Semantic Interpretation and the Upper-Level Ontology of WordNet

Fernando Gomez

School of Computer Science, University of Central Florida
Orlando, Florida 32816, USA; gomez@cs.ucf.edu

ABSTRACT

This paper deals with the ontology for natural language understanding. In particular, the paper focuses on the WordNet noun ontology, proposes some additions, changes and reorganizations, and explains the rationale supporting them. These changes have been pointed out by a semantic interpreter driven by over 3000 verb predicates that use WordNet noun ontology and WordNet verb classes. The selectional restrictions in the semantic roles of the predicates are WordNet ontological categories. The main criterion for changing an ontological category has been failing to interpret some sentences. A detailed analysis of the proposed changes is presented.[1]

KEY TERMS

lexical ontology, WordNet, natural language understanding, cognitive science

1. INTRODUCTION

Two of the most difficult issues in defining an ontology are (a) how to decide on the categories that make up the ontology and (b) how to determine the inclusion of a concept as a member of a given ontological category. For instance, should the category *phenomenon* belong to the ontology and if so, why? Alternatively, should the concept *country* be both a member of the ontological category *social-group* and *location* and if so why? Without appealing to some general algorithm that can test these criteria, it is hard to make these decisions.

In the Cyc® ontology (Lenat & Guha, 1989), the category *Tangibleobject*, "anything that

---

[1] This paper is an updated version of the technical report CS-TR-01-01 *Grounding the Ontology on the Semantic Interpretation Algorithm*. A shorter version of that technical report appeared with the same title in the *Second International WordNet Conference*, Brno, Czech Republic, Jan-2004, pp. 34–35. e-mail: gomez@cs.ucf.edu

does have mass-energy and has no intangible aspects", is a subconcept of *Process*. It makes sense to view rocks as processes that elapse in time with a clear beginning and an end. The problem with this idea is that every physical object can be considered as a process because they are temporally based. From an intuitive point of view, however, one can sense a difference between a rock and an earthquake. An earthquake is not a physical object that stands there, can be touched, etc. An earthquake consists of a set of events and subevents having a strong temporal content. The view that conflates processes and physical objects does not correspond to the ontology underlying the semantics of natural language because language does recognize the semantic distinction between processes and physical things in many ways. For instance, it makes sense to say "She kicked/picked up the rock," but not "She kicked/picked up the earthquake," unless it is being said in a figurative sense. Coalescing this distinction would make the ontology useless for natural language understanding. This view does not mean that there are necessary and sufficient criteria for membership of an entity to an ontological category. In this aspect of cognition, there is ontological relativity (Quine, 1969) and a degree of indeterminacy, not much different from the indeterminacy of translation discussed by (Quine, 1960).

In this paper, we explain some of the reorganizations and changes to the WordNet 1.6 noun ontology (Miller, 1998), which has become an integral part of an algorithm for semantic interpretation (Gomez, 2001). The algorithm solves the following semantic interpretation problems: determination of the meaning of the verb, identification of semantic roles and adjuncts, and attachments of post-verbal prepositional phrases (PPs). The algorithm is based on verb predicates (Gomez, 1998), or verbal concepts, for WordNet 1.6 verb classes (Fellbaum, 1998). The semantic roles of the predicates have been linked to the noun ontology and to grammatical relations. As of this writing, we have defined over 3000 predicates and mapped 95% of WordNet 1.6 verb classes into predicates. The definitions of new predicates have been followed by thoroughly testing them using the algorithm.

In contrast to other ontologies for natural language (Bateman et al., 1990, Bateman, 1995, Mahesh & Niremburg, 1995), or to efforts to extract a concise set of ontological categories from WordNet (Buitelaar, 1998), the principles guiding our changes have been the selectional restrictions in the semantic roles of the 3000 predicates. For testing, we have used the electronic encyclopedia *The World Book Encyclopedia* (World Book, Inc., Chicago, Illinois, USA, 1987). This encyclopedia with its 12 volumes has become our corpus. As a result of this extensive testing, the ontology for nouns has undergone additions, reorganization, and restructuring, which we explain in this paper.[2] Failing to interpret certain constituents of the sentence has been the clue for redefining several ontological categories. For instance in one of our tests, the interpreter

---

[2] My homepage, *www.cs.ucf.edu/~gomez* contains two zip files containing the interpretation of over 1000 sentences that can be downloaded. Another set of 1000 sentences will be made available shortly.

(Gomez, 2001) failed to interpret the sentences "Europe recovered from the war" and "Denmark gave Norway to Sweden but kept Greenland and other Norwegian colonies" because the words "Europe," "Denmark", and "Sweden" in the WordNet 1.6 ontology are categorized only as *location*. In fact, all continents and nations are categorized only as *location* in WordNet 1.6. As a result, the interpreter failed to interpret all those frequent sentences in which countries and nations are used as *organization*, a subconcept of *social-group*. The concept *written-communication*, which includes many words in WordNet 1.6, is categorized as *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 WordNet 1.6. The interpreter failed to interpret the sentence "Dimes, quarters, half dollars, and dollars are made of three layers of metal," because the hypernym (superconcept) of "dime," "quarters," "dollar" is *coin1* and the hypernym of *coin1* is *possession2*, which is not a hyponym (subconcept) of *physical-thing* in WordNet 1.6, but of *abstraction*. Coins are not made of abstraction, however, but rather of something physical like metal. The examples are many and are discussed in detail throughout the paper.

