Semantic Interpretation of Nominalizations

Richard D. Hull and Fernando Gomez
Department of Computer Science
University of Central Florida
Orlando, FL 32816
hull@cs.ucf.edu

Abstract

A computational approach to the semantic interpretation of nominalizations is described. Interpretation of nominalizations involves three tasks: deciding whether the nominalization is being used in a verbal or non-verbal sense; disambiguating the nominalized verb when a verbal sense is used; and determining the fillers of the thematic roles of the verbal concept or predicate of the nominalization. A verbal sense can be recognized by the presence of modifiers that represent the arguments of the verbal concept. It is these same modifiers which provide the semantic clues to disambiguate the nominalized verb. In the absence of explicit modifiers, heuristics are used to discriminate between verbal and non-verbal senses. A correspondence between verbs and their nominalizations is exploited so that only a small amount of additional knowledge is needed to handle the nominal form. These methods are tested in the domain of encyclopedic texts and the results are shown.

Introduction

Quirk(Quirk *et al.* 1985) defines nominalization as "a [noun phrase] which has a systematic correspondence with a clause structure," where the head nouns of nominalization phrases are related morphologically to either a verb (deverbal noun) or adjective (deadjectival noun). In this paper we focus on deverbal nominalizations, a common linguistic device used as a vehicle for describing specific events, generic actions, and *action concepts* such as evolution. Understanding sentences with nominalizations requires the ability to determine the meaning of the nominalization and to make sense of the nominalization's modifiers, namely, the other words in the noun phrase (NP) containing the nominalization and the prepositional phrases that follow it.

An important problem is distinguishing between the verbal and non-verbal senses of the nominalization, as is necessary for words like "support", "decoration", and "publication". In order to distinguish between verbal and non-verbal senses, we will use the term nominalization to refer to only those *senses* of the noun which are derived from verbs. For example, the noun "decoration" has several senses including a military badge of honor, an ornament, and the nominalization sense, which means the act or process of decorating.

There also is the equally serious problem of what to do when the nominalized verb is polysemous. This problem is quite prevalent, and while there is a large body of work discussing word sense disambiguation using knowledge-based approaches(McRoy 1992; Hirst 1992; Jacobs & Rau 1993; Voorhees 1993; Li, Szpakowicz, & Matwin 1995), a computational approach to this problem has not, to our knowledge, been specifically addressed in the literature. As an illustration of the problem, consider the verb *promote*, which has several different meanings, e.g., to promote a product, to promote a person to a higher position of authority, and to promote a cause. The nominalization, "promotion," therefore, can have different meanings as in *the promotion of Peter* or *the promotion of liberalism.*

The framework of a model of semantic interpretation that can be used to solve these problems is described in detail in (Gomez, Segami, & Hull 1997). While a partial treatment of nominalizations involving the attachment of prepositional phrases to them is presented in that work, its focus is not a complete model of nominalization interpretation and as such does not address the issues of interpreting nominalization NP modifiers and deciding between a nominalization's verbal and non-verbal senses, which will be discussed here.

The remainder of this paper is comprised of five sections. Section describes the essential aspects of the semantic interpreter of which the nominalization algorithms, explained in section , are a part. The testing of these algorithms on a collection of nominalizations is discussed in section . A comparison of this approach to other work in the literature is explained in section . Finally, section presents the authors' conclusions.

Semantic Interpreter

The parser used by our semantic interpreter leaves structural syntactic relations underspecified along the lines of D-Theory (Marcus, Hindle, & Fleck 1983)

and minimal commitment parsers (Weinberg 1993). The parser recognizes the following syntactic relations: subject, object1, object2, predicate, and prepositional phrases (PPs). Object1 is built for the first postverbal noun phrases (NPs) of transitive verbs, and object2 for the second postverbal NP of diatransitive verbs. The parser also recognizes temporal adjuncts of the verb; but, as indicated, it does not resolve structural ambiguity. PPs are left unattached in the structure built by the parser until the semantic interpreter finds their meaning and attaches them. As each syntactic relation is identified, it is passed to the semantic interpreter to make semantic sense of it.

