Natural Language Engineering **6** (2): 139–161. Printed in the United Kingdom

© 2000 Cambridge University Press

Semantic interpretation of deverbal nominalizations†

RICHARD D. HULL

Elagent Corporation, 7001 N. Atlantic Ave, Suite 203, Cape Canaveral, FL 32920, USA

FERNANDO GOMEZ

Department of Computer Science, University of Central Florida, Orlando, FL 32816, USA

(*Received April 1998; revised 27 March 2000*)

Abstract

An algorithmic approach to the semantic interpretation of deverbal nominalizations found in encyclopedic texts, such as *support*, *publication* and *control*, is described. Interpreting these nominalizations is crucial because they are quite common in encyclopedic texts, hence a great deal of information is represented within them. Interpretation 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.

1 Introduction

Quirk, Greenbaum, Leech and Svartvik (1985) define 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 such as 'destruction') or adjective (deadjectival noun such as 'truth'). 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. Deverbal nominalizations are a proper subset of all event nominals, as nouns including 'accident' and 'disaster' denote events and yet do not have a systematic correspondence with any verb. Understanding sentences with deverbal nominalizations requires the ability to determine the meaning of the nominalization,

† This paper is an extended version of a paper which appears in the *Proceedings of AAAI-96* entitled 'Semantic interpretation of nominalizations'.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

and to make sense of the nominalization's modifiers, namely, the other words in the Noun Phrase (NP) containing the nominalization and any prepositional phrases that follow it.

One of the most often cited issues associated with interpreting and representing nominalizations is the decision process that leads to the determination of whether a given nominalization refers to an event, process, state or result. This is linked to the Aristotelian categorization of verbs into events, states, etc., that is popular in the literature (Vendler 1967; Dahlgren 1988; Pustejovsky 1995; Grimshaw 1990).

1.1 Selecting between verbal and non-verbal senses

Another important problem is distinguishing between the verbal and non-verbal senses of the nominalization, as is necessary for words like 'support', 'publication' and 'decoration'. 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 'support' has several senses, including a physical thing that bears weight, as in *The bridge supports are corroded*, and the act or process of supporting, as in *The politician's support of education*. The noun publication also has several senses, including a published work and the act or process of publishing. Finally, the senses of 'decoration' include a military badge of honor, an ornament and the nominalization sense, which means the act or process of decorating. While these senses are logically related in that a physical decoration is often the result of some instance of 'decorating', we do not explicitly represent these relations.

When confronted by the phrase 'Maxwell's decoration', one cannot decide between the different senses of decoration. However, consider this phrase in two different contexts:

(1) *Maxwell's decoration of the room took most of the day*, and
(2) *Maxwell's decoration hung appropriately on the wall.*

Disambiguating 'decoration' is accomplished by realizing that the prepositional phrase 'of the room' in sentence 1 provides 'what was decorated' in a decorating event, implying that decoration is being used in the sentence as a nominalization. This is not the case in sentence 2, and the absence of 'what was decorated' suggests a nominal sense of decoration. Of course, when the nominalization stands alone, as in *Maxwell saw the decoration*, there are no thematic roles present, i.e. no evidence is available for disambiguation. In this case, the definite article modifying 'decoration' causes the initiation of anaphoric reference resolution, which may determine that 'decoration' refers to a previously mentioned decorating event or nominal sense.

1.2 Disambiguating verbal senses

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 both corpus-based approaches (Brown, Della Pietra, Della Pietra and Mercer 1991; Yarowsky 1992; Wilks, Guthrie, Guthrie

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

and Cowie 1992) and knowledge-based approaches (McRoy 1992; Hirst 1992; Jacobs and Rau 1993; Voorhees 1993; Li, Szpakowicz and Matwin 1995) an algorithmic approach to deverbal nominalization disambiguation has not, to our knowledge, been implemented and tested on real-world texts. 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.*

When one comes upon the sentence

(3) *General Dwight D. Eisenhower forced Patton to apologize, and Congress temporarily held up his permanent promotion to major general.*

the meaning of the nominalization 'promotion' must be disambiguated from the three possible verb meanings. In this case, it is the prepositional phrase 'to major general', which indicates that this use of promotion is the act of promoting a person to a position of higher authority.

1.3 The importance of interpreting deverbal nominalizations

The framework of a model of semantic interpretation that can be used to solve these problems is described in detail by Gomez, Segami and 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.

