

Grounding the Ontology on the Semantic Interpretation Algorithm

Fernando Gomez

Dept. of Computer Science
University of Central Florida, Orlando, FL 32816
Email: gomez@cs.ucf.edu

Abstract. Some reorganizations and modifications to the WordNet ontology are explained. These changes have been suggested by extensive testing of the ontological categories with an algorithm for semantic interpretation. The algorithm is based on predicates that have been defined for WordNet verb classes. The selectional restrictions of the predicates are WordNet ontological categories.

1 Introduction

This paper, a much shorter version of CS-TR-01-01 with the same title, provides a sample of our reorganizations and changes to the WordNet noun ontology (WordNet 1.6) [6]. These changes have been dictated by a semantic interpretation algorithm reported in [3]. The algorithm is based on predicates, or verbal concepts, that have been defined for WordNet verb classes [2]. The semantic roles of the predicates have been linked to the noun ontology and to syntactic relations. After the initial set up, the definition of new predicates has been followed by testing them using the algorithm. As of this writing, 3000 predicates have been defined and 95% of WordNet verb classes have been mapped into these predicates. In contrast to other ontologies for natural language [1,5], or to efforts to induce a concise set of ontological categories from WordNet [4], the principles guiding our changes have been the selectional restrictions in the semantic roles of the 3000 predicates. Hence, the failure of interpreting a sentence has been the clue for redefining some ontological categories. For instance, the concept *written-communication*, which has many subconcepts, is categorized in Wordnet 1.6 only as an *abstraction*. Thus, the interpreter failed to interpret such simple sentences as “She burned the letter/She put the letter on the table,” because “letter” does not have *physical-thing* as one of its hypernyms (superconcepts). In “The fish frequently hides in a crevice,” the interpreter failed to assign meaning to “hides” because “crevice” is categorized in WordNet 1.6 only as an *abstraction*. In “Blood poured from the wound,” the interpreter fails to assign meaning to “poured” because “wound” and its hypernym, “injury,” are not as a *physical thing* in WN. The examples are many. This paper is organized as follows. Section 2 and its subsections discuss the concept of *physical-thing* and a few of its main subconcepts. Section 3 and subsections explain a few of the subconcepts of *abstraction*, and section 4 gives our conclusions.

2 Physical-Thing

The concept of *physical-thing* corresponds to the WordNet 1.6 (henceforth WN) concept of *entity1*. Most subconcepts of *entity1* are physical things. Those few concepts which are not, such as the synset *variable1* have been extracted from *entity1*. The concept of *physical-thing* is not the same as the concept of *physical-object(object1)* in WN. *Physical-objects* are countable while *physical-thing* includes concepts which are not countable such as the concept of *substance*, and concepts which are not physical objects such as the concepts of *physical-process* and *natural-phenomenon*. The latter two are tangled to *process* and *phenomenon*, respectively. The major subconcepts of *physical-thing* that have undergone some reclassification as a result of our analysis are listed next. (We have used the star (*) and indentation to indicate the subconcepts of a given concept. Besides, we have used the arrow to indicate that a concept is also tangled to another concept. If a WN synset corresponding to our concept exists, it is listed in parentheses next to the concept. We have used the expression concept *a* goes to concept *b* in WN, in order to mean that concept *b* is a hypernym, or superconcept, of concept *a*.)

Physical-Thing

- * *physical-object (object1)*
- * *location (location1)*
- * *substance (substance1)*
- * *physical-group*
- * *physical-process -> process*
- * *natural-phenomenon -> phenomenon*

2.1 Physical-Object

Physical-object has everything in *object1* except *substance1* and *location1*, which have become subconcepts of *physical-thing*. These are the major subconcepts of *physical-object* that have undergone some reclassification.

Physical-Object

- * *physical-part (part7)*
- * *animate (life-form1)*
- * *artifact (artifact1)*

The concept of *part7*, which in our modified WN ontology (henceforth referred as “our ontology”) has been called *physical-part*, has two subconcepts *plant-part*, which in WN goes just to *entity1*, and *animal-body-part (body-part1)* which in WN goes to *part7*. In our ontology, *plant-part* and *animal-body-part* have been tangled to the concept *animate (life-form1)* in WN). Thus, we have:

physical-part(part7)

- * *plant-part(plant-part1) -> animate*
- * *animal-body-part(body-part1) -> animate*

The concept of *animate (life-form1)* has undergone few additions, one being *body-cell (cell2)* which in WN goes directly to *entity1*.