All important semantic decisions are delayed until the meaning of the verb is determined. Determining verb meaning is done by rules that use the types of syntactic relations built by the parser and the semantic categories of the nouns in them. For instance, what determines the meaning of "drive" in (1) is the direct object of the verb and the fact that "bus" is a motor-vehicle (here "drive" means *to operate a vehicle*), and that "nail" is a device (here "drive" means *to hammer*).

(1) *Peter drove the bus/the nail.*

In addition to NP complement rules, prepositional phrase rules (PP rules) are also used to determine the meaning of the verb. In (2), the meaning of "left" is identified by two *for*-rules stored in its lexical entry: the meaning of "left" is identified as *transfer of possession* in the first case because the object of the PP is a human and as *depart* in in the second because the object of the PP is a location.

(2) *Jennifer left the orange groves for her son/for home.*

Once the meaning of the verb is established, additional knowledge is needed to interpret subsequent syntactic relations and those which have been parsed but have remained uninterpreted. This knowledge is stored in the representation of the verbal concept or predicate. Below are examples of NP complement rules and preposition rules for the verb "defend":

```
(defend
  ((object
    ((if% is-a obj location)
       (verbal-concept-is defend-physical-thing))
    ((if% is-a obj idea)
       (verbal-concept-is defend-idea)))
  (prep
    (in ((if% equal obj-of-prep court)
          (verbal-concept-is legal-defend))))
  (end-of-clause
    ((if% is-a obj championship)
       (verbal-concept-is defend-championship))
    ( t (verbal-concept-is defend)))))
```

The first set of rules selects the verbal concept based upon the NP complement. If the direct object is a subconcept or an instance of a *location* within a hierarchy of concepts, then the meaning of the verb is *defend-physical-thing*. This rule is designed to handle constructions such as *Carleton defended Quebec*. If the direct object is a subconcept or instance of an *idea* in the concept hierarchy, then the meaning of the verb is represented by the verbal concept *defend-idea*, as in *her defense of the theorem*. If the direct object does not pass either of these two constraints, then the meaning of the verb is left unspecified in hopes that later evidence will disambiguate it. The next rule is a preposition rule for PPs that follow the verb "defend" and begin with the preposition "in". If the object of such a prepositional phrase (the object of the PP is the head of its complement NP) is equal to the word "court", then the meaning of the verb is represented by the verbal concept *legal-defend*. The last set of rules, called end-of-clause rules, is used if the parser reaches the end of a clause and the meaning of the verb is still unknown. If the direct object is a subconcept or instance of a *championship* in the concept hierarchy, then the meaning of the verb is represented by the verbal concept *defend-championship*. Otherwise, the meaning of the verb is the generic *defend*. These five verbal concepts are displayed below.

```
(defend
 (is-a (action-r))
 (subj (agent (actor)))
 (obj (thing (theme))
 (prep (against (thing (against (strong))))
       (from (thing (against (strong))))))

; WordNet sense defend1,4,7
(defend-idea
 (is-a (defend))
 (obj (idea (theme))))

; WordNet sense defend2,5
(defend-physical-thing
 (is-a (defend))
 (obj (physical-thing (theme))))

; WordNet sense defend3
(defend-championship
 (is-a (defend))
 (obj (championship (theme))))

; WordNet sense defend6
(legal-defend
 (is-a (defend))
 (obj (agent (theme)))
 (prep (in (court (at-loc (strong))))))
```

The first entry in the verbal concept *defend*, *(is-a (action-r))*, places *defend* within the hierarchy of action concepts. The next entry is a restriction: if the subject of defense is subsumed by the *agent* concept in the concept hierarchy, then make it fill the *actor* role. The other entries represent restrictions on the object and prepositional phrases. Each of the subconcepts of *defend*; *defend-idea*, *defend-physical-thing*, *defend-championship*, and *legal-defend*, inherits entries from *defend*.

