Verb Predicates, Meaning and Multilinguality

Fernando Gomez
Dept. of Computer Science
School of EECS
University of Central Florida, Orlando, FL 32816
gomez@cs.ucf.edu
Working Paper CS-TR-03-02
March-26-03

Abstract

This paper summarizes some of the research on verb predicates that has taken place at the University of Central Florida during the last few years. It also describes how this research may be related to the MEANING project. The point is made that the definition of predicates linked to a general ontology, such as that of WordNet, and to grammatical relations can solve the crucial issues of verb and prepositional meaning and thematic roles. Also, it is suggested that a broad ontology of predicates, which would be language-free as far as ontological categories is concerned, could be extended relatively easy to other languages, resulting into ontologies of predicates similar to the different "wordnets" that have emerged from WordNet. This ontology of predicates could facilitate multilingual tasks such as information retrieval and machine translation, as well as deeper monolingual natural language processing tasks, such as understanding.

1 Introduction

For the last few years, we have been working in defining predicates for English verbs. This work has proceeded by mapping WordNet verb classes (Fellbaum, 1998) into generic predicates (Gomez, 1998). The selectional restrictions of the predicates are semantic categories in WordNet noun ontology (Miller, 1998). The thematic roles are also linked to the syntactic relations that realize them. An algorithm that uses the predicate definitions has been designed and implemented (Gomez, 2001a). The algorithm solves the following meaning relations: verb meaning, thematic roles, prepositional attachment and meaning of prepositions, and adjuncts. Noun senses are solved not only in those instances in which the selectional restrictions of the predicate constrain the noun sense, e.g., the meaning of "buck" as an male animal in "He killed a buck," but also in many instances in which the selectional restriction *prefers* (Wilks, 1975) a noun sense over other, e.g., the meaning of "fortune" as possession in "She inherited a fortune."

The goal of this research is to define predicates for every English verb and place them into a hierarchy in which thematic roles and inferences can be inherited. Because the selectional restrictions of the predicates are ontological categories in a well established and multilingual ontology, namely WordNet, we have some ground to believe that the predicates which we are defining for English can be mapped into other languages, in a similar way in which WordNet has been mapped into other languages. This report that summarizes our research in these issues is organized as follows. Section 2 summarizes the definition of generic predicates for WordNet verb classes. Section 3 explains recent developments and the construction of predicates for individual verbs. Section 4 summarizes the rationale behind some changes to the WordNet ontology, and Section 5 explains the relation of this work to the MEANING project (Rigau et al., 2002).

2 Defining Predicates for WordNet Verb Classes

Predicates have been defined for WordNet 1.6 (henceforth, WN) verb classes (Fellbaum, 1998). We have taken a top-down approach by defining first generic abstract predicates subsuming semantically and syntactically a large class of verbs. WN verb classes have been mapped into these generic predicates. Some of this mapping has required us to define new classes and to reclassify and/or redefine some

WN classes and subclasses (Gomez, 1998). The predicates form a hierarchy in which thematic roles and inferences are inherited by subpredicates from their superpredicates. One major consequence that derives from mapping verb classes into abstract semantic predicates is coalescing several WN senses into a predicate, which reduces the systemic polysemy in some WN senses.

The *differentia* between a predicate and its subpredicate are given by one or more of the following: a) different selectional restrictions for the thematic roles, b) different syntactic relations for the thematic roles, and c) different sets of inferences associated with the subpredicates. For instance, the predicate *walk* (walk1 in WN) (not to be confused with the verb "walk") inherits the *agent*, *to-loc*, *from-loc*, and also the *distance* (e.g. "She walks two miles.") from the generic predicate *change-of-location-by-animate-being*, but the *instrument* is very specific to it. One does not want to take "in the ship" as the *instrument* of the predicate *walk* in "He walks in his ship." Compare to "She went to the island in her ship." In our ontology, the predicate *fill* (fill1 in WN) is a subpredicate of *cause-to-change-location*, which expresses a cause of change of location of something other than the agent; although the agent may have also changed location. The *goal* of *fill* is realized by the object, e.g. "She filled the tank" and the *theme* by a [with NP] prepositional phrase , which clearly differs from how those roles are realized in the generic predicate *cause-to-change-location*. Subpredicates inherit all the thematic roles not listed in their definitions from their parent predicates.