Our focus on the semantic interpretation of nominalizations is necessary for automatically acquiring knowledge from encyclopedic texts, which was the goal of a recent investigation. This project, called SNOWY-BIOS, sought to use the knowledge acquired from more than 5000 World Book Encyclopedia articles to answer a large set of questions about famous people (Hull 1997). Each nominalization represents a clausal structure, therefore, for the task of acquiring knowledge, a nominalization is just as important as any other clause. Moreover, ignoring nominalizations can cause problems with the interpretation of the clauses in which they appear because nominalizations are contenders for Prepositional Phrases (PPs), and if these modifiers are not attached to them, PPs may be mistakenly attached to other constituents in the sentence, particularly the clause's main verb. In *Harold heard about Edison's improvements on the original telegraph*, the PP 'on the telegraph' can potentially modify both the nominalization 'improvement' and the verb 'heard'. If the PP is not attached to the nominalization because we are unable to determine an appropriate interpretation of this attachment, then the verb will face no competition for it and the PP may be wrongly attached.

These two points would be academic were it not for the fact that nominalizations are so common in the encyclopedia articles we have been studying. In a random selection of 25 paragraphs from biographical articles of the World Book Encyclo-

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

In 1935, he became army chief of staff. The following year, the leftists won the election and sent Franco to a post in the Canary Islands. Military leaders plotted to overthrow the leftist government in 1936. Franco delayed taking part in the *plot*, but he was promised command of the most important part of the army. The revolt began in July 1936 and it started a total civil war. Two and a half months later, the rebel generals named Franco commander in chief and dictator. Franco's forces, called Nationalists, received strong *support* from Italy and Germany. On April 1, 1939, after 32 months of bitter fighting, the Nationalists gained complete victory. Franco then became dictator without opposition.

Fig. 1. Passage from the World Book Encyclopedia monograph of Francisco Franco.

pedia, 23 had two or more nominalizations. To illustrate the kinds of texts we are analyzing, consider a short passage from an article about Francisco Franco in the World Book Encyclopedia.

Some typical uses of nominalizations are present in figure 1. The 'plot' mentioned in the fourth sentence refers to the event mentioned in the sentence before, and is used to describe Franco's lack of involvement in the plan. Anaphoric usages of nominalizations are commonly employed to further specify the thematic roles of the original predicate, or as a means for relating the original predicate to new information. Another use is illustrated in the seventh sentence. In, "Franco's forces [...] received strong support from Italy and Germany", the nominalization 'support' greatly affects the meaning of the clause. Light verbs such as 'receive' are often used in conjunction with nominalizations, and it is the underlying verb of the nominalization that usually specifies the predicate of the clause. Other nominalizations in this paragraph are election, revolt, fighting and opposition. We do not address the problem of light verbs in this paper. However, we are saying that part of the problem associated with light verbs, namely their lack of predicative force, can be resolved by identifying the clause structure of the nominalizations when these verbs appear associated with nominalizations.

If we do not interpret the nominalizations in the passage above, we will not be able to answer questions like *Which nations supported Franco's forces? Did Franco plot to overthrow the government in 1936? Who was elected in 1936?* Moreover, if we cannot answer these questions, then our attempt to acquire knowledge from this text has been severely compromised.

Failure to interpret a nominalization often means missing the opportunity to acquire knowledge which could not be acquired in any other way. Of the nominalizations tested in World Book biographies, *trade* most clearly illustrates the necessity of nominalization interpretation for the purpose of knowledge acquisition. There are 200 occurrences of *trade* used as a nominalization, and only 47 occurrences of *trade*, *traded*, *trades* and *trading* used as a verb. Therefore, if one is interested in acquiring knowledge about 'who traded what with whom', the nominalization form must be handled, or more than 80% of the trade relations will not even be considered.

The remainder of this paper is comprised of five sections. Section 2 describes the essential aspects of the semantic interpreter of which the nominalization algorithms, explained in section 3, are a part. The testing of these algorithms on a collection of

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

nominalizations is discussed in section 4. A comparison of this approach to other work in the literature is explained in section 5. Finally, section 6 presents the authors' conclusions.