This fine-grained decomposition of "defend" is necessary if one wishes to make specific inferences depending on the type of defense. The structures above embody the knowledge necessary to understand clauses containing the verb "defend" (see (Gomez, Segami, & Hull 1997) for a detailed discussion of VM rules and verbal concepts). This verbal knowledge can be exploited and reused for interpretation of nominalizations, if a small amount of additional information detailing how the nominalization's modifiers relate to the syntactic

relations of the VM rules and verbal concepts is constructed. Knowledge indicating whether any thematic roles are "obligatory," that is, necessary for a verbal sense interpretation, is also stored.

The nominalization "defense" is shown below. *Defense* has one obligatory role, *theme*. Because "defense" has both verbal and non-verbal senses, a requirement is made that the *theme* must be present for a verbal sense to be chosen. A nominalization that has only verbal senses does not need an obligatory-role slot. No special mapping rules for genitives or prepositions are present because "defense" behaves like most nominalizations: genitives represent either the actor or theme of the action; and besides the preposition "of" which represents the theme of the transitive verbs and the actor of intransitive verbs, "defense" inherits the meanings of its PP modifiers from its root verb, "defend."

```
(defense (obligatory-role (theme)))
```

For the majority of nominalizations, no information over and above that of the verbal knowledge is necessary. However, there are exceptions. Genitive modifiers of the nominalization "attendance" only make sense as the actors of "attend". In this case, a slot specifying that the genitive should fill the *actor* role is needed. Another situation where additional information is needed is in the handling of certain prepositional phrases. The nominalization "control" takes PPs using the preposition "over" as the verb's object, as in, *his control over the business*, while the verb "control" does not. To handle this, a slot mapping the preposition "over" to the verb's object is added.

In addition to providing the means for disambiguating between nominal and deverbal senses of the nominalization, verbal knowledge can also be used to disambiguate the underlying verb of the nominalization when it exhibits polysemy. It is the prepositional phrase "in court" that selects the meaning *legal-defend* for "defense" in *her defense in court*, and it is the prepositional phrase "of Richmond" that selects the meaning *defend-physical-thing* in *Lee gave up the defense of Richmond.*

Interpretation Algorithms for Nominalizations

The interpretation algorithms attempt to determine the verbal concept of the nominalization and to fill its thematic roles. Determination of the verbal concept requires disambiguation of the meaning of the nominalization's root verb. This ambiguity may be resolved by examining the noun phrase in which the nominalization occurs, or as is true in many cases, disambiguation can only be accomplished by examining postnominal prepositional phrases. Once the verbal concept has been identified, surrounding nouns are then interpreted as verbal concept arguments. There are three separate interpretation algorithms: the nominalization noun phrase algorithm, the prepositional attachment and meaning determination algorithm, and the end-of-clause algorithm. We will discuss the main points and then the details of each algorithm in turn.

Nominalization Noun Phrase Algorithm

The nominalization noun phrase algorithm is triggered when the head noun[1] of some NP is determined to be a potential nominalization. This is accomplished by consulting WordNet(Miller *et al.* 1993) to see if any of the senses of the noun are hypernyms of either actions or events. Conceptually, the algorithm has two objectives:

1. To determine the verbal concept or predicate of the nominalization, and
2. To identify which thematic roles of the verbal concept, if any, each of the remaining nouns and adjectives of the NP fill.

Determining the verbal concept of the nominalization establishes its meaning within the context of the sentence. Occasionally, the nominalization has a single meaning and in those cases we can immediately determine the verbal concept. This trivial disambiguation is attempted first and works for nominalizations like "invasion" and "murder." More often, however, determining the verbal concept requires disambiguating the nominalization because the root verb of the nominalization is polysemous.

In order to disambiguate polysemous nominalizations, the algorithm uses the root verb to select VM rules. In addition, mapping rules and heuristics are needed to handle the fact that nominalizations do not take bare NPs; the verb's syntactic subject and object reappear as genitival, adjectival, or prepositional modifiers. This algorithm addresses genitives, possessive pronouns, single prenominal nouns, i.e., pairs of the form *noun nom*, and adjectives which fill thematic roles. The prepositional attachment and meaning determination algorithm described later handles prepositional modifiers.