The semantic interpretation algorithm (Gomez, 2001a) is activated by the parser after parsing a clause. The parser does not resolve structural ambiguity, which is delayed until semantic interpretation. Our mapping of WN verb synsets to predicates provides a list containing the predicates for the verb of the clause. The goals of the algorithm are to select one predicate from that list, attach PPs and identify thematic roles and adjuncts. All these tasks are simultaneously achieved. For each syntactic relation (SR) in the clause (starting with the NP complements) and for every predicate in the list of predicates, the algorithm checks if the predicate *explains* the SR. A predicate *explains* an SR if there is a thematic role in the predicate realized by the SR and the selectional restrictions of the thematic role subsume the ontological category of the head noun of the syntactic relation. This process is repeated for each SR in the clause and each predicate in the list of predicates. Then, the predicate that explains the most SRs is selected as the meaning of the verb. The thematic roles of the predicate have been identified as a result of this process. In case of ties, the predicate having the greatest number of thematic roles realized is selected. Every syntactic relation that has not been linked to a thematic role must be an adjunct or an NP modifier. The entries for adjuncts are stored in the root node *action* and are inherited by all predicates. Adjuncts are identified after the meaning of the verb has been determined because they do not belong to argument structure of the predicate.

3 Defining Predicates for Individual Verbs

See (Gomez, 2003) for a detailed discussion of the issues discussed in this section. Our subsequent research has shown that our initial working hypothesis that the definition of a generic predicate for a WN class would also apply to most verb forms under that class does not hold in many cases. The reasons for this is that many verbs under a class realize their thematic roles by different selectional restrictions and syntactic relations. Even, in many instances, verb forms in the same synset list differ semantically and syntactically with their selectional restrictions and syntactic relations. The concept of *troponymy* (Fellbaum, 1998), which has been the criterion used for grouping verb classes in WN differs from the criterion that we have been applying in defining our predicate classes, namely the inheritance of thematic roles in a hierarchy of predicates. Notwithstanding these problems, WN verb classes have facilitated immensely our work and we owe much to them. Moreover, our insistence in working in a top-down fashion by first building generic predicates for WN verb classes has paid off because our mapping of 98% of WN verb classes into predicates has facilitated our subsequent task of defining predicates for individual verbs. The

definition of predicates for an individual verb is similar to the way in which the senses for a verb are listed in a dictionary. A dictionary provides the senses for each verb (and some dictionaries also give the prepositions for that sense if applicable). What dictionaries do not provide is a) a predicate for that sense and a hierarchy in which to insert it, b) the thematic roles for that predicate, and c) the syntactic relations and the ontological categories for the selectional restrictions. Section 5 gives an example of how one constructs predicates for an individual verb, the Spanish verb "dirigir." In (Gomez, 2003), we describe in detailed how these definitions take place. We also explain a software environment for the semiautomatic definition of predicates, consisting of a parser, the semantic interpreter that uses the predicates to interpret the sentence, a corpus (*The World Book Encyclopedia*, World Book, Inc. Chicago) divided into different sections for defining, refining and testing the predicate definitions, a skimmer that searches for sentences in the corpus containing the verb for which predicates are being defined, and a mechanism for dragging in sentences from the corpus. In that paper the reader also can find a description of the upper-level ontology of predicates, and the results of our testing of the predicates that we have defined for individual verbs. We have defined 3017 predicates and mapped 98% of WN verb classes into predicates.