2.2 Artifact (Artifact1)

This concept has not undergone much change. However, many of the hyponyms of *structure1*, a hyponym of *artifact1*, have been tangled to *location* because most of its subconcepts (*hospital*, *building*, *area*, etc.) are used as locations. They fill the roles *to-loc* or *from-loc* of change of location verbs. More importantly, some of the hyponyms of *structure1* have also been tangled to *organization* because they are used as agents. Most of the subconcepts of *building1*, which is a subconcept of *structure1*, are also used as agents. Some of these concepts are: *tavern*, *library*, *hotel*, *restaurant*, This was discovered by failing to interpret sentences such as “The restaurant hired a new chef,” and similar ones.

2.3 Location (Location1)

Location1 is directly a subconcept of *physical-object* (object1) in WN. In our ontology, it is a subconcept of *physical-thing*. It seems that the concept *location* is not as much a *physical-object* as the concept, say, *pencil*. One finds the sentences “Peter threw/kicked the pencil” acceptable, but not “Peter threw/kicked Europe” unless one is using them in a figurative sense. That sense is what the distinction between *physical-object* and *physical-thing* tries to grasp. These comments apply strongly to *substance* because this concept is not a countable entity. Some subconcepts of *location* in WordNet have been tangled to *organization* because they are used as such. For instance, the sentence “France invaded Italy during the Napoleonic wars” and many other similar sentences could not be interpreted because “France” was just as a *location* in WordNet. Below are some of these concepts:

```
location
* district ((district1)(territory2))
* state-or-province (state2)
* country (country1) (state3)
* continent (continent1)
* residential-district
  (residential-district1)
```

State3 contains some few concepts such as *reich*, *carthage*, *holy roman empire*. Some subconcepts of *workplace1*, which in WN go to *location*, have been also tangled to *organization*. Some of these are: *farm* and its subconcepts, as well as *fishery*, *brokerage house* and a few others.

2.4 Physical-Group, Physical-Process, Natural-Phenomenon

WN distinguishes three senses of “group.” The first sense of “group,” *group1*, is a unique class containing many concepts. The problem with this is that *group1* needs to be linked to the hierarchy, and one needs to decide if *group1* must be made a subconcept of *abstraction* or of *physical-thing*. It seems obvious that the concept *group* is an abstraction, meaning a collection of abstract or physical things. However, many subconcepts of *group1* or of some of their subconcepts are collections of physical things, e.g., “fleet,” “flora,” “fauna,” “masses,” etc. which are all subconcepts of *group1* in WN. In the sentence “The hurricane pushed the fleet into the rocks,” “push” is used in its physical sense: an inanimate cause causing a change

of location of physical things, namely ships. Thus, we have created the concept *physical-group* that contains as subconcepts all those concepts under *group1* which are collections of physical things.

In WN, an important immediate subconcept of *group1* is *social-group1*, which contains many subconcepts. Because social groups are frequently used as agents, in our ontology *social-group* has become a subconcept of *human-agent*, which includes individual humans and social groups. The concepts of “people,” “citizenry,” “multitude,” and others have become subconcepts of *social-group*. Another subconcept of *group1*, *animal-group1*, has become a subconcept of *animal*. *Animal-group1* contains such concepts as “pride,” “flock,” “swarm,” “herd,” etc. which are used as referring to the members of the group rather than to the group itself.

3 Abstraction

Next we discuss the following subconcepts of *abstraction* (*abstraction6*), namely: *possession2*, which is not a subconcept of *abstraction6* in WordNet, but a unique class. We also discuss the following concepts: *communication* and *space*, which are subconcepts of *abstraction6* in WN.

3.1 Possession (Possession2)

Possession2 (anything owned or possessed) is a unique class in WN, however in our ontology is a subconcept of *abstraction* (*abstraction6*). A major subconcept of *possession* that is not classified as a subconcept of *possession* in WN is *debt-instrument1*. In WN, *debt-instrument1* is a subconcept of *document3*. In our ontology, it is both a subconcept of *written-communication1* and *possession2*. *Debt-instrument1* contains many subconcepts such as *junk bond*, *note receivable*, etc. Another subconcept of *document3* which has also become a subconcept of *possession* is *letter of credit*.

One of the hyponyms of *possession2*, *territory2*, *dominion*, *territorial dominion*, *province*, *mandate*, *colony*, has been extracted from *possession2* and made a subconcept of *location*. Another subconcept of *possession2*, *real-property1*, which contains such concepts as *hacienda*, *plantation*, etc. has been also extracted and made a subconcept of *location*. Some concepts of *possession2* have been tangled to *physical-thing* and *possession*. The major ones are: *property1*, *belongings*, *holding*, *material possession* which include such concepts as *personal effects*, *public property* and others. Besides, *currency1* (“the metal or paper medium of exchange that is presently used”) and some of the senses of “treasure” have been also tangled to *physical-thing*. The main point to emphasize is that most of the concepts that have remained as subconcepts of *possession* express an abstract relation of ownership, debt, value, liability, etc., although some subconcepts have been tangled to *physical-thing*.

