

# Lexical Annotation for Multi-word Entries Containing Nominalizations

Catherine Macleod

Computer Science Dept.  
New York University  
715 Broadway, 7th Floor  
New York, NY 10003-6806  
macleod@cs.nyu.edu

## Abstract

New York University has produced a dictionary of nominalizations (NOMLEX) whose entries capture the relationship of the nominalization with its associated verb. This dictionary indicates where the verbal arguments may be found in the noun phrase which contains the nominalization. We have now made a study and produced some entries for nominalizations and their co-occurring verbs. These entries are much more complex than NOMLEX entries. In order to express all the relationships between the nominalization and its co-occurring verb, we made use of the terminology of Igor Mel'čuk, whose theories have been used to create dictionaries in French and Russian. His categories were found to be very useful for this task. The verb + nominalization pairs were selected by frequency of co-occurrence and thus do not strictly conform to what are considered support verbs. Support verbs are generally defined as having no semantic content, serving only to carry tense and number which the nominalization cannot express. A typical example of this is "commit a murder". The paper below describes the NOMLEX entry which is the basis of this work and then demonstrates the additional information needed to describe the verb + nominalization pair.

## 1. Introduction

New York University (NYU) developed NOMLEX, a dictionary containing 1,000 entries of nominalizations, with detailed information about their arguments. This dictionary is publicly available at the NYU Proteus website: <http://cs.nyu.edu/cs/projects/proteus/nomlex/index.html>. Given the frequency with which support verbs combine with nominalizations in ways that can only be interpreted by specific lexical information, we have developed an extension to NOMLEX to cover support verbs (and other co-occurring verbs). Support verbs are those verbs which are defined as having no semantic content, serving only to carry tense and number as Maurice Gross defines it (Gross, 1981) p.17. The following examples demonstrate this generally accepted view of support verbs: "he took a walk" *he walked*; "they made a decision" *they decided*; "she made a discovery" *she discovered*. Below we will describe how the information is captured in our dictionaries. Since the nominalization support verb entries are based on the original NOMLEX entries we will briefly outline what is contained in those entries.

## 2. NOMLEX Entries

The notation that had been used for NOMLEX (Macleod et al., 1997) was a quite complicated two part entry with the main arguments of the associated verb being handled differently from the oblique complements. The "main" arguments of the verb (subject, object, direct object) may occur as possessive pronouns, pre-noun noun modifiers, and post-noun prepositional phrases (most often with "of"). The oblique complements of the verb occur post-noun either unchanged or introduced by a preposition. Figure 1 is a typical NOMLEX entry. In this particular entry, the subject of the verb (VERB-SUBJ) may appear as a prepositional phrase where the preposition is "from" ("the inquiry from Bob"/ Bob inquires), as a noun-noun modifier ("the White House inquiry"/the White House inquires) or

a possessive determiner ("Mary's inquiry"/Mary inquires). The type of this particular nominalization is VERB-NOM which refers to the fact some nominalizations behave very verbally and can appear with many if not all of the verbal complements. Other possible NOM-TYPES are (Subject) and (Object); these have one of the verbal arguments absorbed by the nominalization. For example, "teacher" (Subject) and "draftee" (Object).

In the entry, information is also given on the complements of the associated verb which can occur also with the nominalization. The verbal information comes from COMLEX, a large syntactic dictionary, also created by the Proteus group at NYU (Macleod et al., 1998) and (Grishman et al., 1994). The complements are prefixed by "NOM" to make clear that these are now being treated as nominalization complements. The complement NOM-PP-P-WH-S consists of a prepositional phrase (PP) followed by a preposition (P), a wh-word (WH) and a sentence (S). This complement occurs unchanged on both the verb and the nominalization, as in, "John inquired of his tour guide about whether they would arrive on time." and "John's inquiry of his tour guide about whether they would arrive on time [sounded whiney]".

## 3. Complexity of the multi-word entry

The multi-word nominalization entries were based on the NOMLEX entries and thus also on the COMLEX verb entries. There was an attempt made to adapt the original NOMLEX notation for the multi-word entries, but it became very dense and difficult both to enter and to read. The added presence of the support verb made the entry much more complex. Information was now needed to identify the role the nominalization plays with regard to the support verb. It can occur as subject or object of the support verb or as an oblique object where the nominalization occurs in a prepositional phrase; see Table 1.

It was also necessary to state whether the verb was a

```

(NOM :ORTH "inquiry"
:VERB "inquire"
:PLURAL "inquiries"
:SPELLING "American"
:PLURAL-FREQ "not rare"
:NOM-TYPE ((VERB-NOM))
:VERB-SUBJ ((PP :PVAL ("from"))
(N-N-MOD)
(DET-POSS))
:SUBJ-ATTRIBUTE ((COMMUNICATOR))
:VERB-SUBC ((NOM-P-WH-S :PVAL ("about"))
(NOM-PP-P-WH-S :SUBJECT ((DET-POSS)
(N-N-MOD))
:PVAL ("about")
:PVAL1 ("of" "from"))
(NOM-P-POSSING :PVAL ("about"))
(NOM-PP-PP :SUBJECT ((DET-POSS)
(N-N-MOD))
:PVAL ("about")
:PVAL2 ("from" "of"))
(NOM-PP :PVAL ("after"
"about"
"into"))
(NOM-INTRANS :SUBJECT ((DET-POSS)
(PP :PVAL ("from"))))))

```

Figure 1: NOMLEX Entry

Role	Example
Subject:	The <i>destruction</i> took place last year
Object:	He took a long <i>walk</i>
Oblique:	They accepted her for <i>admission</i>

Table 1: Nominalization Roles

“true” support verb (like “make” in “make a decision”) or whether it had some semantic content (as in “bring” “they brought an action against him” where it has a temporal aspect). In searching for a better type of notation we settled on Mel’čuk’s lexical function notation, which suited our purpose admirably.