2 The semantic interpreter

The parser used by our semantic interpreter leaves structural syntactic relations underspecified along the lines of D-Theory (Marcus, Hindle and Fleck 1983; Marcus 1987; Marcus and Hindle 1990) and minimal commitment parsers (Gorrell 1991; Weinberg 1993). Throughout the paper we will use the following syntactic relations analysed by the parser: 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. As indicated, the parser 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. These rules are called verb meaning rules, or *VM rules*. For instance, what determines the meaning of 'drive' in the examples below 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*):

(4) *Peter drove the bus.*
(5) *Peter drove the nail.*

In addition to NP complement rules, Prepositional Phrase rules (PP rules) are also used to determine the meaning of the verb. In the examples below, 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 sentence (6) because the object of the PP is a human, and as *depart* in sentence (7) because the object of the PP is a location:

(6) *Jennifer left the orange groves for her son.*
(7) *Jennifer left the orange groves for home.*

2.1 Verbal knowledge

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. Examples of NP complement rules and preposition rules for the verb 'defend' are shown in figure 2.

The first set of rules selects the verbal concept based upon the NP complement. If the direct object (object1) is a subconcept or an instance of a *location* within a

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

```
VM Rules:

(defend
   ((object1
     ((if% is-a obj1 location)
           (verbal-concept-is defend-physical-thing))
     ((if% is-a obj1 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)))
  )
)
```

Fig. 2. VM Rule for the verb *defend*.

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*, and as we will see to handle *Carleton's defense of 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*. This rule handles *She defended the theorem* and *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 in figure 3.

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 defend is subsumed by the *agent* concept in the concept hierarchy, then make it fill the *actor* role. The following entry allows the object of a defence to be anything. Finally, the last entry describes how to handle PPs that begin with 'against' or 'from'. The object of a PP of the form *against* < *object* >, where < object > is anything, is made to fill the *thing-performing-attack* role. The qualifier *strong* on these entries indicates that the preposition is claimed strongly by the verbal concept (see Gomez *et al.* (1997) for a detailed discussion of this distinction). This handles cases such

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

```
(defend
  (is-a (action-r))
  (subj (agent (actor)))
  (obj1 (thing (theme))
  (prep (against (thing (thing-performing-attack (strong))))
        (from (thing (thing-performing-attack (strong)))))
)

; WordNet sense defend1, 4, 7
(defend-idea
  (is-a (defend))
  (obj1 (idea (theme)))
 )

; WordNet sense defend2, 5
(defend-physical-thing
  (is-a (defend))
  (obj1 (physical-thing (theme)))
)

 ; WordNet sense defend3
(defend-Championship
  (is-a (defend))
  (obj1 (championship (theme)))
)

; WordNet sense defend6
(legal-defend
  (is-a (defend))
  (obj1 (agent (theme)))
  (prep (in (court (at-loc (strong)))))
)
```

Fig. 3. The verbal concepts encapsulating the meaning of *defend*.

as: *Peter defended Mary against the bullies*, where the object of the preposition is an agent; *Peter defended Mary against the criticism*, where the object of the PP is an idea; and *The walls defended the city from the elements*, where the object of the PP is a phenomenon. 'From' PPs are handled identically.

The ontology used by the system described in this paper is not the WordNet ontology, although there are strong relations. Our ontology was developed over the last thirteen years, and was motivated by the semantic interpretation algorithms. A partial description of these algorithms appears in Gomez and Segami (1989). The category *idea* in our ontology corresponds to the category *abstraction* in WordNet, and *abstraction* is the top hypernym of one of the WordNet senses of 'criticism'. However, none of the senses of 'idea' has *abstraction* as one of its hypernyms in WordNet.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

Each of the subconcepts of *defend*, *defend-idea*, *defend-physical-thing*, *defend-championship* and *legal-defend*, inherits entries from *defend*. Therefore, the only entries specified in the subconcepts are those that differentiate them from their parent. For example, the theme of *defend-physical-thing* is a *physical-thing*, so an object entry states this restriction. The syntactic relations, *subj*, *prep against* and *prep from* are inherited from *defend*. Still others are inherited from *action-r*. In *The Texans defended the Alamo from the Mexican Army in 1836*, the verbal concept selected by the VM rules for defend is *defend-physical-thing*. This verbal concept does not have entries for the prepositions 'from' and 'in', however, our representation allows inheritance of entries through *is-a* slots. Therefore, because *defend-physical-thing is-a defend*, the entry for 'from' is inherited from *defend*. Furthermore, *defend is-a action-r* and an entry for the preposition 'in' is inherited from *action-r*.

This fine-grained decomposition of 'defend' is necessary if one wishes to make specific inferences depending on the type of defence. For example, one might want to infer that the theme of *defend-idea* is believed by the defender, or that the theme of *legal-defend* is having legal trouble. Inference rules can be attached to each verbal concept, and are fired once processing of the clause has completed.

