Catherine MACLEOD, Ralph GRISHMAN, Adam MEYERS, Leslie BARRETT and Ruth REEVES, New York University

NOMLEX: a lexicon of nominalizations

Abstract

NOMLEX (NOMinalization LEXicon) is a dictionary of English nominalizations currently under development at New York University. NOMLEX seeks not only to describe the allowed complements for a nominalization, but also to relate the nominal complements to the arguments of the corresponding verb. We consider both the main verbal arguments (subject, direct object, and indirect object), which may map into a variety of nominal positions, and the oblique verbal complements, which map more directly into nominal complements. The argument correspondences are specified through a combination of explicit information in the lexical entries and general linguistic constraints on the correspondences.

Keywords: nominals, nominalization, lexicon, machine readable dictionary.

1. Introduction

NOMLEX is a dictionary of English nominalizations being developed by the Proteus Project at New York University. In order to aid natural language processing tasks such as information extraction, we felt it was important to have a lexical resource that would relate the arguments of a nominalization to the predicate argument structure of its associated verb. We wanted to capture the fact that the information in a noun phrase like "Rome's destruction of Carthage" is the same as that in the sentence "Rome destroy(ed) Carthage". To that end we have developed a notation for encoding where the verbal subject, object, and indirect object can be found in the noun phrase and what other verbal complements can appear with the nominalization.

2. Background

The Proteus Project was responsible for COMLEX Syntax (Macleod, Meyers and Grishman 1996), and therefore has considerable experience with the development of dictionaries with detailed complement information. Nonetheless, the design of NOMLEX was quite a challenge because there was little precedent for a dictionary of nominalizations. In particular, there were no ready-made classes like those we adapted from the Linguistic String Project at New York University for COMLEX, and the type of notation we used in COMLEX, which is based upon the fixed combinations of elements in fixed word orders, was not appropriate for representing nominalizations. For instance, verbal complements such as NP-PP (that is a Noun Phrase followed by a Prepositional Phrase) when mapped into nominalizations may be divided into pre and post-noun positions. We have gone through several restructurings of the dictionary entries in order to capture these relationships, before arriving at the notation presented here.

3. Methods

We have adapted the menu-based entry program used for entering COMLEX Syntax for NOMLEX. This is coded in Common Lisp and uses the Garnet GUI package. The dictionary entries are being created by two part-time "elves"[1] who have now been working for more than a year. The elves consult various hard-copy dictionaries and the same on-line concordance used in entering COMLEX. This corpus consists of about 100 MB, including most of the Brown Corpus, 27 MB of the Wall Street Journal, 30 MB of the San Jose Mercury, 29.5 MB of the Associated Press and 1.5 MB of miscellaneous literary texts. But the most important resource of all is their knowledge as native speakers of English and linguistics graduate students.

Unfortunately, no word list of nominalizations was readily available, so we have developed our own by searching for common suffixes in lists of nouns from a combination of the Brown Corpus and the Wall Street Journal (about 1 million words of each). These words were ordered by frequency. We doubt that we will be able to do all the nominalizations in COMLEX, so it seemed logical to take samples of the most frequent words first. We started out with nominalizations with the *-ion*, *-ment* and *-er* suffixes, and have now entered *-ee* and *-al* nominalizations and some noun-verb homonyms.[2] We expect to have entered one thousand nominalizations by September, 1998.

4. Notation

Like COMLEX, NOMLEX has a Lisp-like notation, organized as a typed feature structure. A sample dictionary entry for NOMLEX can be seen in Figure 1. The notation for nominalizations is by necessity quite complex. It can be thought of in two parts: one that concerns the main nominal arguments and the main roles of the verb (subject, direct object, and indirect object) and one which deals with all the other complementation of the verb.