It is critical to get the upper-level ontology correct, not only for resolving ordinary verb polysemy and semantic roles (Gomez, 2004) but also for finding out the metaphoric senses of the verb. In most cases, metaphoric senses are extensions of the normal senses of the verb and intrinsically linked to them. If we want to get the sense of "kill" as *terminate, end* not only in the sentence "The Congress killed the bill" but also in so many others such as "She killed the project/the plan/the proposal/poverty...," we have to rely on general distinct ontological categories in the selectional restrictions of the verb predicate. Many metaphoric verb senses are created when we shift from the ontological categories in the selectional restrictions for ordinary verb senses to very different ones, e.g., a shift from *physical-thing* to *abstraction*, *action*, or *event*. Thus, it is crucial to recognize when this shift has occurred. In the sentence, "Roofs are killed by hail," "kill" is not used in its normal sense of *cause to die*, or in its sense of *terminate, end* but rather in the sense of destroying a physical thing, and this is recognized because the upper-level ontology of "roof" is an *artifact*, which in turn is a *physical-object*. If we want to get the sense of "burn" as *waste, blow, squander,* not only as in "He has money to burn," but also as in "He burned her inheritance/fortune", it is essential that "inheritance" and "fortune" be categorized in the ontology as *possession* and *possession* as a subconcept of *abstraction*. The design of the ontology must deal not only with conventional metaphors, which many dictionaries treat as cases of polysemy, but also with those metaphors not covered in dictionaries that are freely created by users of the language. For instance, we would like the system to make sense of "She burned her vacation time," as "she used up, wasted her vacation time". The goal is that the

ontology and the semantic interpreter in combination would be able to go beyond enumerative lexical semantics and infer word senses.

Our analysis of the WordNet 1.6 ontology has produced six primitive categories, or six distinct hierarchies, which are as follows: *physical-thing, abstraction (abstraction6) state-r (state4), action (act2), process (process2), and event (event1)*. The terms in parentheses are the corresponding WordNet 1.6 ontological categories. The only super-category of these is that of *thing*, an empty category that encompasses everything. The categories that have undergone additions and modifications are explained in detail in the sub-sequent sections. The major changes have been to the categories of *physical-thing* and *abstraction*, as explained in Secs. 3 and 4, respectively. These two sections constitute the main bulk of this paper. In Sec. 5, we discuss the concepts of *act2 state4, event1, phenomenon1, and process2*, and in Sec. 6 we provide our conclusions.

2. SOME NOTES ON THE TERMS USED

Throughout the remainder of the paper, the term WN refers to WordNet 1.6, and the expression "our ontology" stands for "our modified WordNet 1.6 ontology". Some notes on WordNet 2.0 are provided in the conclusions. When we refer to concepts or senses of a word, we italicized the word (e.g. *pencil* is a subconcept of *writing-implement.*). When we refer to the word, we quote it (e.g "pencil" is a monosemous word in WN.). We have used the expression *a* is a subconcept of *b* to indicate that *a* is a hyponym of *b*, and the expression *a* is a superconcept of *b* to indicate that *a* is a hypernym of *b*. We have also used the notation $a \Rightarrow b$ to indicate that *a* is a hyponym of *b* as it is standard in WN. In the hierarchies depicted in Figs. 1, 2, and 3, we have used the star (*) and indentation to indicate the subconcepts of a given concept. In addition, we have used the arrow (->) to indicate that a concept in a hierarchy has another superconcept that does not belong to that hierarchy. In the text, we have used the expression "concept *a* has been tangled to concept *b*" to refer to this. For instance in the hierarchy depicted in Fig. 1, *injury-1* is listed as a subconcept of *physical-thing*, but it also has an arrow to *physiological-state* indicating that it is also a subconcept of *physiological-state*. Yet, *physiological-state* is not a subconcept of *physical-thing* but of *state-r* (*state4*), which is a separate hierarchy. If a WN synset corresponding to our concept exists, then it is listed in parentheses next to the concept. Thus, "*physical-thing (*entity1)" means that all words that, in WN, have *entity1* as a hypernym have *physical-thing* as a superconcept in our hierarchy, unless we have extracted a hyponym from the *entity1* hierarchy, and moved it to a different hierarchy that does not have *entity1* as a superconcept.

3. PHYSICAL-THING

The major subconcepts of *physical-thing* that have undergone reorganization are depicted in Fig. 1. The concept of *physical-thing* corresponds to the WN concept of *entity1*. Most subconcepts of *entity1* are physical things. Those few concepts that are not, such as the synset *variable1* have been extracted from *entity1*. The first subconcept of *physical-thing* is that of *injury-1*, which corresponds to *injury1*. In WN, *injury1* is a subconcept of *ill-health1* ⇒ physiological-state1 ... ⇒ *state4*. In our modified ontology, it has also been tangled to *physical-thing*. *Injury1* includes many words such as "wound," "sting," "bump," etc. The semantic interpreter failed to interpret many sentences in which these words are used in a *physical-thing* sense, rather than in a *state4* sense. For example, "Bandage is any material used to wrap or cover a wound", "The insect's bite leaves a small, pus-filled, itchy bump that is easily infected," etc.

The next concept *movement-3* (*movement3* in WN) has only *event1* as its superconcept in WN. However, in our ontology it has also been tangled to *physical-thing*. *Movement-3* has many subconcepts such as *avalanche1* ("a slide of masses of snow"), *tide1* ("the periodic rise and fall of the sea level"), *wave1, moving ridge*, *sonic-boom1*, etc. Typical sentences that the interpreter failed to interpret were "The tides bore the flatboats to the river", in which the *inanimate-cause* (the inanimate agent carrying the flatboats to the river) is realized by a *physical-thing*. Two other examples among many other sentences are "The famous tidal wave of the Petitcodiac River rushes past Moncton twice a day" and "The incoming tide sweeps up the narrowing channel, causing a bore that reverses the flow of the River Severnm," taken from *The World Book Encyclopedia* (World Book, Inc., Chicago, Illinois, USA, 1987).