2.2 Leveraging verbal knowledge for nominalizations

The structures above embody the knowledge necessary to understand clauses containing the verb 'defend' (see Gomez *et al.* (1997) for a detailed discussion of VM rules and verbal concepts). This verbal knowledge can be exploited and reused for the 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', i.e. necessary for a verbal sense interpretation, is also stored.

The representation for 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 subjects of 'attend'. In this case, a slot specifying that the genitive should be meet the restrictions of the subject

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

of the verb 'attend' is needed. The mapping rules for attendance are:

```
(attendance
  (obligatory - role (theme))
  (genitive (subj))
  (prep (at (obj1))  ; at --> obj in ''Peter's attendance at the dance''
        (in (obj1))  ; in --> obj in ''Peter's attendance in the House''
        (of (subj))) ; of --> subj in ''The attendance of Peter''
```

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.

Let us now look at how the verbal knowledge along with the nominalization's information are used. Consider the sentence below:

(8) *His closing argument for the* **defense** *in a case against Henry Ward Beecher lasted eight days.*

'Defense' is ambiguous, and may indicate a nominalization, or as in this case, a noun meaning 'the defendant and his/her lawyer(s)'. It is the absence of the obligatory thematic role *theme*[1], which is necessary for a verbal sense interpretation, that determines that this is not a nominalization. Hence, the meaning of 'the defense' is not the act of defending something or somebody, but rather 'the defense team'. Alternatively, the next examples show how the presence of the *theme* of the nominalization indicates a verbal reading.

(9) *Leading abolitionists contributed money for her* **defense** *in court.*

(10) *When Lee gave up the* **defense** *of Petersburg and Richmond, Ewell took command of the remnant of a corps.*

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 sentence (9), and it is the prepositional phrase 'of Petersburg and Richmond' that selects the meaning *defend-physical-thing* in sentence (10). Note that the meaning of 'defense' in sentence (9) is not determined by the pronoun 'her', because even though the genitive entry of 'defense' may represent the subject or object of 'defend', no VM Rules are triggered by a human subject (in this case there are no subject rules) or a human object. Therefore, the thematic role for 'her' is left undetermined until the end of the clause is reached. At that time the meaning of 'defense' is determined to be *legal-defend*, and the genitive mapping is reapplied using the verbal concept's specification for selectional restrictions on the object and subject. While 'her' satisfies

[1] It should be noted that the *theme* need not be specified in the current sentence, but may have been provided previously in an anaphoric reading of 'the defense'.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

the restrictions of both the subject and object, the object entry is preferred based upon the order in which the algorithm tries them, and 'her' would be made to fill the *theme* role. If the sentence had been 'her defense *of Peter* in court', 'of Peter' would have already specified the *theme* thereby causing the algorithm to select the subject entry, making 'her' fill the *actor* role. The algorithms described in the next section explain exactly how polysemy and role determination are handled.

3 Interpretation algorithms for nominalizations

The interpretation algorithms attempt to determine the verbal concept of the nominalization and to fill its thematic roles. The algorithms are activated by the noun phrase interpreter, a component of the semantic interpreter, when the parser has recognized a noun phrase in the sentence. Determination of the verbal concept requires disambiguation of the meaning of the nominalization's root verb.

For example, the nominalization 'execution' is generated from the verb 'execute', which has several meanings: to fulfill or perform; to administer; to put to death; to complete or make valid (to execute a deed or lease); and to create with an idea or plan (to execute a statue in marble).

The 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.

It is the prepositional phrase 'of Louis Riel' that determines the meaning of the nominalization in 'The Macdonald administration's execution of Louis Riel was widely protested'.

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.

3.1 The nominalization noun phrase algorithm

The nominalization noun phrase algorithm is triggered when the head noun[2] of some NP is determined to be a potential nominalization. This is accomplished by consulting WordNet (Miller, Beckwith, Fellbaum, Gross and Miller (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.
2. To identify which thematic roles of the verbal concept, if any, each of the remaining nouns and adjectives of the NP fill.

[2] The algorithm does not currently handle nominalizations in positions other than the head such as 'election results'.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

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.

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.

3.1.1 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 th 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.

