit multiple inheritance if compatible with its regimentation.

3.7 Difficult ontological issues

Although all the pairs retrieved by our tests point at (one or several) errors, in a few cases, these are not solved easily. In particular, difficult ontological issues are faced with fictional entities. WN classifies most of these under *psychological_feature%1:03:00* (AB). However, these fictional entities often show very similar properties to those of concrete entities. As a result, some of them are classified as ED or PD, e.g., *acheron%1:17:00* is an instance of *river%1:17:00* (ED), while being somehow recognized as fictional since it is a meronym of *hades%1:09:00*, a subclass (here again, not an instance, an additional problem) of *psychological_feature%1:03:00* (AB), something pointed out by Test 1. Others have concrete parts, e.g. we find the pair $\langle \textit{wing}\%1:05:00, \textit{angel}\%1:18:00 \rangle$ among the cases of $\langle \text{ED,AB} \rangle$, i.e. Test 1 results. Angel wings (and feathers, etc.) are of course of a different nature than bird wings, and hellish rivers are not real rivers, but how to distinguish them without duplicating most concrete concepts under *psychological_feature%1:03:00* (AB) is unclear.¹⁰

Another regular anomaly is found with roles and relations, e.g., with pairs like $\langle \textit{customer}\%1:18:00, \textit{business_relation}\%1:24:00 \rangle$, an $\langle \text{ED,AB} \rangle$ case (Test 1). A straightforward analysis saying that meronymy has been confused with participation (cf. 3.3) would overlook the fact that the customer role is defined by the business relation itself, i.e., that the dependence is even tighter. Since currently in WN, *customer%1:18:00* simply is a sub-class of *person%1:03:00* (ED), in any case the classical issues related to

¹⁰Although the ontological nature of fictional entities is discussed in metaphysics (see, e.g., (Thomasson, 1999)), how to deal with their “concrete” aspects is not a central issue.

the representation of roles are not addressed, and a more general solution should be looked for, perhaps along the lines of (Masolo et al., 2004).

3.8 Small errors

Finally, our tests identify a few isolated WN errors, which can be seen as small slips, such as for instance a wrong sense selected in the meronymy, e.g., $\langle seat\%1:06:01, seating_area\%1:06:00 \rangle$ where $seat\%1:15:01$ (the area, not the chair) should have been selected,¹¹ or a wrong taxonomical attachment, that is, a wrong sense selected for an hypernym, e.g., $infrastructure\%1:06:01$ is an hyponym of $structure\%1:07:00$, a property, instead of $structure\%1:06:00$, an artifact (from the pair $\langle infrastructure\%1:06:01, system\%1:06:00 \rangle$ extracted with Test 1).

3.9 Types of solutions

As can be observed, tests do not all point at a unique type of problem, nor suggest a unique type of solution. Basically, there are five kinds of formal issues underlying the types of errors analyzed above, each calling for different modifications of WN:

- a synset is considered as a class but should be an individual (3.1): need to change its direct hypernym link into an instance-of link, possibly changing as well the attachment point in the taxonomy;
- a synset is not attached to the right place in the taxonomy (3.4, 3.8): need to move it in the taxonomy;
- a synset mixes two senses (3.2, 3.5): need to introduce a missing sense, either attached elsewhere in the taxonomy or as instance of the synset at hand;
- the meronymy relation is confused with another one (3.3): need to remove it (or change it for another sort of relation when this is introduced in WN);
- the meronymy relation is established between the wrong synsets (3.8): need to change one of the two synsets related by another sense of a same word.

In some cases, the problems should be addressed through more general cures, at a higher level in the taxonomy (3.4) or by imposing more systematic modeling choices (3.6, 3.7).

4 Looking forward

We showed in this paper that automatic methods can be developed to spot errors in WN, especially in the hyperonymy relations in the lower levels of the taxonomy. The query system based on ontological principles and semantic constraints we proposed was very effective, as all the items retrieved did point to one or more errors. With such generic tests though, a manual analysis of the extracted examples by lexicographers, domain or ontological experts is necessary to decide on how the error should be solved. However, this same analysis showed many regularities pointing at standard ontological errors, which suggested that the tests can be much refined to limit the variety of issues caught by a single test and that simple repair guidelines can be written.

This work can therefore be developed in several directions. On the one hand, the same tests can be exploited further by expanding the meronymy datasets, for instance if some annotated corpus similar to the SemEval2007 datasets becomes available. The range of tests can be extended as well. For instance, one can make further coherence tests exploiting meronymy data, refining or complementing the Tests 0–4 presented here. The class of abstract entities AB groups a variety of concepts, so incompatible combinations of subclasses are certainly present in $\langle AB, AB \rangle$ pairs (e.g., across $relation\%1:03:00$, $psychological_feature\%1:03:00$, or $measure\%1:03:00$), suggesting new tests. Without considering to remove groups from abstract entities, cases of incoherence involving groups could also be addressed by checking

¹¹This is extracted with Test 1, because an additional problem appears with $seating_area\%1:06:00$ (or rather with its direct hypernym $room\%1:23:00$), which is under $spatial_relation\%1:07:00$ (AB) rather than area and location (ED). This shows that the error in the meronymy relation would in principle require finer-grained tests to be found.

the compatibility of the ontological categories of their members. Among the class of physical entities ED, we disregarded the presence of location entities, so new tests could also examine incompatible combinations of subclasses of ED. Finally, we could check whether the “substance” meronymy relation indeed involves substances, in a similar way as Test 4 for groups. Additional tests can be considered using other knowledge sources than meronymy data. Within WN, we could exploit the semantics of tagged glosses (cf. Princeton WordNet Gloss Corpus) in order to check the coherence with the taxonomy. And since WN is more than a network of nouns, others relations can be exploited, for instance between nouns and verbs. Similarly, SemEval datasets deal with other relations than the one exploited here: from other subtypes of meronymy (e.g., “place-area”), to any of the semantic relations analyzed in the literature (e.g., “instrument-agency”). In particular, relations involving thematic roles are quite easily associated with ontological constraints and so can constitute the basis for further tests.