Determining the Verbal Concept

The VM rules of the root verb do not include any for handling genitives, pronouns, or noun/adjective modifiers. Therefore, if these rules are to be reused, some way of selecting the appropriate ones is needed. The central problem associated with disambiguation of the nominalization within the NP then becomes *Which VM rules should be fired?* Consider the case where the NP is of the form: (*genitive nominalization*). The genitive may correspond semantically to the verb's subject or object, as is illustrated in the examples below.

(3) *Lincoln's election; The representatives elected Lincoln.*

[1]The algorithm does not currently handle nominalizations in positions other than the head.

(4) *Metternich's resignation; Metternich resigned.*

In (3), the genitive corresponds to the object position, while in (4), it corresponds to the subject position. The verb *resign* is intransitive, except for colloquial expressions such as "resigned his office", and therefore, genitive modifiers of the nominalization *resignation* correspond to the verb's subject. This idea forms the first rule selection heuristic. Passive nominals behave differently; their genitive modifiers correspond to the verbal object. A mapping rule selects the verb's object rules when the nominalization is passive. If the nominalization's verb is neither intransitive nor is the nominalization passive, then the genitive could correspond to either the subject or object. Consequently, the selection of VM rules is postponed, in hopes that following prepositional phrases will disambiguate the nominalization.

If the nominalization is modified by either a noun or adjective, it is not possible to determine exactly which disambiguation rules must be fired. Any ordering of the rules is guaranteed to be wrong in a large percentage of cases. In addition, it would be unproductive to try all of the rules in hopes that only the appropriate one would be triggered. Therefore, disambiguation is postponed until more evidence is available in the form of roles filled by prepositional phrases.

Filling Thematic Roles

Once the verbal concept of the nominalization has been determined, the next step is to determine which of the other constituents of the NP fill thematic roles of the verbal concept and what those roles are. A syntactic relation is said to fill a thematic role if the concept it represents in the concept hierarchy passes the selectional restrictions associated with that role. For example, *humans* elect *humans* to *institutions* as *social-roles*. The hierarchy of concepts is consulted to determine if the argument of the syntactic relation under consideration is subsumed by the subhierarchy of the selectional restriction. If it is, then the restriction is passed.

Even when the verbal concept of the nominalization has been determined, identifying the role that the nominalization's modifier plays is difficult. For this reason it seems appropriate to wait until all of the roles stemming from prepositional phrases have been identified before trying to resolve the nominalization's modifiers in the NP. That way, candidate roles, if already filled, can be weeded out. Thus, only the remaining unfilled roles need to be checked. Below is a detailed description of noun phrase algorithm.

Noun Phrase Algorithm

Let n_0 be the head noun of some noun phrase that is currently under consideration, and one or more of the senses of n_0 represent a nominalized verb. Apply the algorithm below to determine the meaning of n_0:

1. If n_0 has a single sense, set the verbal concept vc_0 to this meaning and continue.
2. If the NP containing n_0 includes a genitive or a possessive pronoun, and there is no modifying "of" PP, attempt to realize the qualifier (or its anaphoric referent) as a thematic role of a verbal concept from the nominalization's root verb as follows:
   (a) If the verbal concept has been determined, then
      i. Fire nominalization mapping rules.
      ii. If no rules are triggered, check the genitive/pronoun against the selectional restrictions of the verbal concept's object and subject entries (subject entries only for intransitive verbs).
      iii. If no meaning was found for the gènitive/pronoun, goto step 4.
   (b) Else, (the verbal concept has not been determined)
      i. Fire nominalization mapping rules.
      ii. Else, if the nominalization is an -ing nominalization, fire the subject rules of the verb.
      iii. If no rules are triggered and the verb is intransitive, fire the subject rules of the verb.
      iv. Else, try both the object and subject rules of the verb. If only one type has a rule that fires, take the triggered verbal concept.
      v. If no rules are triggered, the verbal concept cannot be determined, exit.
3. If the NP containing n_0 includes some other qualifier and the verbal concept is unknown, exit.
4. Else, if the NP containing n_0 includes some other qualifier(s) and the verbal concept is known:
   (a) If the modifier is a noun, attempt to determine which role the noun plays in the verbal concept underlying the nominalization as follows:
      i. Examine the selectional restrictions found in the representation of the verbal concept.
      ii. If the modifier satisfies a single role, make that the interpretation of the modifier.
      iii. Else, procrastinate until more evidence is available.
   (b) Else, if the modifier is an adjective, determine if it can: fill an at-time role, e.g., the 1972 election; is derived from a noun which may fill a role; or is an ordinal adjective, in which case, mark the adjective as a temporal indicator.