(11) *Lincoln's election; The representatives elected Lincoln.*
(12) *Metternich's resignation; Metternich resigned.*

In sentence (11), the genitive corresponds to the object position, while in sentence (12), it corresponds to the subject position. The verb *resign* is intransitive, except for collocative expressions such as 'resigned his office' or 'resigned his seat', 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, i.e. when the nominalization has a systematic correspondence to a passive sentence as in *The play's hostile reception by the critics* vs. *The play was received in a hostile manner by the critics.* 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, i.e. subject rules, object rules, or rules for other syntactic relations. Any preference on the rules, say to try the subject rules first, will be wrong in a significant percentage of these

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

situations. Moreover, it is unproductive to try all of the rules in hopes that only the appropriate one would be triggered because in many verbal concepts the subject and object rules will have similar selectional restrictions. Therefore, disambiguation is postponed until more evidence is available in the form of roles filled by prepositional phrases which modify the nominalization. Of course, if such PPs are not present in the sentence we may be faced with not interpreting the nominalization or making mistakes in some cases.

The VM rules described in this article were developed for the purpose of automatically acquiring knowledge from encyclopedic biographies. The ultimate goal of this investigation was to answer English questions involving a diverse set of verbal relationships as opposed to constructing a single hierarchy of closely related ones. 123 verbal concepts and VM rules for over 150 verbs were created. Verbal concepts can have more than one verb mapped to them (due to synonymy), and a single verb can be mapped to several verbal concepts (due to polysemy). These newly created verbal concepts were then integrated into the existing VC hierarchy of SNOWY (Gomez *et al.* 1997).

The taxonomy of verbs was motivated by the application of our system, namely acquiring knowledge from biographies from the encyclopedia. WordNet verb classes were consulted in building this taxonomy, but no effort was made to base our verbal concepts, or predicates, on WordNet verbs classes, and much less to link the selectional restrictions of our verbal concepts to WordNet hypernyms. However, in our subsequent research we have based our taxonomy of verbal concepts on WordNet verbs classes and used WordNet hypernyms in the selectional restrictions of the verbal concepts. See Gomez (1998) for a detailed discussion.

3.1.2 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 (Katz 1990) 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.

This task is not a simple one, however. Filling thematic roles presents a challenge because of the tremendous difficulty associated with interpreting noun phrases consisting of multiple nouns (Finin 1980; McDonald 1982). Consider the range of possible roles that the prenominal modifier represents in the examples below:

- knowledge representation, shoe sale (theme)
- boat ride, grenade attack (instrument)
- UNICEF donation (recipient or actor)
- bank acquisition (actor or theme)
- Cambridge transportation (location).

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

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. In that way, candidate roles, if already filled, can be weeded out and only the remaining unfilled roles need to be checked.

The definitions of several terms and functions are necessary before presenting a detailed description of the noun phrase algorithm. These definitions will also apply to components of the prepositional phrase and end-of-clause algorithms.

Let a noun phrase under consideration be defined as $NP = q_n \ldots q_1 n_0 pp_1 \ldots pp_m$, where n_0 is the head of NP, with $n \geq 0$ qualifiers and $m \geq 0$ prepositional phrases. Let S_0 denote the set of all senses $s_1 \ldots s_{|S_0|}$ of n_0, and suppose that one or more of these senses represent a nominalized verb v_0.

Several functions access lexical information in the lexicon. Function *IngNominal*(n_0) returns True if n_0 is an ing-nominalization (e.g. *killing*), False otherwise. Function *Intransitive*(v_0) returns true if the underlying verb of the nominalization does not take a direct object. Function *Genitive*(q_i) returns True if q_i is a genitive possessive, False otherwise. Function *PossPronoun*(q_i) returns True if q_i is a possessive pronoun, False otherwise. Function *Preposition*(pp_i) returns the preposition of prepositional phrase pp_i.

Two functions use the VM and mapping rules described in the previous section to determine the verbal concept. Function *ApplyMappingRules*(n_0, q_i) returns the verbal concept of the nominalization or NIL if no mapping rules are fired. Function *Apply Verb Rules*(v_0, (*syntactic cases*), *modifier*) returns the verbal concept of the nominalization or NIL if no verb rules are fired.

Finally, two procedures determine the thematic role of the verbal concept filled by a modifier of the nominalization. Procedure *ApplySelctionalRestrictions*(vc_0, q_i, (*syntactic cases*)) attempts to realize q_i as a thematic role of vc_0 using vc_0's syntactic case entries for the elements of the third parameter. Procedure *FindRoleForQual* (vc_0, q_i) checks whether the qualifier can fill an open thematic role of vc_0.