On the other hand, methods aiming at improving the quality of WN can be concretely built on the basis of these tests. A semi-automatic tool for “cleaning-up” WN could be fully developed, which could contribute to the next, improved, version of WN. The analysis of regular errors made in WN could simply lead to *guidelines* to help lexicographers avoid classical ontological mistakes. Such guidelines could be used for the extension of Princeton WN, e.g., for new domains. They could be used also during the creation of new WordNets for other languages, suggesting at the same time to abandon the common practice of simply importing the taxonomy of Princeton WN, importing also its errors. These two ideas could be combined in creating a tool to assist the development of WordNets by automatically checking errors and pointing out them in the development phase. This could well complement the TMEO methodology, based on ontological distinctions, used during the creation of the Sensocomune computational lexicon (Oltramari et al., 2010).

Acknowledgements

We wish to thank Alessandro Oltramari for his contribution to the initial stages of this work, Laurent Prévot for fruitful discussions on this topic and comments on a previous draft, Emanuele Pianta and three anonymous reviewers for their comments. This work has been supported by the LOA-ISTC-CNR and the ILIKS joint European laboratory.

References

- Alvez, J., J. Atserias, J. Carrera, S. Climent, E. Laparra, A. Oliver, and G. Rigau (2008). Complete and consistent annotation of WordNet using the Top Concept Ontology. In *Proceedings of LREC2008*, pp. 1529–1534.
- Bentivogli, L., I. Dagan, H. T. Dang, D. Giampiccolo, and B. Magnini (2009). The Fifth PASCAL Recognizing Textual Entailment Challenge. In *Proceedings of TAC 2009 Workshop*, Gaithersburg, Maryland, USA.
- Budanitsky, A. and G. Hirst (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics* 32(1), 13–47.
- Casati, R. and A. Varzi (1999). *Parts and Places - The Structures of Spatial Representation*. Cambridge, MA: MIT Press.
- Clark, P., P. Harrison, T. Jenkins, J. Thompson, and R. Wojcik (2006). From WordNet to a Knowledge Base. In C. Baral (Ed.), *Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering. Papers from the 2006 AAI Spring Symposium*, pp. 10–15. AAI Press.
- Fellbaum, C. (Ed.) (1998). *WordNet. An Electronic Lexical Database*. Cambridge (MA): MIT Press.

- Gangemi, A., N. Guarino, C. Masolo, and A. Oltramari (2003). Sweetening WordNet with DOLCE. *AI Magazine* 24(3), 13–24.
- Girju, R. and A. Badulescu (2006). Automatic Discovery of Part-Whole Relations. *Computational Linguistics* 32(1), 83–135.
- Girju, R., V. Nastase, and P. Turney (2007). SemEval-2007 Task 04: Classification of Semantic Relations between Nominals. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 13–18. Association for Computational Linguistics.
- Guarino, N. (1998). Some ontological principles for designing upper level lexical resources. In A. Rubio, N. Gallardo, R. Castro, and A. Tejada (Eds.), *First International Conference on Language Resources and Evaluation*, pp. 527–534. European Language Resources Association.
- Guarino, N. and C. Welty (2004). An overview of OntoClean. In S. Staab and R. Studer (Eds.), *Handbook on Ontologies*, pp. 151–159. Springer-Verlag.
- Kaplan, A. N. and L. K. Schubert (2001). Measuring and Improving the Quality of World Knowledge Extracted From WordNet. Technical Report 751, University of Rochester.
- Masolo, C., S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari (2003). The WonderWeb library of foundational ontologies and the DOLCE ontology. WonderWeb (EU IST project 2001-33052) deliverable D18, LOA-ISTC-CNR.
- Masolo, C., L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi, and N. Guarino (2004). Social roles and their descriptions. In D. Dubois and C. Welty (Eds.), *Proceedings of the 9th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2004)*, pp. 267–277. Menlo Park (CA): AAAI Press. Whistler June, 2-5, 2004.
- Neel, A. and M. Garzon (2010). Semantic Methods for Textual Entailment: How Much World Knowledge is Enough? In *Proceedings of FLAIRS 2010*, pp. 253–258.
- Oltramari, A., G. Vetere, M. Lenzerini, A. Gangemi, and N. Guarino (2010). Senso comune. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, and D. Tapias (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, pp. 3873–3877. European Language Resources Association (ELRA).
- Pease, A. and C. Fellbaum (2009). Formal ontology as interlingua: the SUMO and WordNet linking project and Global WordNet. In C.-R. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, and L. Prévot (Eds.), *Ontology and the Lexicon. A Natural Language Processing Perspective*, pp. 31–45. Cambridge University Press.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge (MA): MIT Press.
- Simons, P. (1987). *Parts - A study in ontology*. Oxford: Clarendon Press.
- Thomasson, A. (1999). *Fiction and Metaphysics*. Cambridge University Press.
- Vieu, L. and M. Aurnague (2007). Part-of relations, functionality and dependence. In M. Aurnague, M. Hickmann, and L. Vieu (Eds.), *The Categorization of Spatial Entities in Language and Cognition*, pp. 307–336. Amsterdam: John Benjamins.
- Winston, M., R. Chaffin, and D. Herrmann (1987). A taxonomy of part-whole relations. *Cognitive Science* 11(4), 417–444.