4 The WordNet Noun Ontology

See (Gomez, 2001b) for a detailed discussion of the issues discussed here. The noun ontology has undergone changes and reorganizations as dictated by the semantic interpretation algorithm. These changes have been very important because without them the algorithm would fail to interpret many sentences. Hence, the criteria for changing or reorganizing an ontological category has been dictated by the failure of interpreting a sentence. For instance, upon looking into the reasons why the interpreter failed to interpret the sentence "She placed the letter on the table," we found that the interpreter could not identify the *theme* of the predicate *place-put*. This was due to the fact that "letter" was not in the ontology as a *physical-thing*, and the selectional restriction for the *theme* of *put-place* requires a *physical-thing*. Of course, this is not only true of "letter" but also of all concepts that have *written-communication1* as a superconcept (e.g, "They hid the charter in a tree."), because in WN *written-communication1* is a subconcept of *abstraction*, and not of *physical-thing*. The solution has been to tangle *written-communication* to *physical-thing* and *abstraction*.

We illustrate the importance of these changes for semantic interpretation with two modifications to the WN noun ontology. The reader interested in these issues may want to look into the aforementioned paper. First of all, we have the distinction between *physical-object* and *physical-thing*. The concept of *physical-thing* corresponds to the WN concept of *entity1*. Most subconcepts of *entiy1* 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-object* subsumes objects that 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*. This distinction is very relevant in many many definitions of predicates. For instance, the concept *Location1* is directly a subconcept of *physical-object* (object1) in WN. However, 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/the cup" acceptable, but not "Peter threw/kicked Europe/the river" unless one is using them in a figurative sense. There are many predicates besides those expressed by the verbs "throw" and "kick" that prefer the concept of *physical-object* over the concept of *physical-thing*. The distinction is relevant to select not only the sense of verb, but also the sense of the noun. For instance, in the sentence "She broke the table," everybody would think of "table" as a piece of furniture, (senses 2 and 3) and not as a location "mesa, table" or any of the other senses for "table" listed by WN.

Consider the WN concept of space3 ("an empty area (usually bounded in some way between things)"), which is a subconcept of *ab-*

straction (abstraction6) in WN, and includes the concepts of *crevice*, *crack* and others. However, instances of this concept are used many times in the selectional restrictions of the thematic roles of the predicates *change-of-location* and *cause-to-change-location* and others, which need a *physical-thing* in the selectional restriction. For instance, "The female lays one white egg in a crevice," "The fish hides in a crevice." The solution is to tangle *space3* to *location*, which is a *physical-thing* and to *abstraction*. See (Gomez, 2001b) for a detailed discussion of the concept *space* in WN and its changes. The distinction between *physical-thing* and *abstraction* has been critical in our modified upper-level WN ontology, since many figurative senses depend on this distinction.

5 Relation to the Meaning Project

The relation of our work to the MEANING project (Rigau et al., 2002) is rather strong. Our work deals with the issues of verb meaning, thematic relations, prepositional meaning, adjuncts and to a lesser extent with noun senses. We plan to quantify in a near future how much our selectional restrictions contribute to determining the meaning of the noun. It would be interesting to find out how much they can improve the percentages achieved by some of the noun senses methods (see (Ide and Veronis, 1998; Rigau et al., 1997; Yarowsky, 1995)). Another aspect in which our work relates to the MEANING project is the important role that Wordnet plays within the MEANING research group (Vossen, 1998; Daude et al., 2000) and (Agirre and Martinez, 2002; Agirre et al., 2001). Another aspect in which we find intersection is the work reported in (McCarthy et al., 2001) about the automatic acquisition of selectional preferences and the disambiguation of noun and verb senses. Finally, our work can benefit greatly from one of the goals of the MEANING project, the construction of corpora across different languages. This is something that may impact the entire NLP community. But, perhaps the most fascinating aspect of our work is how the verb predicates that we have been building for English can be used to construct verb predicates for some of the languages targeted by the MEANING Project.