The underlying mechanisms of these supplementary functions are described in an earlier monograph and will not be presented here. Given these definitions, apply the algorithm in figure 4 to determine the meaning of n_0 and the meaning of its qualifiers.

3.2 The 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.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

```
if |S0| = 1 then vc0 = s1; /* trivial case */
for i = 1 to n loop /* loop through qualifiers */
   if vc0 = NIL then
      if (! (vc0 = ApplyMappingRules (n0,qi))) then
         if (IngNominal (n0) || Intransitive (v0)) then
            vc0 = ApplyVerbRule (v0, subject, qi)
         else
            vc0 = ApplyVerbRule (v0, (object1 subject), qi)
         end else
      end if
   end if
   else /* the verbal concept has been determined */
      if Genitive (qi) or PossPronoun (qi) then
         if Preposition (pp1) !="of" then
            if (!ApplyMappingRules (n0,qi)) then
               ApplySelectionalRestrictions (vc0, qi, (object1 subject))
            end if
         end if
      else
        FindRoleForQual (vc0, qi);
      end else
    end else
end loop
```

Fig. 4. Noun phrase algorithm.

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 *et al.* (1997).

One additional definition is necessary: the function *SuperConcept*(vc_0) returns the parent concept of vc_0 in the verbal concept hierarchy or NIL if one is not defined. Apply the algorithm in figure 5 to determine whether pp_i attaches to (modifies) n_0 and what the meaning of pp_i is, i.e. its thematic role.

Discussion

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

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

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

```
for i = 1 to n loop
  if vc0 = NIL then
    if (!(vc0 = ApplyMappingRules(n0, ppi))) then
      ifPreposition (pp1) ="of" then
        ApplyVerbRule(v0,(object),ppi)
      else
        ApplyVerbRule(v0,Preposition(ppi),ppi)
      end if
    end if
  else /* vc0 has been determined */
    do
      if(!ApplyMappingRules(n0,ppi))then
        if Preposition(pp1) ="of"then
          ApplySelectionalRestrictions(vc0, ppi, (object1))
        else
          ApplySelectionalRestrictions(vc0,ppi,Preposition(ppi))
        end if
      end if
    until(attachment is established or !(vc0=SuperConcept(vc0)))
  end else
end loop
```

Fig. 5. Prepositional phrase attachment algorithm.

In sentence (13), the meaning of 'control' cannot 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, mapping rules are applied. 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 applying the selectional restrictions stored under the preposition 'in', which attaches pp_2 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 left uninterpreted by the noun phrase algorithm is handled by the end-of-clause algorithm, which is described in the next section.

3.3 The end-of-clause algorithm

The end-of-clause algorithm is activated when the parser reaches the end of a clause containing a nominalization.[3] 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.

[3] 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.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

```
if (!(vc0=ApplyEndOfClauseRules))then
  if (! (vc0=WordNetCollocation(n0,q1 ... qn))) then
    if (exactly one sense of n0,si, is a verbal sense) then
      vc0 =si
    end if
  end if
end if
if vc0!=NIL then
  for i=1 to n loop
    if Uninterpreted (qi) then
      Activate Noun Phrase Algorithm for qi
    end if
  end loop
  for i=1 to m loop
    if Uninterpreted (ppi) then
      Activate Prepositional Phrase Attachment Algorithm for ppi
    end if
  end loop
end if
```

Fig. 6. End-of-clause algorithm.

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 (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.

The end-of-clause algorithm has several additional supplementary functions which must be defined. Function *ApplyEndOfClauseRules*( ) makes a last effort to determine

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

the verbal concept if it has not been determined, returning the verbal concept is a rule fires or NIL otherwise. Function *WordNetCollocation*(n_0, $q_1 \ldots q_n$) returns the verbal concept suggested by a collocation in WordNet or NIL otherwise. Function *Uninterpreted*(*constituent*) returns TRUE if the constituent has not been associated with a thematic role of the verbal concept, and FALSE otherwise. Apply the algorithm in figure 6 to complete interpretation of the nominalization and its modifiers.

4 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 1997).

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. Additionally, a list of proper nouns representing proper names was used for recognizing people and locations.

4.1 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.

4.2 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

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

Table 1. *Algorithm results*

Nominalization	*n*	Num Senses	Disambiguate Correctly (%)	Interp Genitive (%)	Interp NP (%)	Interp 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

Table 2. *Results of selecting the most frequent sense*

Nominalization	*n*	Num Senses	Disambiguate correctly (%)
Arrest	29	3	100
Birth	78	4	22
Capture	34	5	0
Control	421	11	10
Defense	120	10	38
Execution	25	4	92
Murder	59	1	100
Nomination	212	1	100
Publication	69	3	22
Trade	200	7	6
Total	1247		

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.