Prepositional Attachment and Meaning Algorithm

The prepositional attachment and meaning determination algorithm is activated for each prepositional phrase found within the scope of some nominalization, which is defined to be any postnominal position within the same sentence clause as the nominalization, up to the main verb (for nominalizations before the verb) and to the end of the clause (for nominalizations after the verb). This algorithm has two objectives:

1. To determine if the prepositional phrase attaches to the nominalization, and
2. To determine the meaning of the prepositional phrase attachment within the context of the nominalization.

As each such prepositional phrase is parsed from left to right, the preposition is used to select either VM rules, if the verbal concept has not been established, or to select verbal concept selectional restrictions, if it has been established. If one of these rules fires, indicating that the nominalization takes the preposition, the PP is attached to the nominalization, and its thematic role is noted. The prepositional phrase attachment algorithm, shown below, is part of a general semantic interpretation algorithm that is described in (Gomez, Segami, & Hull 1997).

Prepositional Phrase Attachment Algorithm

Let n_0 be the head noun of some noun phrase where one or more of the senses of n_0 represent a nominalized verb, and let the verbal concept of that verb be vc_0. Let pp_1, pp_2, ..., pp_i be a list of one or more prepositional phrases that follow n_0 in the sentence. Apply the algorithm below to determine whether pp_i attaches to (modifies) n_0 and what the meaning of pp_i is, that is, its thematic role:

1. If the verbal concept, vc_0, underlying the nominalization is not known, then

(a) If the preposition is "of,"
   i. Use the "of" mapping rules of the nominalization, if any exist.
   ii. If no "of" mapping rules exist, attempt to fire the obj rules of the nominalized verb.
(b) Else, (the preposition is not "of")
   i. Use the appropriate mapping rules of the nominalization, if any exist.
   ii. Else, attempt to fire the preposition rules of the nominalized verb for the given preposition.

2. Else, if vc_0 is known,
   (a) If the nominalization has mapping rules for the preposition, use them to select the appropriate verbal concept entry. If the entry's selectional restriction is passed, goto step 3 else goto step 4.
   (b) Else, if the preposition is "of," try the obj entries of vc_0. If the entry's selectional restriction is passed, goto step 3 else goto step 4.
   (c) Else, use the entries of vc_0 indexed under the appropriate preposition. If the entry's selectional restriction is passed, goto step 3 else goto step 4.
3. pp_i attaches to vc_0, therefore, save the attachment to the nominalization, save its meaning, and exit.
4. If vc_0 does not claim pp_i, see if any superconcept of vc_0 has an entry, under the appropriate preposition or using the mapping, that vc_0 inherits, which determines attachment and meaning. Repeat step 2 with the ancestor of vc_0 recursively until either an attachment is found, or the list of superconcepts is exhausted.

Discussion

We examine the progress of the interpreter as the nominalization and its modifying constituents of the sentence below is parsed:

(5) *The king sent another fleet to break the* $\underbrace{\textit{Muslims'}}_{genitive}\ \underbrace{\textit{control}}_{n_0}\ \underbrace{\textit{over spice}}_{pp_1}\ \underbrace{\textit{in that country}}_{pp_2}$.

In (5), the meaning of "control" can not be determined by the noun phrase interpretation algorithm because "Muslims' control" may have several different interpretations. Now the PP algorithm is called with "over spice," and because the verbal concept is unknown, step 1 executes. The nominalization "control" has mapping rules for the preposition "over," which ultimately take the preposition strongly as its *theme*, and determine the verbal concept of "control" to be *control-physical-thing*. The last constituent, pp_2, is handled by step 2c, which attaches it to "control" as the location of the action. PP_2 can also be attached to "spice," but preference is given to the nominalization. The genitive modifier is handled by the end-of-clause algorithm, which is described in the next section.