In the remainder of this section, we make some *suggestions* of how a mapping from the English verb predicates onto verb predicates for other languages could take place. Because of our knowledge of Spanish, we will consider Spanish as the target language. Our *very preliminary* look into these issues shows that there is a great similarity between English and Spanish regarding the hierarchy of predicates and their selectional restrictions. The differences, of course, lie in the grammatical relations. Let us consider a large class of predicates, say *change-of-location-by-animate*, whose prototypical verbs are "walk,"("andar") "run" ("correr"), "go" ("ir,") etc. For these predicates, not only the selectional restrictions are identical, but also the grammatical relations: both classes of verbs are used mostly intransitively, subcategorize distance NPs (e.g. "She walked many miles") and almost the same class of prepositions. Consider the generic predicate **cause-to-change-location**. The primary event expressed by this predicate is a cause of change of location of something other than the agent, although the agent may have also changed location. In "Mary carried the boxes to the house " the agent also has also moved, but the primary event is the fact that Mary caused a change of location of the boxes. In a sentence such as "The moon circles the earth," "the moon" is the *theme* and the *agent* or *inanimate-cause* is unknown. Within this predicate and its subpredicates, the similarities between both languages are very high. If one considers one of its subpredicate *fill*. Not only the selectional restrictions are the same "He filled the tank," ("Llenó el depósito"), but also the grammatical relations. In both languages, the *goal* is realized by a direct object, and the *theme*, by the preposition "with" in English and "de" in Spanish. This is a very minor difference. When one considers *change-of-state* predicates, the similarities are striking. One may find some differences, which are very minor when one considers the relevance and scope of the mapping we have in mind. For instance, in English the *theme* can be expressed by a subject in "The door broke" while in Spanish needs also the marker "se," e.g., "La puerta se rompió." The same observations apply to other generic predicates such as *think*, *decide*, *judge* and their subpredicates. The selectional restrictions are identical, the dif-

ferences are in the grammatical relations which, in our preliminary opinion, would be very easy (that is, it will take little time) to map. The mapping can be automated and then corrected by editing it.

At the individual verb level there are differences, of course. Some verbs may have more senses/predicates in one language than in another, or different predicates. For instance, if one considers a Spanish verb such as "comer," ("eat" in English) one can find that many of the predicates for "eat" also exist in Spanish, e.g., ingest food, corrode, use up, and others. The only thing needed for building those predicates for "comer" would be to copy those predicates from English, and do some minor editing. But, Spanish overloads the meaning of "comer" much more than English. We will illustrate that situation by defining some of the predicates for the Spanish verb "dirigir," which does not correspond to a single clear verb in English. This is a bottom up method that takes a Spanish-English dictionary entries as keys to provide the predicates.

The first sense that the dictionary provides for "dirigir" is to manage people, organizations. Our software already provides automatic ways to display all the predicates for a given English verb. This is as simple as (display-predicates $<$ *verb* $>$), where $<$ *verb* $>$ is any verb. The only thing one needs to do is to display the predicates for "manage" and insert under "dirigir" the predicate that means to manage people, organizations, which is *manage-people-organizations*. No syntactic changes are needed. Another sense for "dirigir" is to "address somebody." Again, one displays the predicates for "address" and brings in the predicate *address-somebody*. A syntactic change would be needed the *recipient* in Spanish is realized by a PP [a NP], while in English is by an object. The selectional restrictions are the same. Another sense is "to conduct a musical group." The predicates for "conduct" already include that predicate, which is identical in Spanish. Another sense for "dirigir" is "to go towards." The predicates for "go" includes that predicate. The syntactic changes needed would be to add the prepositions "a" and "hacia," and to ignore the personal pronouns in Spanish. We have listed in figure 1 the predicates for "dirigir" that we have discussed. The predicate is in Spanish, while its superpredicates and selectional restrictions are in English. The syntactic relations are: subj (subject), obj (object), prep (prepositional phrase) followed in parentheses by the prepositions that may introduce it, pre-verbal-np, a NP preceding the verb. The super-predicates for the predicate *dirigir-personas-organizations* are *manage* $\Rightarrow$ *exert-control-over* $\Rightarrow$ *decide* and the ones for *dirigirse-a-alguien* are *speak-to-somebody* $\Rightarrow$ *speak-talk* $\Rightarrow$ *trans-infor* $\Rightarrow$ *communicate* $\Rightarrow$ *interact* $\Rightarrow$ *action*. Every one of them applies also to Spanish. Note that this method of mapping does not only produce verb predicates for Spanish from English verb predicates, but also, with the same effort, a translation of English verb predicates onto Spanish verb predicates.

```
[dirigir-personas-organizations
(is-a(manage))
(agent(human-agent) (subj))
(theme(human-agent organization) (obj))]

[dirigirse-a-alguien
(is-a(speak-to-somebody))
(agent(human-agent animal)(subj))
(ignore(personal-pronoun)  (pre-verbal-np))
(recipient(human-agent animal)
        ((prep a)))]

[dirigir-un-grupo-musical
(is-a(perform-play))
(agent(human)(subj))
(theme(musical_organization1) (obj))]

[dirigirse-a-un-lugar
(is-a(change-location-by-animate-being))
(agent(human animal) (subj))
(ignore(personal-pronoun) (pre-verbal-np))
(towards-loc(location physical-thing)
      ((prep a hacia)))]
```