4.1. The Relationship of Verb Roles to Noun Positions

The verbal roles of subject, object and indirect object can be realized in the nominal positions of the possessive determiner, the pre-noun noun modifier and the post-noun prepositional phrase (usually headed by "of"). One of the problems we found early on is that the verbal subject and object can sometimes occur in all noun positions, i.e. the subject and object can appear as the possessive determiner (DET-POSS), the pre-noun noun modifier (N-N-MOD) or the post-noun pp (PP-OF). For instance, DET-POS for 'destruction' can be both subject and object:

> (1) "Rome's destruction" [DET-POSS is object]
> (2) "Rome's destruction of Carthage" [DET-POSS is subject]

Clearly a simple statement about what verbal role DET-POSS represents, independent of the other nominal arguments, is not possible. On the other hand, enumerating all the possible combinations of nominal argument positions in each lexical entry, and specifying the correspondence to verbal roles for each combination, would produce an enormous entry for each word. We therefore rely on a combination of information explicitly provided in the

entry, along with various constraints, to establish the correspondence between nominal and verbal arguments.

One such constraint is the uniqueness criterion, which states that all verbal roles must be uniquely filled.[3] In particular, at most one of the nominal positions can map to the verbal object, and at most one to the verbal subject. It follows that if for 'destruction' PP-OF can only be interpreted as the object of 'destroy', then in example (2) above, the DET-POSS can only be interpreted as the subject of 'destroy'. The uniqueness criterion also applies to the nominal positions DET-POSS and PP-OF, blocking such constructs as "Rome's Carthage's destruction", even though both subject and object could independently map into the DET-POSS position.

Figure 1 gives the NOMLEX entry for 'destruction', which is the nominalization of the simple transitive verb, 'destroy'. The VERB-SUBJ feature states that both N-N-MOD and DET-POSS can be the subject of 'destroy'. You can say both "The FMLN's destruction of the town" and "The FMLN destruction of the town". In both cases 'FMLN' is the subject of the act of destroying. The VERB-SUBC feature lists the complements of the corresponding verb (represented by the COMLEX categories prefixed by "NOM-") and provides information about how each verbal complement is realized as a nominal complement. 'Destroy' has one verbal complement, NP, a direct object. The :OBJECT feature indicates that the direct object can appear in all three noun positions, DET-POSS, N-N-MOD, and PP-OF.

We have already noted how the uniqueness condition limits some of the possible correspondences between nominal and verbal arguments. One additional constraint is the word order constraint, that is, we assume that the pre-nominalization arguments appear in the order: subject, object, indirect object, oblique argument. Therefore, if we have the phrase "The White House Secretary of State appointment", the "White House" must be the entity which is appointing the Secretary of State. If we switch the nouns as follows: "The Secretary of State White House appointment", the result is either that the "Secretary of State" is appointing someone or that someone is appointing the "Secretary of State" to the "White House" (in some capacity). It cannot be interpreted as the "White House" appointing anyone.

These constraints do not address the case where only one nominalization argument position is filled. To handle these cases, the dictionary entry may include the features DET-POSS-NO-OTHER-OBJ and N-N-MOD-NO-OTHER-OBJ. DET-POSS-NO-OTHER-OBJ specifies what verbal role is filled by the possessive, in case no other core argument position[4] is filled. For 'destruction' this keyword has the value OBJECT. That means that if no other core argument is present, then the possessive must be the object. This is the case in example (1) above. Thus we can disambiguate this situation. Similarly, N-N-MOD-NO-OTHER-OBJ stipulates what the pre-noun noun modifier must correspond to in the absence of another object. Taken together, the uniqueness criteria, the word order constraint, and the NO-OTHER-OBJ features resolve the ambiguity for all combinations of arguments of 'destruction'.