3.2 Communication

The major restructuring in the category *relation* (*relation1*) has been the subconcept of *communication*. This is the final hierarchy:

```

communication
* act-of-communicating
    (communication1)
* something-communicated
    (communication2)
* written-communication -> physical-thing
    (written_communication1)
* print-media (print-media1)

```

In WN, *communication1* goes to *act2*, *human action*, *human activity* and *communication2* goes to *social-relation1*, which goes to *relation1*. Our analysis for these concepts is similar to the ones we have been just discussing, namely creating the concept *communication* to which we have not mapped any WN synset, and making *communication1* and *communication2* subconcepts of *communication*. A major concept under *communication2* is that of *written-communication*. In WN, this concept is a subconcept of *communication2*. In our ontology, *written-communication* is also tangled to *physical-thing*. The interpreter was failing to interpret many sentences such as “He burned the prescription/letter ...” because “prescription,” “letter” were not subconcepts of *physical-thing*.

We have also made *print-media1*, which includes *newspaper* and its subconcepts (a total of 20 concepts), a subconcept of *written-communication*. In WN, *print-media1* is a subconcept of *artifact*. We have also mentioned that *debt-instrument* has become a subconcept of *written-communication* and *possession*.

3.3 Space

The first three senses of “space” in WN have undergone some reorganization. The first sense, *space1*, has no subconcepts, and has *abstraction6* as its immediate superconcept. *Space2*, *topological-space1* is mathematical space and has a few mathematical subconcepts. The immediate super-concepts of *space2* are: *set2* (an abstract collection of numbers or symbols) \Rightarrow *abstraction6*. *Space3* (“an empty area usually bounded in some way between things”) has many subconcepts such as *crack*, *rip*, *hole*, *crevice*, *fault*, ... The superconcepts of *space3* are *amorphous-shape1* \Rightarrow *shape2* \Rightarrow *attribute2* \Rightarrow *abstraction6*. Our reorganization is:

```

space (space1)
* mathematical-space (space2)
* empty-area (space3) -> location.
* outer-space (space5) -> location

```

The other senses of “space” in WN remain as they are. We have made *mathematical-space* (*space2*) and *empty-area*(*space3*) subconcepts of *space* (*space1*). More importantly, we have tangled *space3* to *location*, because *space3* and its subconcepts are used most times as *location*. Note that *location* is a *physical-thing*, and we need a *physical-thing* as the selectional restriction of *change-of-location* and *cause-to-change-location* predicates. In fact, if *space3* were just a subconcept of *abstraction*, the interpreter would not be able to assign meaning to the PPs (“in a crevice,” “in the space,” “into the space”) in the sentences: “The fish frequently hides in a crevice,” “Pleural effusion is an accumulation of excessive amounts of liquid in the space between the two parts of the pleural membrane,” “Peridural anesthesia

is caused by injecting the anesthetic into the space just outside the covering of the spinal cord.”

In WN, *space5* (outer-space) is a subconcept of *location* while in our ontology is also a subconcept of *space*. Basically, our representation is capturing the duality of the concept *space* as an *abstraction* and as a *location*. Most times, however, “space,” is used as a *location* in ordinary language, e.g., “Some neutron stars, called pulsars, give off beams of radiation into space.”

4 Conclusions

We have explained some reorganizations and changes to the WN noun ontology. These changes have been pointed out by a semantic interpretation algorithm which is based on predicates linked to the WN noun ontology. Space limitations have prevented us from discussing other important concepts in the WN upper-ontology (See CS-TR-01-01 with the same title.). These changes are very much within the principles that have been guiding Wordnet, and can be easily integrated into the Wordnet ontology. As our testing of the predicates continues, we expect to make additional changes although we do not think that they will be major ones.

References

1. Bateman, J. A. and Kasper, R. T. and Moore, J. D. and Whitney, R. A.: A General Organization of Knowledge for Natural Language Processing: the PENMAN upper model. (1990).
2. Fellbaum C.: English Verbs as a Semantic Net. In *WordNet: An electronic Lexical Database and some of its applications*. Fellbaum, C. (ed.), MIT Press, (1998).
3. Gomez, F: An Algorithm for Aspects of Semantic Interpretation Using an Enhanced WordNet. In the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics NAACL-2001. pp. 87–94 (2001).
4. Buitelaar, P.: CoreLex: systematic polysemy and underspecification. PhD thesis. Dept. of Computer Science, Brandeis University (1998).
5. Mahesh, K. and Niremburg, S.: A situated Ontology for Practical NLP. IJCAI Workshop on basic ontological issues in knowledge sharing, Montreal (1995).
6. Miller, G.A: Nouns in WordNet. In *WordNet: An electronic Lexical Database and some of its applications*. Fellbaum, C. (ed.), MIT Press, (1998).