Figure 1: Some of the Predicates for the Spanish verb "dirigir."

Our very preliminary analysis indicates that the English verb predicates can be used to semiautomatically construct Spanish verb predicates, and, as an extension, predicates for other Romance languages such as Catalan, Galician, Italian etc. without involving a great effort. We think that this is also the case for the construc-

tion of verb predicates for Germanic languages and other Indo-European languages.

6 Conclusions

We have summarized some of our work on English verb predicates, and indicated briefly how this work relates to the MEANING project. We have also described some very preliminary analysis, indicating that the verb predicates built for English could be used to construct verb predicates for other languages in a relatively simple manner. We have provided some examples of predicates for the Spanish verb "dirigir," and shown how they could be derived from the predicates for the English verbs into which "dirigir" can be translated.

References

E. Agirre and D. Martinez. 2002. Integrating selectional preferences in wordnet. In *Proc. of First International WordNet Conference*, Mysore, India.

E. Agirre, O. Ansa, D. Martinez, and E. Hovy. 2001. Enriching wordnet with topic signatures. In *Proc. of the NAACL Workshop on WordNet and other lexical resources*, Pittsburgh, USA.

J. Daude, L. Padro, and G. Rigau. 2000. Mapping wordnets using structural information. In *Proc. of 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong.

C. Fellbaum. 1998. A semantic network of english verbs. In C. Fellbaum, editor, *WordNet: An electronic Lexical Database and some of its applications*, page chapter 3. MIT Press, 1998, Cambridge, Mass.

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

F. Gomez. 2001a. An algorithm for aspects of semantic interpretation using an enhanced wordnet. In *NAACL-2001*.

F. Gomez. 2001b. Grounding the ontology on the semantic interpretation algorithm. Technical report, Dept. of Computer Science, SEECS, University of Central Florida. CS-TR-01-01, Feb. 2001.

F. Gomez. 2003. Building verb predicates: A computational view. Technical report, Dept. of Computer Science, SEECS, University of Central Florida. CS-TR-03-01, Feb. 2003.

N. Ide and J. Veronis. 1998. Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics*, 24.

D. McCarthy, J. Carrol, and J. Preiss. 2001. Disambiguating noun and verb senses using automatically acquired selectional preferences. In *Proc. of the SENSEVAL-2 Workshop at ACL-01*, Toulouse, France.

George Miller. 1998. Nouns in wordnet. In C. Fellbaum, editor, *WordNet: An electronic Lexical Database and some of its applications*, page chapter 1. MIT Press, 1998, Cambridge, Mass.

G. Rigau, J. Atserias, and E. Agirre. 1997. Combining unsupervised lexical knowledge methods for word sense disambiguation. *Computational Linguistics*, 04:237–251.

G. Rigau, B. Magnini, E. Agirre, P. Vossen, and J. Carroll. 2002. Meaning: A roadmap to knowledge technologies. Technical report, Informatika Fakultatea, Basque Country. www.lsi.upc.es/nlp/meaning.html.

P. Vossen. 1998. *EuroWordNet: A multilingual database with lexical semantics*. Kluwer Academic Publishers, Dordrecht.

Y.A. Wilks. 1975. Preference semantics. In E.L. Keenan, editor, *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge, UK.

D. Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proc. of 33th Annual Meeting of the Association for Computational Linguistics*.