End-of-Clause Algorithm

The end-of-clause algorithm is activated when the parser reaches the end of a clause containing a nominalization[2]. This algorithm has two objectives:

1. To determine the verbal concept of the nominalization, if it is still unknown, and to determine if the nominalization is being used in a non-verbal sense, and
2. To reevaluate each nominalization modifier to ensure that an interpretation has been found.

If the verbal concept is still undetermined, this algorithm makes one last effort to establish it. First the algorithm fires the end-of-clause rules of the root verb. If none of the rules fires, it may be that the nominalization is part of a collocation. The NP of the nominalization is used to search WordNet's list of collocations. This will provide the verbal concept in cases such as "primary election" and "free trade." If no matching collocation can be found and the nominalization has both verbal and non-verbal senses, a set of heuristics based on work by Grimshaw(Grimshaw 1990) is used to reject any verbal sense. In the absence of any thematic roles, a verbal sense can be rejected if the nominalization is plural or has an indefinite article, e.g., *Maxwell moved the controls* and *Tasha wanted a decoration.* If the verbal sense of the nominalization can be rejected and the nominalization has only one non-verbal sense, then that sense can be selected.

If these heuristics are unsuccessful or a non-verbal sense is selected, a verbal concept will not be found and further processing is abandoned. However, if the verbal concept is already known or is established by the first step of the end-of-clause, each prepositional phrase within the scope of the nominalization and each noun within the nominalization NP is reexamined to verify that it has been interpreted. Reexamination means to reactivate the appropriate algorithm for the nominalization modifier. This is necessary because the determination of the verbal concept might come after several prepositional phrases have been parsed.

End-of-Clause Algorithm

1. If end-of-clause rules do not determine the verbal concept,
   (a) Look up NP (minus articles, quantifiers, etc.) in WordNet's list of collocations.
   (b) If the meaning of the NP is found, save it and goto step 2.
   (c) Else, if the nominalization is plural and there are competing non-verbal senses, assume that this is a non-verbal use of the nominalization and exit.
   (d) Else, if there is only one verbal sense, make it the verbal concept.
2. Reevaluate each nominalization modifier (if verbal concept has been determined)
   (a) See if the modifier has either been assigned a thematic role or has been attached to some other constituent.
   (b) If a modifier that has not been attached is found, fire the appropriate rules (depending on whether the modifier is a prepositional phrase or resides within the nominalization's noun phrase).
   (c) If a rule fires, be sure that the thematic role indicated by the rule has not already been filled.
   (d) If the role has been filled, reject that role and continue firing any other appropriate rules.

Testing

The algorithms were tested to determine how successful they were in disambiguating the nominalization, and in recognizing the underlying verbal concept of the nominalization and filling its thematic roles. The discourse domain was comprised of biographical articles from the World Book Encyclopedia, which are being used in an ongoing research project to acquire historical knowledge from encyclopedic texts(Hull 1994).

[2] Actually, a general end-of-clause algorithm is activated when any clause ends. For the sake of brevity we will describe only those parts of the general end-of-clause algorithm related to the interpretation of nominalizations, and will treat them as a separate algorithm.

Table 1: Algorithm Results

nominal	n	senses	disambig.	gen.	NP	PP
arrest	29	3	100%	100%	33%	91%
birth	78	4	81%	100%	50%	97%
capture	34	5	100%	100%	50%	100%
control	421	11	72%	80%	40%	97%
defense	120	10	97%	96%	38%	97%
execution	25	4	92%	100%	0%	100%
murder	59	1	100%	100%	29%	98%
nomination	212	1	100%	87%	100%	91%
publication	69	3	79%	100%	47%	98%
trade	200	7	94%	89%	57%	97%
total	1247		88%	93%	71%	96%

The algorithms assume the existence of rules for disambiguating the root verb of each of the nominalizations, as well as the mapping rules for those syntactic constructions which are specific to the nominalization. The verb disambiguation rules had already been written as part of our ongoing research, and therefore, the effort needed to handle the nominalizations of these verbs was quite small. Moreover, a list of proper nouns representing proper names was used for recognizing people and locations.