The nominalization entry in Figure 1 also specifies the type of nominalization with the feature NOM-TYPE. The lexicon identifies four types of nominalizations: *verb-nom*, where the nominalization refers to the action or state of the verb; *verb-part*, which is like verb-nom but incorporates a particle or preposition which is separate in the verbal form (such as 'takeover'); *subj-nom*, where the nominalization refers to the subject of the verb (such as

'teacher'); and *obj-nom*, where the nominalization refers to the object of the verb (such as 'appointee').

```
(nom  :orth "destruction"
      :verb "destroy"
      :plural *none*
      :nom-type((verb-nom))
      :verb-subj ((n-n-mod) (det-poss))
      :det-poss-no-other-obj ((object))
      :n-n-mod-no-other-obj ((object))
      :verb-subc (nom-np :object ((det-poss)(n-n-mod)(pp-of)))))
```

Figure 1: Sample NOMLEX dictionary entry for "destruction".

4.2. The Relationship between Verbal Complements and Nominal Arguments

The relationship between the oblique verbal complements (or subcategorization) and the nominal arguments is much more straight-forward than that for the main verbal roles (subject, object, and indirect object) discussed above. As we have said, we base our NOMLEX subcategorization on the COMLEX verbal complements. These are defined in our COMLEX Syntax Word Class Manual[5] (Wolff, Macleod and Meyers 1998). These symbols are prefixed with NOM for use in NOMLEX. The *translation* of these complements into nominal arguments consists of taking the first NP (if it exists) and treating it as the direct object (in the manner described in the previous section). Thus, a complement such as NP-TO-INF-OC (a Noun Phrase followed by a TO INFinitive whose subject is the object of the matrix verb) is analyzed as a direct object (which can appear as a DET-POSS, N-N-MOD or PP-OF) and a TO-INF. The verb 'disqualify' is classified as having an NP-TO-INF-OC. The sentence, "They disqualified him to run the race." [NP-TO-INF-OC] has a counterpart in the noun phrase "Their disqualification of him to run the race" or "His disqualification to run the race". The first noun phrase shows the object NP appearing in an of-phrase followed by the TO-INF noun argument and the second noun phrase shows the object as the possessive determiner with the to-infinitive still appearing in the post-noun position. This demonstrates that while the main verbal arguments may roam around and even change part of speech, the oblique complements occur virtually unchanged (that is, a to-infinitive is still a to-infinitive) and are always post-noun (the object, of course, may precede).

Figure 2 shows the NOMLEX entry for 'promotion' and Figure 3 shows the COMLEX entry of its associated verb 'promote'. This verb is far more complex and as a result the nominalization entry is richer. The NOM-TYPE is again VERB-NOM. This class allows the richest complementation for the nominalization. The verbal subcategorization is listed under VERB-SUBC. Examples of these complements are shown in Figure 4. For each VERB-SUBC the default subject position is given on the VERB-SUBJ list. This can be overwritten by a SUBJECT keyword under the particular subc. The object is always listed under the subc because, if it occurs, it is part of the COMLEX complement. Note that for NOM-NP-AS-NP there is no interpretation of the N-N-MOD position as the object, even though the object corresponding to the verbal complement NP is allowed in that position. It seems to be a fact that NP-AS-NP complements in general do not allow the verbal object to map to the N-N-MOD position.

```
(nom   :orth "promotion"
       :verb "promote"
       :nom-type((verb-nom))
       :verb-subj ((n-n-mod) (det-poss))
       :verb-subc ((nom-np :object ((det-poss)(n-n-mod)(pp-of)))
                   (nom-np-as-np :object ((det-poss) (pp-of)))
                   (nom-possing :nom-subc ((p-possing :pval ("of"))))
                   (nom-np-pp :object ((det-poss) (n-n-mod) (pp-of))
                      :pval ("into" "from" "for" "to"))
                   (nom-np-pp-pp :object ((det-poss) (n-n-mod) (pp-of))
                                 :pval ("for" "into" "to") :pval2 ("from"))))
```

Figure 2: Sample NOMLEX dictionary entry for "promotion".

```
(verb  :orth "promote"      ((np-pp :pval ("to" "for" "into" "from"))
                             (np-pp-pp :pval ("for" "to" "into" "from"))
                             (possing)(np)(np-as-np)(np-tobe)))
```

Figure 3: Sample COMLEX Syntax verb entry.