Physical-Object

A very important addition to the ontology has been the concept of *physical-object* (*object1*) as distinct from *physical-thing*. *Physical-object* includes every subconcept (hyponym) of the WN concept *object1* except *substance1* and *location1*, which have become subconcepts of *physical-thing* in our ontology. It also has several important subconcepts (shown in the hierarchy) that in WN do not have *object1* as their hypernym. Entities under *physical-object* are countable while *physical-thing* includes concepts that are not countable such as the concept of *substance*, and concepts that are not physical objects such as the concepts of *physical-process*, *natural-phenomenon*, *physical-group,* and *location*. This is relevant for determining the selectional restrictions of many predicates. For instance, suppose that we want to define selectional restrictions for the role *theme* of the verb "break" when it means destroying the physical integrity of something. If we choose *physical-thing* as its selectional restriction, then it would be too general, resulting in assigning meaning to sentences that are semantically unacceptable such as "The sugar/paper/school ... broke."

```
  Physical-thing (entity1)

* Injury-1 (injury1)        -> physiological-state

* Movement-3 (movement3)  -> event

* Physical-object (object1 minus substance1 and location1)
  * card-1 (card1)
  * sheet-2 (sheet2)
  * list-1 (list1)          -> abstraction
  * written-communication -> communication
  * physical-part (part7)
     * plant-part(plant-part1)      -> animate
     * animal-body-part(body-part1) -> animate
  * animate (life-form1)
    * body-1 (body1)
    * cell-2 (cell2)
  * artifact (artifact1)
    * structure1
      * establishment-4(establishment4) -> organization
      * building1
        * hotel1            -> organization
        * restaurant1       -> organization

* Location (location1)
  * district (district1 territory2)   -> social-group
  * continent (continent1)            -> social-group

* Substance (substance1)
  * chemical (chemical1)
    * chemical-element (chemical_element1)
       * chemical-agent (agent1 agent3)
    * chemical-agent (agent1 agent3)
    * drug (drug1)           -> artifact
  * fluid (fluid1 fluid2)
  * coating-1 (coating1)     -> artifact

* Physical-group (group1)
 * social-group (social_group1 people1)     -> human-agent
 * biological-group (biological_group1)
   * animal-group (animal_group1)            -> animal (animal1)
   * plant-group (vegetation1)               -> plant (plant2)

* Physical-process (process2)               -> process

* Natural-phenomenon (natural_phenomenon1) -> phenomenon
```

Fig. 1: Hierarchy for main upper-level concepts of physical-thing

In WN, the concept *location* (*location1*) has *object1* as its immediate hypernym. In our modified ontology, it is a subconcept of *physical-thing*. Many of the subconcepts of *location* cannot be used in the selectional restrictions of many verb predicates that clearly accept subconcepts of *physical-object*. For instance, the following sentences are not semantically acceptable "Tom threw/kicked/ ... Asia/Texas/Paris," unless the verbs are used in some metaphoric sense. If one replaces "Asia" with "book" or "cup" or "lamp" etc., however, the sentences are perfectly acceptable. Prototypical instances of concepts under *physical-object* are *pencil*, *lamp*, *chair*, etc. The category of *physical-object* as distinct from *physical-thing* has become very useful in defining the selectional restrictions of many predicates.

Card-1 (*card1* in WN "card - one of a set of small pieces of stiff paper marked in various ways and used for playing games or for telling fortunes"), and *sheet2* ("piece of paper, sheet of paper - used for writing or printing") have *substance1* in WN as hypernym. This obviously caused some problems because *physical-object* and *substance* fill different selectional restrictions for many predicates. In WN, the concept *list1* has *abstraction* as its super-concept. In our ontology, it is also a subconcept of *physical-object*. Many words under *list1* such as "calendar," "agenda" etc. have a physical-object sense that is not captured in WN. The same comments apply to the concept *written-communication1*, which has only *abstraction* as its upper-level hypernym in WN. In our ontology, *written-communication1* is also a sub-concept of *physical-object*. The interpreter failed to interpret many sentences in which words in the class *written-communication* were used in their physical sense, rather than in their abstract sense; e.g., "She burned/hid/... the letter/message/..."

The concept of *part7* has *entity1* as its hypernym in the WN ontology. In our ontology, it has *physical-object* as its superconcept. This concept has two major subconcepts *plant-part*, (a subconcept of entity1 in WN), and *animal-body-part* (*body-part1*) (a subconcept of *part7* in WN). In our ontology, *plant-part* and *animal-body-part* are also subconcepts of *animate (life-form1)*, as well as *body-1* (*body1*) "body, organic structure ...") and *cell-2* (*cell2*), which in WN are subconcepts of *entity1*, but not of *life-form1*. This restructuring is necessary to be able to interpret correctly many sentences in which the words under these hyponyms are used in their *animate* sense, such as "The cells/lymphocytes/leukocyte ... die", in which "die" is used in its normal sense of cease to live. Note that this reorganization will also allow the semantic interpreter to select the correct sense of polysemous words under these hyponyms. For instance, it will select the correct sense of "cell," namely "cell-2," among the six senses of "cell" in the sentence "The cells die."

Artifact (artifact1)

The most important change in the concept *artifact* is the tangling of some of the subconcepts of *structure1*, a subconcept of *artifact1* in WN, to *organization* ⇒ *social-group* in our ontology. The concept *establishment4* ("a public or private structure, business or governmental or educational") has been tangled to *organization* ⇒ *social-group* because most of the words "store", "nursing home" and many others under it are frequently used as organizations. Additionally, many of the subconcepts of *building1*, a subconcept of *structure1* in WN, have been tangled to *organization* ⇒ *social-group*. Some of the words under these WordNet concepts are "theater", "library", "tavern", "hotel", "restaurant", etc. This tangling was discovered by failing to interpret sentences such as "The hotel hired a new manager," and similar ones, in which "hotel" is used as a social agent and not as location.