Procedure

The results of the testing are shown in Table 1. Ten nominalizations were selected randomly from a list of nominalizations with at least 20 occurrences in 5000 biography articles from the World Book Encyclopedia. The column n shows how many occurrences of the nominalization were found in those articles. The algorithms were applied to each occurrence, and the results of the interpreter were examined to see if the nominalization was correctly disambiguated, if the genitive and the rest of NP was correctly interpreted, and how successfully the algorithms interpreted prepositional phrases modifying the nominalization.

Analysis

The results in Table 1 illustrate the strengths and the one limitation of the algorithms. The correct sense of each nominalization was selected more than 70% of the time, with the worst disambiguation score, 72%, occurring when testing "control," the most ambiguous nominalization with 11 WordNet senses. Failures to disambiguate were most often caused by situations where the verb rules could not be directly applied. For example, in the sentence *Court was noted for her endurance and control*, nothing triggers any of the verb rules. Further, because "control" has both verbal and non-verbal senses, one can't assume that this is an instance of either one. Other disambiguation errors resulted from rules that didn't fire or selected the wrong verbal concept, or that missed a non-verbal sense. On the whole, however, these algorithms provide an effective means of nominalization disambiguation.

The results of determining the thematic roles of deverbal nominalizations are given by the next three columns of Table 1. The thematic roles of genitives were found 93% of the time, showing how regular genitives are. The only statistically relevant problem involved two possessives used together, as in "his party's nomination" or "their country's trade." This problem could be easily handled in a general manner.

Interpreting the other elements of the noun phrase shows a limitation of the algorithms, which shouldn't be surprising considering the difficulty of NP interpretation. The most significant problem was the interpretation of adjectives which do not fill thematic roles but portray a *manner* of the action. Examples include "sole control," "tight control," "profitable trade," "mass murder," and "powerful defense." Related to this problem are other adjectives which are not manners of the action but could not be interpreted as thematic roles, e.g., "foreign trade," "extraordinary breath control," and "important capture."

PPs were correctly attached and their meaning determined over 90% of the time. This shows that the verb's mechanism for handling PPs can be readily used by the nominalization interpretation algorithms. Most failures were due to ambiguous PP heads and cases were the nominalization took prepositions different from the verb, which were unanticipated.

Related Research

Several knowledge-based approaches to interpretation of nominalizations can be found in the current literature. PUNDIT is a system for processing natural language messages, which was used for understanding failure messages generated on Navy ships(Dahl, Palmer, & Passonneau 1987). Nominalizations in PUNDIT are handled syntactically like noun phrases but semantically as clauses, with predicate/argument structure. In fact, PUNDIT uses the same decomposition as the associated verb. Special nominalization mapping rules are used to handle the diverse syntactic realization of constituents of nominalizations. Some components of our approach are similar; nominalizations inherit selectional restrictions and syntactic mappings from their associated verbal concepts and can have their own specialized mappings when appropriate. PUNDIT avoids handling the ambiguity of the nominalization, including the ambiguity between the verbal and non-verbal senses and the polysemy of the nominalized verb. KERNEL(Palmer *et al.* 1993), a successor of PUNDIT, treats nominalizations in much the same way. Voorhees(Voorhees 1993) and Li *et al.*(Li, Szpakowicz, & Matwin 1995) both use WordNet as a source of disambiguation information, but neither addresses interpretation of nominalizations.