In comparison, Table 2 shows the results of applying a simple algorithm for nominalization disambiguation. This algorithm selects the most frequent sense of the noun under consideration from version 1.5 of WordNet. The most frequent sense was effective for four of the ten nominalizations tested, but faired poorly on the other six. Clearly, the poor showing of this strategy shows that choosing the most 'frequent' sense is successful only when WordNet's idea of most frequent agrees with the use of that sense in the encyclopedia article. Moreover, establishing which sense is used does not provide a means for determining the thematic roles of the nominalization.

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.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

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 where the nominalization took prepositions different from the verb, which were unanticipated.

The results for all 10 nominalizations are presented in Table 3. Shown are the number of occurrences and the number correctly interpreted in the areas of clauses, nominalizations, noun phrase modifiers, prep phrase modifiers, and genitives. There were 310 errors in 305 clauses. Included in Table 3 is a breakdown of what the problems were. Percentages are based on the total number of errors.

Table 3. *Causes of errors*

```
Clauses    : 1247, completely correct:  942 (75·5%)
Nominals   : 1099, disambig correctly:  962 (87·5%)
Total NPs  :  437, correct:   310 (70·9%)
Total PPs  :  875, correct:   841 (96·1%)
Total Gen  :  149, correct:   139 (93·3%)

Problem Categories:

Disambiguation
  56  18·1%  No verb rules attempted
  44  14·2%  Special preposition construction
  24   7·7%  Missed non-verbal sense
   4   1·3%  No verb rules fire
   4   1·3%  Wrong verbal concept selected
   2   0·6%  No verb rules defined for verbal concept

Noun Phrase Interpretation
  84  27·1%  No interpretation for adjective modifier (manner)
  34  11·0%  No interpretation for adjective modifier
  10   3·2%  Role not found because ambiguous
   4   1·2%  Assigned incorrect role

Prep. Phrase Attachment and Meaning
  29  9·4%  No prep rules for preposition or missing
             selectional restriction
   3  1·0%  Wrong meaning assigned

Genitives
   5  1·6%  Double genitive, one is ignored

Misc.
   7  2·3%  Anaphoric reference
```

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

5 Related research

A major similarity between our work and that of Pustejovsky (1995) is that the syntactic realization of the verb's complements determine how syntactic relations are interpreted semantically. This principle is an essential component of our algorithms for determining the meaning of PPs. However, we also base the semantic interpretation process on the following additional principle: the determination of the meaning of the verb depends on the meaning of its complements. The determination of the meaning of the verb, on one hand, and of the meaning of the PPs and NP complements, on the other hand, are synergistic and interdependent processes as we have illustrated throughout the paper.

We strongly disagree with Pustejovsky's view that "the distinctions possible with thematic roles are much too coarse-grained to provide a useful semantic interpretation of a sentence" (Pustejovsky 1995; ch. 2, p. 6). This is a view that is not supported in *The Generative Lexicon*. Many of the views claiming that thematic relations are not adequate meaning representation structures for capturing the meaning of the sentence are based on reductionists views of verb meaning, namely verb meaning can be reduced to a small set of primitives from which one can derive all inferences needed (Jackendoff 1990; Dowty 1979; Schank 1975). Our view, however, is based on Quine's (1960) critiques to reductionist views of meaning. Meaning is holistic and intersentential (see Gomez (1998) for a discussion). In our opinion, we have gathered some evidence in this paper supporting the view that thematic relations are a useful, and perhaps, necessary step in the stairway that leads to the goal of meaning.

Several implementations of knowledge-based approaches to the interpretation of nominalizations can be found in the current literature (Proc. COLING-ACL Workshop on the Computational Treatment of Nominals 1998). PUNDIT is a system for processing natural language messages, which was used for understanding failure messages generated on Navy ships (Dahl, Palmer and 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, Passonneau, Weir and Finin 1993), a successor of PUNDIT, treats nominalizations in much the same way. Voorhees (1993) and Li, Szpakowicz and Matwin (1995) both use WordNet as a source of disambiguation information, but neither addresses interpretation of nominalizations.