Location (location1)

By mapping *disctrict1* and *territory1* to *district* and making *district* a subconcept of both *location* and *social-group*, many words—such as "country", "province", etc.—that have only *location* as their hypernym in WN now have also *social-group*. In WN, words that refer to continents, e.g., "Europe", "Africa", etc. have only *land4*, *dry land* ⇒ *object1* ⇒ *entity1* as their hypernyms. Nevertheless, these words are used as social groups. Before these changes, the interpreter failed to interpret such simple sentences as "Italy conquered Ethiopia in, 1936" and many others in which "country" is used as a *social-group*. The same applies to many words under *state2*, which includes the American states among others. Again, this error was revealed by the semantic interpreter failing to interpret sentences such as "Texas elected/impeached/... the new governor" because "Texas" was not as a *social-group* in the WN ontology.

Extensive feedback from the semantic interpreter was required before which concepts to reorganize in the upper-level ontology of WN became clear to us. For instance, in an earlier version of our ontology, we made *country* (country1 in WN) a subconcept of *social group* without realizing that by making *district1* a subconcept of *social-group* we did not have to modify *country1* because it is a subconcept of *district1* in WN. Further experimentation with the semantic interpreter on many randomly selected sentences indicated to us which concepts and at what level of the WN hierarchy to modify or reorganize.

Substance (substance1)

In WN, *substance1* is a subconcept of *object1* ⇒ *entity1.* Because most concepts under *substance1* are mass nouns, we have made *substance* a subconcept of *physical-thing* rather than a

subconcept of *physical-object*. We have made *chemical-element1* and *chemical-agent1* subconcepts of *chemical1*, a subconcept of *substance1* in WN. The concepts *agent1* and *agent3* have been made subconcepts of *chemical-agent*. In WN, *agent1* and *agent3* have *causal-agent* as their superconcept not *chemical1*. *Agent1* contains such words as "germicide", "antifungal", "disinfectant" and *agent3* such words as "viricide" and "antacid".

In WN, *drug1* is not a subconcept of *substance1* or *chemical1*, but rather a subconcept of *artifact1*. We have made *drug1* a subconcept of *chemical1* and tangled it to *artifact1*. In WN, *coating1* is only a subconcept of *artifact1*, whereas in our ontology *coating1* is also a subconcept of *substance*. This subconcept includes such words as "paint", "coat", etc. Finally, we have created the concept *fluid*, which does not correspond to any WN noun sense, and made *fluid1* ("a substance that is fluid at room temperature") and *fluid2* ("a continuous amorphous substance") subconcepts of *fluid*.

Physical-Group

WN distinguishes three senses of "group". The second synset of group, *group2, radical1* refers to "two or more atoms bound together as a single unit and forming part of a molecule" and contains such concepts as *allyl group*, *benzyl group*, etc. The third sense of group, *group3*, *mathematical-group1*, is the mathematical concept of group. There are no problems with these two senses of "group," which are technical terms. The problem is with the first sense of "group", *group1, grouping1*, which is defined in WN as "any number of entities (members) considered as a unit." This concept of *group1* is a unique category (does not have any superconcepts) and contains many words. Many of the words under *group1* are collections of physical things, e.g., "fleet", "flora", "pride" (a group of lions), "fauna", "masses", etc., and they are used in that sense in most sentences. Thus, the semantic interpreter was failing to interpret sentences in which these words are used in their physical sense, as in "The storm scattered the fleet." Thus, we have created the concept *physical-group* that contains as subconcepts all those concepts under *group1* that are collections of physical things, and we have made it a subconcept of *physical-thing*.

In WN, an important immediate subconcept of *group1* is *social-group1*, which includes many words. 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. Another major concept of *physical-group* is *biological-group* which contains *animal-group* ("pride", "flock", "swarm", "herd", ...) and *plant-group* ("brush", "grove", "forest", ...). *Animal-group* has been tangled to *animal*, and *plant-group* to *plant (plant2)*.

Physical-Process and Natural-Phenomenon

In WN, *process2* has *phenomenon1*, a unique category in WN, as its hypernym. Most words under *process2* are used in their physical thing. Thus, we have created the concept of *physical-process* as a subconcept of both *physical-thing* and *process*. The reasons for doing so are identical to those for the creation of the category *physical-group* that we discussed in the previous section. Likewise, we have also made the concept *natural-phenomenon1*, a subconcept of *phenomenon1* in WN, a subconcept of both *physical-thing* and *phenomenon*. Section 5 contains a more detailed description of *process2* and *phenomenon1*.

4. ABSTRACTION

Next, we discuss some of the main subconcepts of *abstraction* (*abstraction6*), namely: *psychological-feature1* and *possession2*, which are not subconcepts of *abstraction6* in WordNet, but rather unique classes. We also discuss the following concepts: *property (property2, property4), relation (relation1) time*, and *space*, which are subconcepts of *abstraction6* in WN. The hierarchy for *abstraction* is depicted in Fig. 2.

