A System for the Semantic Interpretation of Unrestricted Domains using WordNet

Fernando Gomez
School of Computer Science
University of Central Florida, Orlando, FL 32816
gomez@cs.ucf.edu

Carlos Segami
Dept. of Mathematics and Computer Science
Barry University, Miami Shores, Fl 33161
segami@euclid.barry.edu

In this demonstration, we show an algorithm for the semantic interpretation of unrestricted texts. The algorithm presents a solution for the following interpretation problems: determination of the meaning of the verb, identification of thematic roles and adjuncts, and attachments of prepositional phrases (PPs). An interesting aspect of the algorithm is that the solution of all these problems is interdependent. The interpretation algorithm uses WordNet (Miller *et al.* 1993) as its lexical knowledge-base. Predicates, or verbal concepts, have been defined for WordNet verb classes (Fellbaum 1993), which have been reorganized considerably following the criteria imposed by the interpretation algorithm. The WordNet ontology for nouns has also undergone some reorganization and redefinition to conform with the entries in the thematic roles of the predicates. We have taken a top-down approach that defines generic abstract predicates subsuming semantically and syntactically a large class of verbs. WordNet 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 WordNet classes and subclasses (Gomez 1998a). The predicates form a hierarchy in which thematic roles and inferences are inherited by subpredicates from their superpredicates. Two major consequences derive from anchoring verb classes in abstract semantic predicates: coalescing several WordNet senses into a predicate, which reduces the systemic polysemy in some WordNet senses, and mapping the same WordNet synset into distinct predicates. For instance, all the 5 synsets listed by WordNet for "travel": "travel1, go, move, locomote;" "travel2, journey;" "travel3, take a trip, make a trip;" "travel4, journey;" and "travel5 (undergo transportation, as in a vehicle)" are coalesced into the abstract semantic predicate *change-of-location-by-animate*. This predicate defines a class of verbs containing the most generic properties shared by all members of the class. The *differentia* between this predicate and its subpredicates are given by one or more of the following: a) specific selectional restrictions for the thematic roles, b) different syntactic realizations of the thematic roles, and c) specific sets of inferences associated with the subpredicates. For instance, the *instrument* of *drive* is always a vehicle,

while the *instrument* of *change-of-location-by-animate* can be an animate, an animate body part, etc. The *instrument* of *drive* is never realized by a subject, but the *instrument* of the generic predicate can be realized by a subject, e.g., "This bus goes to Cambridge every Wednesday." On the other hand, *migrate* differs from *change-of-location-by-animate* only by the specific inferences associated with this predicate. Subpredicates inherit all the thematic roles not listed in their definitions from their parent predicates. The entry for the predicate *change-of-location-by-animate* is listed on the next column. The *wn-map* entry means that all the five senses of "travel" and all the verb forms under them are, in principle, coalesced into the concept *change-of-location-by-animate*. These synsets group under them about 1,200 verb forms. However, subpredicates of this predicate will be recognized. Any WordNet synset that satisfies any of the *differentia* given above is mapped into a subpredicate of *change-of-location-by-animate*. In fact, we have defined 48 subpredicates of *change-of-location-by-animate*. But, if they were not recognized, then any form that is mapped by WordNet into any of the five senses of "travel" would be mapped into the predicate *change-of-location-by-animate*. Of course, if the form is also mapped by WordNet to another sense besides any of the senses of "travel," then it will be mapped into whatever predicate we identify for that sense. The generic meaning expressed by all these verbs is that of the change of location by an animate being, in which the *agent* and the *theme* are the same. The entry *agent* indicates that the agents of this predicate are entities belonging to the classes *animal* or *social-group* (a group of humans or animals). It also expresses that the agent is realized syntactically by the subject of the sentence. The *theme* is also realized by the subject. The selectional restrictions for the role *to-loc* are, in order of preference, *location* and *physical-thing*. This role is realized syntactically by a PP containing any of the prepositions listed. The different ways of expressing a path, namely *toward-loc*, *near-loc*, etc. have been collapsed here into *to-loc* for space reasons. The sign "-" preceding a selectional restriction means that any noun sense that is subsumed by the semantic category preceded by "-" does not realize that thematic role.

```
[change-of-location-by-animate(is-a (action))
 (wn-map (travel1) (travel2) (travel3)
         (travel4) (travel5))
 (agent(animal social-group) (subj))
 (theme(animal social-group) (subj))
 (to-loc
   (location physical-thing) ((prep towards
     toward around through into back-to along
     over beside above by under below throughout
     beyond past across near up down))
   (location -social-group -animal
             physical-thing)     ((prep to))
   (location) ((prep for)))
 (from-loc (location -animal -social-group
         physical-thing) ((prep from out-of))
         (location) (obj))
 (instrument
   (conveyance)(subj obj (prep with in on))
   (animal) ((prep on))
   (extremity1) ((prep on with)))
 (medium (road-or-path)((prep on)))
 (distance(distance linear-measure)
          (obj(prep for)))]
```