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

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

events or result nominals, do not have argument 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 cannot 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:

(14) *Some of Johnson's accusers tried to implicate him in Lincoln's* **murder**, *but failed.*
(15) *When the news of Pompey's* **defeat** *at Pharsalus in 48 B.C. reached him, Cato fled to North Africa.*
(16) *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 been written as *the murder of Lincoln*, *the defeat of Pompey* and *the destruction of that city*.

6 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. We have tested the algorithms in a real-world domain consisting of biographical articles from the World Book Encyclopedia.

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.

Acknowledgement

We would like to acknowledge two anonymous referees whose insightful comments greatly improved this manuscript.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

References

Anderson, M. (1983) Prenominal genitive NPS. *Linguistic Review* **3**: 1–24.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J. and Mercer, R. L. (1991) Word-sense disambiguation using statistical methods. *Proc. 29th Annual Meeting of the ACL*, pp. 264–270.

Proc. of the COLING-ACL Workshop on the Computational Treatment of Nominals (1998) Montreal, Quebec, Canada.

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

Dahlgren, K. (1988) *Naive Semantics for Natural Language Understanding*. Kluwer Academic Press, Boston.

Dowty, D. R. (1979) *Word Meaning and Montague Grammar*. Reidel, Dordrecht.

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

Finin, T. (1980) The Semantic Interpretation of Nominal Compounds. *PhD thesis*, University of Illinois at Urbana-Champaign.

Gomez, F. (1998) Linking wordnet verb classes to semantic interpretation. *Proc. COLING-ACL Workshop on the Usage of WordNet in NLP Systems*, pp. 58–64. Universite de Montreal, Quebec, Canada.

Gomez, F. and Segami, C. (1989) The recognition and classification of concepts in understanding scientific texts. *J. Experimental and Theoretical Artificial Intelligence* **1**: 51–77.

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

Gorrell, P. (1991) Subcategorization and sentence processing. In: Berwick, R. and Tenny, S. A. C. (editors), *Principle-Based Parsing: Computation and Psycholinguistics*. Kluwer Academic, Dordrecht.

Grimshaw, J. (1990) *Argument Structure*. MIT Press, Cambridge, MA.

Hirst, G. (1992) *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge University Press, New York.

Hull, R. (1997) Acquisition of historical knowledge from encyclopedic texts. *PhD thesis*, University of Central Florida.

Jackendoff, R. (1990) *Semantic Structures*. MIT Press, Cambridge, MA.

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

Katz, J. (1990) *The Metaphysics of Meaning*. MIT Press, Cambridge, MA.

Li, X., Szpakowicz, S. and Matwin, S. (1995) A wordnet-based algorithm for word sense disambiguation. *Proc. IJCAI-95*, pp. 1368–1374. Montreal, Canada.

Marcus, M. (1987) Deterministic parsing and description theory. In: Whitelock, P., Wood, M., Somers, H., Johnson, R. and Bennett, P. (editors), *Linguistic Theory and Computer Applications*. Academic Press.

Marcus, M. and Hindle, D. (1990) Description theory and intonation boundaries. In: Altmann, G. (editor), *Cognitive Models of Speech Processing*. MIT Press.

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

McDonald, D. (1982) Understanding noun compounds. *PhD thesis*, Carnegie-Mellon University.

McRoy, S. (1992) Using multiple sources for word sense discrimination. *Computational Linguistics* **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, 1993. (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.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press

Pustejovsky, J. (1995) *The Generative Lexicon*. MIT Press, Cambridge, MA.
Quine, V. (1960) *Word and Object*. MIT Press, Cambridge, MA.
Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. Longman, New York.
Schank, R. (1975) *Conceptual Information Processing*. North Holland, Amsterdam.
Vendler, Z. (1967) *Linguistics in Philosophy*. Cornell University Press, Ithaca, NY.
Voorhees, E. (1993) Using wordnet to disambiguate word senses for text retrieval. *Proc. 16th Annual Int ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 171–180. Pittsburgh, PA.
Weinberg, A. (1993) Parameters in the theory of sentence processing: Minimal commitment theory goes east. *J. Psycholinguistic Research* **22**(3): 339–364.
Wilks, Y., Guthrie, L., Guthrie, J. and Cowie, J. (1992) Combining weak methods in large-scale text processing. In: Jacobs, P., (editor), *Text-Based Intelligent Systems*. Lawrence Erlbaum.
Yarowsky, D. (1992) Word-sense disambiguation using statistical models of roget's categories trained on large corpora. *Proc. COLING-92*, pp. 454–460.

https://doi.org/10.1017/S1351324900002436 Published online by Cambridge University Press