#### 4. Multi-word entries using Mel’čuk’s lexical functions

For our multi-word entries we took two parts of Mel’čuk’s notation (Mel’čuk, 1996) and (Mel’čuk, 1988), his deep syntactic representation and some of his functions. We augmented these as necessary, from Thierry Fontenelle’s work (Fontenelle, 1995) and our own.

Mel’čuk’s syntactic representation seemed particularly adapted for use in expanding NOMLEX coverage to nominalization phrases. First his terminology indicates what syntactic relationship the nominalization has with the co-occurring verb. These are *oper*(i), *func*(i) and *labor*(i,j). *Oper*(i) specifies that the nominalization occur as the direct object of the verb. A subscript (i=1,2) indicates whether the subject of the main verb is also the subject of the nominal-

ization or the object of the nominalization. For example, in the sentence “He paid a visit to Jane.” the combination “pay” “(a) visit” would be marked “oper(1)”. “He” is the subject of both “pay” and “visit”, i.e. “he visited”. On the other hand, “He had a visit from Jane” is *oper*(2), since the subject of the main verb is the object of “visit”, i.e. Jane visits him. These verbs (“pay” and “have”) are true support verbs, in that they share at least one argument with the nominalization, and they express no semantic content in themselves. They serve only to express number and tense. In our notation, we refer to these verbs as “transp” because they are transparent as to semantics.

*Func*(i) describes a nominalization that occurs as the subject of the main verb. Examples: “the accusation came from John.” is *func*(1): John accuses (someone). “The accusation names John.” is *func*(2): (Someone) accuses John.

*Labor*(i,j) is assigned to a main verb plus nominalization where the nominalization occurs as the oblique complement of the main verb. Because of this, there are two subscripts to stand for the main verb subject position (i) and the main verb object position (j). For example, “He subjected John to an interrogation.” represents *labor*(1,2). The subject of the main verb is “He” which is the subject of “interrogate”, the object of the main verb is “John” which is also the object of “interrogate”. The sentence can be interpreted as “He interrogated John”.

We added another term to describe the reciprocal relationship, *operrecip*(i+i). “John and Mary traded insults.” would be marked (*operrecip*(1+1)) that is, “Mary and John

insulted each other.”

Other Mel’čuk functions that we made use of in our notation (besides Transp) can be seen in Table 2.

We added or changed functions like the expansion of Mel’čuk’s artifact into three different classes, see Table 2 at the end.

If the entry has no lex-func listed, it is unclassified, not transparent. For example, “approve” and “license” in the entry for “authorization” are listed simply as (oper1) with no other function.

Although this is still a syntactic dictionary, we were forced to take sense distinctions into consideration, since the selection of support or co-occurring verbs is dependent on the senses of the nominalization. “Admission” meaning “allowing entry to a place”, co-occurs with verbs like “offer”, “grant”, “deny”, “refuse”, “accept”; while the “admission” meaning “make a statement that some fact is true”, selects the support verb “make” and the collocate verb plus preposition “come from”. This is an interesting syntactic test for the division of senses. It is “real” because the selection of the verb forces one of two different readings of the noun.

## 5. NOMLEX2 Entries

We analyzed 20 nominalizations completely and they are the basis for what we call our NOMLEX2 lexicon. We hope over time to expand our coverage. This project was mainly to ascertain whether a lexicon of this sort could be made. Figure 2 shows a NOMLEX2 entry that is typical of the multi-word “support” plus nominalization entries. The significant addition to NOMLEX are the support verbs (SUPP-V). As you can see, the particular support verb (SVERB#) can be one verb as in the transp example “make” or a list of similar verbs as in the artifact2 example “deny” “reject” “contest” “fight” etc. Each SVERB (or group) has a LEX-FUNC label with the exception noted above. The LEX-FUNC establishes the role that the nominalization plays in regard to the support verb. The subcategory defines whether the SVERB is transparent or has some “meaning” connected with it.

## 6. Method

We made a selection of verbs which co-occurred with the particular nominalization frequently. We used Dekang Lin’s Collocation (Dependency) Database (Lin, 1995) which provided us with frequency information on the support plus nominalization pairs. Had the database been larger we would have raised the minimum frequency. As it is, some verbs were allowed that probably would not have been considered with a larger data base. We erred on the side of over-selection of co-occurring verbs, reasoning that we could always make finer cuts and eliminate certain verbs but it would be much more complicated to add to the entries afterwards. We also consulted the British National Corpus, and a combination of the Brown and Wall Street Journal corpus.