Psychological Feature

Psychological-feature1 is a unique class in WN. In our ontology, it is a subconcept of *abstraction*. Figure 2 depicts the major subconcepts of *psychological-feature* that have undergone reorganization. In WN, *psychological-state1* is a subconcept of *condition1* ⇒ *state4*, but not of *psychological-feature1*. In our ontology, *psychological-state1* is a subconcept of *psychological-feature* and *state-r* (*state4*). Some of the words under *psychological-state1* are "nervousness", "anxiety", "agitation", "depression", etc. *Cognitive-State (cognitive-state1)* is only a subconcept of *condition1* ⇒ *state4* in WN. In our ontology, *cognitive-state1* is also a subconcept of *psychological-feature*. *Cognitive-state1* includes such words as "curiosity", "consciousness", "doubt", "inwardness", etc. The observation here is that the ontology of natural language does not make those fine distinctions between cognitive-state and psychological-state on one hand and psychological-feature on the other. We need the generic category of *psychological feature* that includes every word sense that expresses feelings, thoughts, traits, etc. Without this generic concept, the interpreter was failing to identify the semantic roles of many predicates that need *psychological feature* in their selectional restrictions.

Trait1 ("a distinguishing feature of one's personal nature") is a subconcept of *attribute2* ⇒ *abstraction6* in WN. Yet, most of the words under *trait1* are used in the sense of *psychological-*

feature. In fact, some words under *Trait1*, e.g. "bravery", are listed in WN with two senses, one of them being *psychological-feature1* and the other one *abstraction6*.

Nevertheless, most words under *Trait1* do not have the sense of *psychological-feature* in WN. For instance, "heroism", "valor", "courage", "cowardice", "audacity", etc. have only the sense of *abstraction6*.

Personality1 and *attractiveness2* are subconcepts of *attribute2* ⇒ *abstraction6* in WN. In our ontology, both concepts are also subconcepts of *personal-trait* (*trait1*), making all of them subconcepts of *psychological-feature* also. Some of the words under *attractiveness2* are "charisma", "charm", "appeal", etc.

Possession (possession2)

Possession2 ("anything owned or possessed") is a unique class in WN, however in our ontology it is a subconcept of *abstraction* (*abstraction6*). The WN concept of *possession2* caused the semantic interpreter to fail to interpret many sentences. The main reason for this was that possession2 included many words that are used in its physical sense rather than in its abstract sense. In addition, many word senses under *possession2* had to be tangled to abstraction and physical-object because they can be used in any of the two senses. Consider the following three sentences "The characters were also scratched on coins," "Stamping machines are used to emboss coins," and "Lydia was one of the first countries to cast coins." The semantic interpreter failed to interpret the constituents in which the word "coin" appears because "coin" in WN has only *possession2* as its hypernym, and "coin" is used in these sentences in its physical sense not in its abstract sense. However, the sentence "She burned the money" can be ambiguous between she physically burned the money and she squandered the money. Most concepts that have remained as subconcepts of *possession* express an abstract relation of ownership, debt, value, liability, etc., and some subconcepts have been tangled to *physical-thing*.