Grimshaw(Grimshaw 1990) states that a subclass of nouns, which she refers to as *process* or *event nominals*, have argument structure that is filled by grammatical arguments. Further, these arguments are obligatory to the same extent to which they are obligatory for the nominal's associated verb. Other nouns, which she calls simple events or result nominals, do not have ar-

gument structure though they may take either complements or modifiers. Grimshaw explains that the common belief that nouns take arguments optionally (Anderson 1983; Dowty 1989) is really just a case of confusing ambiguous nouns that have both event and simple or result senses, e.g., examination. She then provides a comprehensive list of evidence supporting the notion that nouns take obligatory arguments, including methods for disambiguating these nouns. While knowing that a particular nominalization does or does not have argument structure can help in choosing between its verbal and non-verbal senses, it can not disambiguate the nominalization further. Moreover, the restriction that argument structure is obligatory begs the question of what to do when not all the arguments are present and the nominalization clearly describes an action. This phenomenon, illustrated by the sentences below, occurs quite frequently:

(6) *Some of Johnson's accusers tried to implicate him in Lincoln's* **murder**, *but failed.*

(7) *When the news of Pompey's* **defeat** *at Pharsalus in 48 B.C. reached him, Cato fled to North Africa.*

(8) *He saw that city's* **destruction** *by British and American bombing in 1945.*

Although the nominalizations "murder," " defeat," and "destruction" in the sentences above do not meet Grimshaw's criteria for having argument structure, they do take the arguments "Lincoln," "Pompey," and "city" respectively as their *themes* and they do denote events. Instead of portraying these nouns as passive nominals and calling their arguments *adjuncts*, our approach handles them in the same manner as if they had written as *the murder of Lincoln*, *the defeat of Pompey*, and *the destruction of that city.*

Conclusions

We have provided knowledge-based algorithms for the semantic interpretation of nominalizations. These algorithms address the problems of differentiating between the nominalization's verbal and non-verbal senses and interpreting the nominalization when it occurs as the head noun of NPs. Interpreting the nominalization involves determining the predicate of the nominalization when it is polysemous, and determining the attachment of the nominalization's PP modifiers and the identification of their thematic roles.

One major limitation of this approach is the need for having hand-crafted representations of VM rules, verbal concepts and a general ontology. We are working on a parallel project that integrates WordNet's lexical knowledge-base into our system. Preliminary results indicate that the task of defining VM rules and verbal concepts can be highly simplified by interfacing our ontology with the WordNet noun ontology and verb hierarchy.

Acknowledgments

This research has been funded by NASA-KSC Contract NAG-10-0120.

References

Anderson, M. 1983. Prenominal genitive nps. *Linguistic Review* 3:1–24.

Dahl, D.; Palmer, M.; and Passonneau, R. 1987. Nominalizations in PUNDIT. In *Proc. of the 25th Annual Meeting of the ACL*, 131–137.

Dowty, D. R. 1989. On the semantic content of the notion 'thematic role'. In Chierchia, G.; Partee, B. H.; and Turner, R., eds., *Properties, Types, and Meaning.* Dordrecht: Kluwer.

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

Grimshaw, J. 1990. *Argument Structure.* Cambridge, Mass.: MIT Press.

Hirst, G. 1992. *Semantic interpretation and the resolution of ambiguity.* New York: Cambridge University Press.

Hull, R. 1994. Acquisition of historical knowledge from encyclopedic texts. Technical report, University of Central Florida, Department of Computer Science. CS-TR-94-05, dissertation proposal.

Jacobs, P., and Rau, L. 1993. Innovations in text interpretation. *Artificial Intelligence* 63:143–191.

Li, X.; Szpakowicz, S.; and Matwin, S. 1995. A wordnet-based algorithm for word sense disambiguation. In *Proc. of IJCAI-95*, 1368–1374.

Marcus, M.; Hindle, D.; and Fleck, M. 1983. D-theory: Talking about talking about trees. In *Proc. of the Annual Meeting of the ACL.*

McRoy, S. 1992. Using multiple sources for word sense discrimination. *Computational Lingusitics* 18:1–30.

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.

Palmer, M.; Passonneau, R.; Weir, C.; and Finin, T. 1993. The kernel text understanding system. *Artificial Intelligence* 63:17–68.

Quirk, R.; Greenbaum, S.; Leech, G.; and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language.* New York: Longman Group, Inc.

Voorhees, E. 1993. Using wordnet to disambiguate word senses for text retrieval. In *Proc. of the 16th Annual Int'l ACM SIGIR Conference on Research and Development in Information Retrieval*, 171–180.

Weinberg, A. 1993. Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research* 22(3):339–364.