The *to-loc* can also be realized by the preposition "to" and the semantic category *location*, but not by *social-group* nor *animal*, and yes by *physical-thing*. This is a way of saying that it is realized by *location* and *physical-thing* excluding *social-group* and *animal* from *physical-thing*. The *from-loc* role contains two pairs of lists. The first pair expresses this role when it is realized by a PP, and the second pair when it is realized by obj. The reason for using two pairs of lists is that "from" expresses a *from-loc* role in most cases. Hence, if the object of the PP is not classified as a *location* in the ontology but just as a *physical-thing*, one can assume, quite safely, that the PP expresses a *from-loc*. However, in the case of obj, one needs to be certain that the head noun of the NP is a location; otherwise one may incorrectly identify a thematic role and, possibly, a predicate. Similar explanations apply to the remainder of the thematic roles. As of this writing, 1198 predicates have been defined for 70% of WordNet verb forms.

The semantic interpretation algorithm (Gomez 1998b) is activated by the parser after parsing a clause. The parser recognizes clausal and NP complements, relative clauses, resolves gaps resulting from questions and relativization, and handles coordination and subordination. It does not resolve structural ambiguity, which is delayed until semantic interpretation. Our mapping of WordNet 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 that has the greatest number of thematic roles realized is selected. Every syntactic relation that has not been mapped 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 adjuncts are not part of the argument structure of the predicate. Prepositions that are listed in the *action* node as *weakly* claimed by the verb may modify the verb or a NP. Heuristic rules decide in those cases (Gomez, Segami, & Hull 1997). We have integrated the semantic interpreter into SNOWY and begun to look into the task of acquiring biographic knowledge from the *The World Book Encyclopedia*.

The system has been implemented in Franz Common Lisp and runs on Sun Sparc workstations. The WordNet lexicon for nouns and verbs has been converted into a Lisp-based representation. The semantic interpreter takes 0.96 MB, the parser 0.48 MB, the WordNet lexicon in the Lisp representation 35.00 MB, the parser lexicon 4.40 MB and the verbal predicates that we have defined for the WordNet verb class 0.25 MB. In the demo, we will type some sentences in order to illustrate the organization of the lexical knowledge, the predicates and the different aspects of the algorithm. Then, we will let the audience actively participate by forming their own sentences and trying them in the system.

References

Fellbaum, C. 1993. English verbs as a semantic net. Technical report, Princeton. CSL Report 43, revised March 1993.

Gomez, F.; Segami, C.; and Hull, R. 1997. Determining prepositional attachment, prepositional meaning, verb meaning and thematic roles. *Computational Intelligence* 13(1):1–31.

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

Gomez, F. 1998b. Semantic interpretation of unrestricted domains using wordnet. Technical report, University of Central Florida, Department of Computer Science. CS-TR-98-2, Jan. 1998.

Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D.; and Miller, K. 1993. Introduction to WordNet: An on-line lexical database. Technical report, Princeton. CSL Report 43, revised March 1993.