Some of the hyponyms of *possession2* that have been extracted from *possession2* and made subconcepts of *location* are *territory2, dominion, territorial dominion, province, mandate, colony*, and *real-property1* (which contains such concepts as *hacienda, plantation*, etc.) Some subconcepts of *property1, belongings, holding, material possession* ⇒ *possession2*, have been tangled to *physical-object*. In addition, *currency1* ("the metal or paper medium of exchange that is presently used") and some of the senses of "treasure" have been tangled to *physical-object* as well.

```
Abstraction (abstraction6)

 * Psychological-feature
  * psychological-state (psychological-state1)  -> state-r (state4)
  * cognitive-state (cognitive-state1)          -> state-r (state4)
  * personal-trait (trait1)

 * Possession (possession2)
  * debt-instrument (debt_instrument1)    -> written-communication
  * letter-of-credit (letter-of-credit1)  -> written-communication
  * property-belongings (property1)       -> physical-object
  * currency (currency1)                  -> physical-object

 * Property (property2 property4)
  * quantity (quantity2)
    * mathematical-quantity(quantity3)
    * measure
     * measure-quantum (measure3)
     * measurement (measure1)
     * magnitude-relation (magnitude-relation1)
  * spatial-property (spatial-property1)
    * shape
     * shape-contour(shape1)
     * shape-form(shape2)

 * Relation
  * social-relation
    * communication
     * act-of-communicating (communication1)
     * something-communicated (communication2)
     * speech-act (speech_act1)                      -> act2
     * written-communication (written_communication1) -> physical-object
     * print-media (print-media1)                    -> physical-object

 * Time
  * time-continuum(time5)
  * time-unit (time_unit1)      -> measure
  * time-period (time-period1)
    * indefinite-period (time2)
    * time-interval (time-interval1)
  * clock-time (clock-time1)

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

Fig. 2: Hierarchy for main upper-level concepts of abstraction

A major subconcept of *possession* that it is not classified as a subconcept of *possession2* 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* includes many words such as "junk bond", "note receivable", etc. Another subconcept of *document3*, which has also become a subconcept of *possession,* is *letter of credit*. The reorganization of all these concepts was pointed out by the semantic interpreter.

Property (property2 and property4)

Property2 is a subconcept of *abstraction6* in WN, as well as in our ontology. *Property4* is a subconcept of *psychological-feature1* in WN. We have coalesced this concept into *property*. *Property4* has about 40 word senses. One major concept of *property* that has undergone reorganization and additions is the concept of *quantity2*. The hierarchy for *quantity* (*quantity2*) is depicted in Fig. 2. The superconcepts of *quantity* are *amount2* ⇒ *magnitude1* ⇒ *property2* ⇒ *attribute2* ⇒ *abstraction6*. The subconcept of *mathematical-quantity (quantity3)* is a subconcept of *psychological-feature* in WN. *Mathematical-quantity (quantity3)* has such subconcepts as *quotient, binomial, constant* etc. Our concept of *measure*, to which we have not mapped any WN synset, has the following subconcepts: *measure-quantum (measure3)* which in WN has *abstraction6* as its immediate and only superconcept; *measurement (measure1)* that has *magnitude1* as its superconcept in WN; and *magnitude-relation (magnitude-relation1)*, which is a subconcept of *relation1* ⇒ *abstraction6* in WN. All these concepts have become subconcepts of our concept *measure*, which is the concept being used in the selectional restrictions of the predicates.

Spatial-Property

Another concept that has undergone some restructuring is that of *spatial-property*. The analysis of *shape1* and *shape2* is similar to the one for *measure*. In WN, *shape1* ("any spatial attribute especially as defined by outline") is directly a subconcept of *spatial-property*. *Shape2* ("the spatial arrangement of something as distinct from its substance"), however, is a subconcept of *attribute2* and this of *abstraction6*. *Shape2* does not even have *spatial-property1* as one of its superconcepts. *Shape1* includes such words as "concavity", "angularity", and "convexity", whereas *shape2* has all geometric figures, lines, angles, amorphous shapes etc. In our ontology, we have created the concept *shape*, which does not correspond to any WN synset, and made *shape1* and *shape2* subconcepts of *shape*. *Shape* is a generic ontological category that includes all senses of "shape" as spatial attributes, whether they are defined by outline or by arrangement.

Communication

A major restructuring in the category *relation* (*relation1*) has been the subconcept of *communication*. In WN, *communication1* is a subconcept of *act2, human action, human activity* and *communication2* is a subconcept of *social-relation1* ⇒ *relation1*. Our analysis for the concept of *communication* is similar to the ones we have provided for *shape* and *measure*, namely to create the concept of *communication* to which we have not mapped any WN synset, and making *communication1* and *communication2* subconcepts of *communication*. In this way, the category *communication* unifies both senses of "communication". Yet, if the most specific senses of *communication*, namely *communication1* and *communication2*, are needed in some selectional restrictions of some semantic roles, then they can be used. We have also made *speech-act1* a subconcept of communication. In WN, *speech-act1* is a subconcept of *act2*, which is a unique category in WN. Although many words in WN under *speech-act* have senses that have also *communication* as their hypernym, some of them have only *speech-act1*.

A major concept under *communication2* is that of *written-communication1*. In WN, this concept is a subconcept of *communication2*. In our ontology, *written-communication* is also a subconcept of *physical-object* as discussed in Sec. 3 under *physical-object*. The interpreter was failing to interpret many sentences such as "He burned the letter ..." or "She put the prescription on the table" because "letter" and "prescription" do not have *physical-thing* as one of its senses in WN. We have also made *print-media1* a subconcept of *written-communication*. In WN, *print-media1* is only a subconcept of *artifact*.

Time

The concept of *time* is essential in semantic interpretation in order to recognize temporal adjuncts. Time NPs, e.g., "She studied 2 hours" can be easily confused with arguments of the verb predicate if the ontology of *time* is not adequate. We have approached this concept in an identical manner to the concepts of *measure*, *shape,* and *communication* explained previously. We have created the concept *time*, to which we have not mapped any WN synsets, and made some of the WN synsets of "time" subconcepts of this concept. In this way, our concept of *time* unifies all diverse temporal senses of the word "time" in WN.