The nominalization complement is, by default, the same as the verbal complement (without the object NP). For instance, in NP-AS-NP, the realization of the direct object is specified explicitly and, by default, the remainder of the verbal complement, an AS-NP, is realized by the same structure in the nominalization. However, this direct *translation* sometimes does not occur. In that case, the nominal argument is given (prefaced by the key-word NOM-SUBC). In Figure 2, you can see NOM-POSSING where the NOM-SUBC is not POSSING but P-POSSING with the preposition "of" required. An example of this is, "He promoted working late on Fridays" which corresponds to "His promotion of working late on Fridays". Another difference between the nominal and verbal complement notation is the distinction made in the prepositions for the complements with two preposition phrases (PP-PP, for instance). In COMLEX, mostly for historical reasons, there is only one pval list. For example, PP-PP for 'agree' has a preposition list of ("among" "about" "between" "on" "upon" "with") with no indication of co-occurrence constraints. You cannot agree "among" yourselves "between" something; but you can agree among or between yourselves about/on/upon something. We do make these distinctions for NOMLEX, therefore PP-PP complements have two pval lists PVAL and PVAL2 (PVAL1 is used elsewhere). The lists are not ordered, reflecting the fact that PP's are very mobile, i.e. you can say, "Their promotion of her from assistant manager to manager" or "Their promotion of her to manager from assistant manager".

Verb-Subc	Example Sentence
np-as-np	They promoted him as a manager.
np-tobe	They promoted her to be liaison officer.
np-pp	They promoted him to chairman.
np-pp-pp	They promoted him to manager from assistant manager.
np	They promoted excellence in scholarship.
possing	They promoted (his) studying long and hard.

Figure 4: Examples of COMLEX Syntax Verbal Complements

5. Conclusion

Ever since Chomsky's *Remarks on Nominalization* (Chomsky 1970), the correspondence between nominal and verbal forms has been generally treated as being outside the central concerns of linguistics, and so — while there have been some studies of nominalization — the topic has received much less attention than other linguistic phenomena. Our primary goal has been to create a broad-coverage resource of value for natural language processing. However, we hope that this resource will also allow us and others to explore more systematically the relations between nominal and verbal usages, and so enhance our understanding of nominalizations. We hope to release an alpha version of NOMLEX in the Fall of 1998. This dictionary will be freely available from our web site.

Acknowledgments

6. Notes

[1]   The linguistics graduate students who worked on COMLEX and now on NOMLEX are referred to as Enterers of Lexical Features, hence ELFs or elves.

[2]   Surprisingly few of the latter are really nominalizations; consider "sight" as in "He sighted the ship". The homograph noun is not a nominalization. You cannot say "His sight of the ship", but must instead use the gerund, "His sighting of the ship".

[3]   We base our uniqueness criteria on the stratal uniqueness theorem of Relational Grammar and Arc-Pair Grammar (see for example Perlmutter 1984 and Johnson 1980).

[4]   Core arguments include all verbal arguments except oblique complements.

[5]   Available through the COMLEX WWW page: http://cs.nyu.edu/cs/projects/proteus/comlex.

7. References

Chomsky, Noam (1970). Remarks on Nominalization. *Studies on Semantics in Generative Grammar*. Mouton, The Hague.

Johnson, David E. and Paul M. Postal (1980). *Arc Pair Grammar*. Princeton University Press, Princeton.

Macleod, Catherine, Adam Meyers and Ralph Grishman (1996). COMLEX Syntax: An On-Line Dictionary for Natural Language Processing. In *Proceedings of Euralex96*: pp. 131—140.

Perlmutter, David. M. (1984). *Studies in Relational Grammar 1*. The University of Chicago Press, Chicago.

Wolff, Susanne Rohen, Catherine Macleod and Adam Meyers (1998). Comlex Word Classes Manual, Proteus Project, New York University.