There are 10 senses of "time" in WN. *Time1* has no subconcepts, and is a subconcept of *happening* ⇒ *event1*, e.g., "This time he succeeded." This sense of "time" is not a time concept. *Time2* (an indefinite period) is a subconcept of *time-period1* ⇒ *measure* ⇒ *abstraction6. Time2* has about 30 subconcepts; this is clearly a time concept. *Time3* ("a period of time considered as a resource under your control ...") also is a subconcept of *time-period1* ⇒ *measure3* ⇒

abstraction6. This is also a time concept but it has no subconcepts. *Time4* ("a suitable moment; `it is time to go") has also *time-period1* as its superconcept. *Time4* has only two subconcepts: *occasion* and *high time*. *Time5* ("the continuum of experience in which events pass from the future through the present to the past") has *abstraction6* as its immediate superconcept. *Time5* contains such concepts as *past times* and its subconcepts, *cosmic time* and its subconcepts (*equinox, solstice, sidereal time*). This sense of "time" is clearly a time concept, and we have subcategorized it as a subconcept of *time* (see hierarchy). *Time6* (*clock-time1*) is a subconcept of *psychological-feature* in WN. *Clock-time* contains such concepts as *mealtime, dinnertime, bedtime, sunset, etc.* We have subcategorized it as a subconcept of *time*. *Time7*, *fourth-dimension1* is a subconcept of *measure, magnitude, attribute, abstraction* in WN; it has no subconcepts. *Time8* ("a person's experience on a particular occasion ...") is a subconcept of *event* in WN; it has no subconcepts. *Time9, meter4* ("rhythm as given by division into parts of equal time") is a subconcept of *abstraction6* in WN. None of these three senses of "time" is a time concept. *Time10, prison-term1* is a subconcept of *time-period1*; it has one subconcept, namely *hard time*. *Time-unit1* is a subconcept of *measure3* ⇒ *abstraction6* in WN and contains such concepts as *hour, minute, second*, etc. We have subcategorized this concept as a subconcept of *time* and have tangled it to *measure*. *Time-period1*, which contains many subconcepts, is a subconcept of *measure3* ⇒ *abstraction6* in WN. *Time-period1* has also become a subconcept of *time* in our modified ontology. Another concept that has been treated in the same way as *time-unit1* is *time-interval1* that is a subconcept of *measure3* ⇒ *abstraction6* in WN, and contains such concepts as *latent period, reaction time*, etc. Thus, we have the hierarchy for time depicted in Fig. 2. The concept *time* does not correspond to any WordNet synset, and the others are its subconcepts.

Space

The first three senses of "space" in WN have undergone some reorganization, while the other senses of "space" remain as they are. The first sense, *space1* ("the unlimited 3-dimensional expanse in which everything is located") has no subconcepts and has *abstraction6* as its immediate superconcept. The second sense of "space", *space2, topological-space1*, is the concept of mathematical space and has five mathematical subconcepts. *Space2* is a subconcept of *set2* ("an abstract collection of numbers or symbols") ⇒ *abstraction6*. *Space3* ("an empty area usually bounded in some way between things") includes many words, such as "crack", "rip", "hole", "crevice", "fault," ... *Space3* is a subconcept of *amorphous-shape1* ⇒ *shape2* ⇒ *attribute2* ⇒ *abstraction6*. Our reorganization is depicted in Fig. 2. We have created the concept of *space* and mapped the WN synset *space1* to it. Then, 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*, then the interpreter would not be able to assign meaning to the PPs ("in a crevice", "into the space") in the sentences: "The fish frequently hides in a crevice," "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*, whereas in our ontology it is also a subconcept of *space1*. Our representation is capturing the duality of the concept of space as an *abstraction* and as a *location*. Most times, however, the word "space", is used as a *location* in ordinary language, e.g., "Piles are placed in two rows, and the space between them is filled with rocks."

5. ACT2, STATE4, EVENT1, PHENOMENON1, AND PROCESS2

Next, we explain the concepts of *state4, event1, phenomenon1,* and *process2.* All these are unique categories in WN as well as in our modified ontology. The category of *act2, human action, human activity* ("something that people do or cause to happen") has not undergone any major change, so it is not mentioned in this discussion.

State-r (state4)

The category of *state4* ("the way something is with respect to its main attributes; the current state of knowledge; his state of health; in a weak financial state") is a unique category in WN as well as in our modified ontology. This concept has undergone little modification. However, we have added some concepts to it. *Temperature1* ("the degree of hotness or coldness of a body or environment ...") is a subconcept of *physical-property1* ⇒ *property2* ⇒ *... abstraction6*. However, *temperature1* and its subconcepts are classified in our ontology as subconcepts of *state-r* (*state4*) (see Fig. 3).

Temperature1 is a changing state of physical things as it is revealed by the following sentences: "The temperature plunged to -40 degrees F" and "The body temperature rises very high," both of which the semantic interpreter failed to interpret.

Animation3 is a subconcept of *attribute2* ⇒ *abstraction6* in WN; it includes such words as "vigor", "vitality", "exuberance", "enthusiasm", "resilience", "vivacity", ... Some of these words have senses that are clearly subconcepts of *state-r*(*state4*), and WN does include the sense *state4*

in many of them, but misses some. Some important additions have taken place to one subconcept of *state4*, namely *ill-health1*, which is a subconcept of *state4* in WN, as well as in our ontology. WN, however, fails to include the concept *symptom1* and its many subconcepts (e.g., *abscess1, inflammation1, atrophy1* etc.) as a subconcept of *ill-health1*.

Many words are included under *symptom1*. In WN, *symptom1* is a subconcept of *evidence1* ⇒ *information3* ⇒ *cognition1* ⇒ *psychological-feature1*. Another subconcept of *ill-health* in our ontology, but which is not a subconcept of *ill-health* in WN, is that of *affliction3*. In WN, *affliction3* is a subconcept of *trouble4* ⇒ *happening1* ⇒ *event1*. Some of its subconcepts are *embolism2* and its subconcepts, *thrombosis1* and its subconcepts, *seizure1* and its subconcepts, and others. Finally, in WN *disorder1* is a subconcept of *state4*, but not of *ill-health*. Words included under *disorder1* are "epilepsy", "insomnia", "anorexia", and others.

```
Action (act2)
  (no major changes)

State-r (state4)
 * temperature (temperature1)
 * animation-3 (animation3)
 * ill-health (ill-health1)
    * symtom-1 (symptom1)
    * affliction-3 (affliction3)
    * disorder-1 (disorder1)

Event (event1)
 * social-event (social_event1)
   * party-gathering (party4)        -> social-group
 * meeting (meeting1 meeting2)   -> social-group
 * meal-1 (meal1)                -> food
 * movement-3 (movement3)        -> physical-thing

Process (process2)
 * physical-process                 -> physical-thing
  * natural-process (natural-process1)
 * cognitive-process (process3)  -> psychological-feature
  * unconscious-process (process5)
  * psychoanalytic-process (psychoanalytic_process1)
 * phenomenon (phenomenon1)
  * natural-phenomenon(natural_phenomenon1) -> physical-thing
    * periodic-event (periodic_event1)
```

Fig. 3: Hierarchy for the concepts of action state-r, event, process, phenomenon

Event (event1)

Which role the category *event1* plays in the ontology is unclear because *event1* ("something that happens at a given place and time") does not seem to be part of the ontology of concepts regardless of how they are sententially used. There is much redundancy between the concept of *event1* and those of *state4*, *act2*, *process2*, and *phenomenon1*. In fact, any of the latter concepts can be used as an event in language. For instance, "fire" is listed in WN as *event1* and as a *process2* among other senses. In the sentence, "He disproved the theory that air, earth, water, and fire were the basic elements of all matter", "fire" is used as a *process2*. In "Many people died in the fire as a result of the earthquake in San Francisco", however, "fire" is used as an *event*, a particular fire that occurred in a specific time slot. Every deverbal nominalization can be used as an *event* or an *action*, e.g., "The rescue took place during the night" vs "Rescues are dangerous operations". It seems that the category *event1* could be eliminated in favor of *process2* or *act2* in most cases and of *state4* in some cases. In the selectional restrictions of the predicates, *event1* is used in those cases in which *state4, act2*, and *process4* are used. WN lists some of the well known complex actions, phenomena, and states as events. For instance, "coronation" is listed only as a *social-event*, although it should be also listed as *act2*.

WN has two major subconcepts of *event*: *happening1* and *social-event1*. *Party4* that in WN is a subconcept of *social-group1*, but not of *event1*, has also become a subconcept of *event* (*event1*) in our ontology. *Party4* includes such words as "dance", "banquet", "ball", "open house", etc. Because *social-group1* is an *agent* in our ontology, the interpreter was interpreting incorrectly some adjuncts, (e.g., "At the ball, Romeo meets Juliet") in which these words occur. For the same reasons, *meeting1* and *meeting2*, which also have *social-group1* as a superconcept in WN, have been coalesced into the concept *meeting* that has become a subconcept of *event1* in our ontology. The concept *meal1* ("the food served and eaten at one time") is a subconcept of *food1* ⇒ *substance1* ⇒ *object1* ⇒ *entity1* in WN. In our ontology, it is also a subconcept of *event1*. This synset includes such words as "lunch", "dinner", "supper", etc., which are frequently used as events or activities. For instance, "At lunch, we will talk about your proposal." Finally, we have made *movement3*, which is only a subconcept of *event1* in WN, a subconcept of *physical-thing* as discussed in Sec. 3.

Phenomenon1 and Process2

Phenomenon1 is a unique category in WN and is defined as "any state or process known through the senses rather than by intuition or reasoning". *Process2*, however, is a hyponym of *phenomenon1* in WN. We have extracted *process2* from *phenomenon1* and made it an

independent ontological category. Moreover, we have made *phenomenon1* a subconcept of *process2*. For that reason, *process2* is a more inclusive ontological category than that of *phenomenon1*, whereas *phenomenon1* is preferred by the more specific selectional restrictions of semantic roles.

We have created the subconcept *physical-process*, which has become a subconcept of *physical-thing*. *Natural-process1* has become a subconcept of *physical-process*. The concept of *cognitive-process* has been tangled also to *psychological-feature*. Finally, we have made *periodic-event* (*periodic-event1*) a subconcept of *natural-phenomenon*. In WN, *periodic-event1* is only a subconcept of *event1*, and includes such concepts as *tide1* ("the periodic rise and fall of the sea level ..."), *diastole,* and others.

6. CONCLUSIONS

We have explained some reorganizations and changes to the WordNet noun ontology. These changes have been pointed out by a semantic interpretation algorithm that is based on verb predicates linked to the WordNet noun ontology. As our testing of the predicates continues, we expect to make some additional changes, although we have gathered considerable evidence (Gomez, 2001; 2004) now indicating that the changes will not be major.

The changes to the Wordnet ontology discussed in this paper constitute important improvements to that ontology, but they are within the methodological principles guiding WordNet. Upon receiving an earlier version of this paper (CS-TR-01-01 Grounding the Ontology on the Semantic Interpretation Algorithm), George Miller informed us (e-mail communication May-1-2001) that he will be introducing some of the changes that we propose in this paper in the new release of WordNet 2.0, but that others could not be introduced because of implementation problems. Indeed, WordNet 2.0 has incorporated some of the changes discussed in this paper. In Lenat et al. (1995, 48), Miller says, "Our assumption, which may prove too optimistic, has been that the lexical component of language can be studied more or less independently of other components." Our work has shown that WordNet upper-level ontology is essentially adequate for semantic interpretation and that it provides the foundation for ontology-based approaches to natural language understanding.

REFERENCES

Bateman, J.A., Kasper, R.T., Moore, J. and Whitney, R.A. 1990. A general organization of knowledge for natural language processing: the penman upper model, Technical report, UCS/Information Science Institute, Marina del Rey, California, USA.

Bateman, J.A. 1995. On the relationship between ontology construction and natural language: a sociosemiotic view, International Journal of Human-Computer Studies 43, 929–944.

Buitelaar, P. 1998. CoreLex: systematic polysemy and underspecification. Ph.D. Dissertation, Brandeis University, Department of Computer Science.

Fellbaum, C. 1998. A semantic network of English verbs. In: WordNet: An Electronic Lexical Database, edited by Fellbaum, C., Cambridge, Massachusetts, USA, MIT Press, 1998, 69–104.

Gomez, F. 1998. Linking wordnet verb classes to semantic interpretation. In: COLING-ACL Workshop on Usage of WordNet in NLP Systems, 58–64.

Gomez, F. 2001. An Algorithm for aspects of semantic interpretation using an enhanced wordnet. Proceedings of the 2nd North American Meeting of the North American Association for Computation Linguistics, NAACL-2001, 87–94.

Gomez, F. 2004. Building verb predicates: A computational view. Proceedings of the 42nd Meeting of the Association for Computational Linguistics, ACL-04, 351–358.

Lenat, D. and Guha, R. 1989. Building Large Knowledge-Based Systems: represen-tation and inference in CYC project. Reading, Massachusetts, USA, Addison-Wesley Publishing Company.

Lenat, D., Miller, G. and Yokoi, T. 1995. CYC, WordNet and EDR: Critiques and responses. Communications of the ACM 38, 45–48.

Mahesh, K. and Niremburg, S. 1995. A situated ontology for practical nlp. In IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing.

Miller, G. 1998. Nouns in WordNet. In: WordNet: An Electronic Lexical Database, edited by Fellbaum, C., Cambridge, Massachusetts, USA, MIT Press, 23–46.

Quine, V. 1960. Word and Object. Cambridge, Massachusetts, USA, MIT Press.

Quine, V. 1969. Ontological Relativity and Other Essays, New York, NY, USA, Columbia Univ. Press.