## 7. Conclusion

Mel’čuk’s classes have, to our knowledge, been applied to lexicons in French (Mel’čuk, 1984) and Russian

(Mel’čuk and Xolodovič, 1984). We show that they are very useful when adapted to English. We chose a group of nominalizations from NOMLEX. We have done 20 nominalizations, of which 8 were divided for different senses. Therefore, in the nominalization with support verbs dictionary NOMLEX2, we have 28 entries containing 158 support verbs or verb groups. We found that this type of entry was slow but replicable. Three different lexicographers worked on the entries and they were in agreement most of the time on how to assign the classes. Therefore, we feel that the decision to use Mel’čuk’s work was a good one and we have seen that the creation of a multi-word dictionary is a doable project.

## 8. Acknowledgement

This research was supported in part by the National Science Foundation under Grant 99-82069 from the Division of Information and Intelligent Systems.

## 9. References

- Thierry Fontenelle. 1995. *Turning a bilingual dictionary into a lexical-semantic database*. Ph.D. thesis, University of Liège, Belgium.
- Ralph Grishman, Catherine Macleod, and Adam Meyers. 1994. COMLEX Syntax: Building a Computational Lexicon. In *Proceedings of Coling 1994: The 15th International Conference on Computational Linguistics*, pages 268–272.
- Maurice Gross. 1981. Les bases empiriques de la notion de prédicat sémantique. In A. Guillet and C. Leclère, editors, *Formes Syntaxiques et Prédicat Sémantiques*, volume 63 of *Langages*, pages 7–52. Larousse, Paris.
- Dekang Lin. 1995. A Dependency-based Method for Evaluating Broad-Coverage Parsers. In *IJCAI-95*.
- Catherine Macleod, Adam Meyers, Ralph Grishman, Leslie Barrett, and Ruth Reeves. 1997. Designing a Dictionary of Derived Nominals. In *Proceedings of Recent Advances in Natural Language Processing*, Tzgov Chark, Bulgaria.
- Catherine Macleod, Ralph Grishman, and Adam Meyers. 1998. COMLEX Syntax. *Computers and the Humanities*, 31(6):459–481.
- Igor’ A. Mel’čuk and Aleksandr A. Xolodovič. 1984. *Tolkovo-kombinatornyi Slovar’ Sovremenogo Russkogo Jazyka*. Opyty Semantiko-Sintaksičeskogo opisanija Russkoj Leksiki. WSA, Vienna. Translation: Explanatory Combinatory Dictionary of Modern Russian. Semantico-Syntactic Studies of Russian Vocabulary.
- Igor’ A. Mel’čuk. 1984. *Dictionnaire explicatif et combinatoire du français contemporain*. Les Presses de l’Université de Montréal, Montréal.
- Igor’ A. Mel’čuk. 1988. *Dependency Syntax: Theory and Practice*. State University Press of New York, Albany.
- Igor’ A. Mel’čuk. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins Publishing Company, Amsterdam.

```

(NOM :ORTH "accusation"
:VERB "accuse"
:PLURAL "accusations"
:PLURAL-FREQ "not rare"
:NOM-TYPE ((VERB-NOM))
:NOUN-SUBC ((NOUN-PP :PVAL ("about")))
:VERB-SUBJ ((DET-POSS)
(N-N-MOD))
:SUBJ-ATTRIBUTE ((COMMUNICATOR))
:OBJ-ATTRIBUTE ((COMMUNICATOR))
:N-N-MOD-NO-OTHER-OBJ ((SUBJECT))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))
(P-P-OF)))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))
:SUPP-V
((SVERB1 :ORTH "make"
:LEX-FUNC ((OPER1 :TRANSP T))
:REQUIRED ((NOM-DET :INDEF T))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))
(SVERB2 :ORTH "level, lob, fling, hurl"
:LEX-FUNC ((OPER1 :TRANSP T))
:REQUIRED ((NOM-DET :INDEF T))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against" "at"))))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))
(SVERB3 :ORTH "come forward with, bring"
:LEX-FUNC ((OPER1 :INCEP T))
:REQUIRED ((NOM-DET))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))
(SVERB4 :ORTH "fabricate"
:LEX-FUNC ((OPER1 :ANTIFACT0 T))
:REQUIRED ((NOM-DET))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))
(SVERB5 :ORTH "press, renew, revive"
:LEX-FUNC ((OPER1 :CONT T))
:REQUIRED ((NOM-DET :DEF T))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))

```

```

(SVERB6 :ORTH "disprove"
:LEX-FUNC ((OPER2 :ANTIFACT0 T))
:REQUIRED ((NOM-DET :DEF T))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))

(SVERB7
:ORTH "deny, laugh off, reject, dispute, contest, rebut, counter, fight"
:LEX-FUNC ((OPER2 :ANTIFACT2 T))
:REQUIRED ((NOM-DET :DEF T))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))

(SVERB8 :ORTH "respond to, answer, address, face"
:LEX-FUNC ((OPER2 :TRANSP T))
:REQUIRED ((NOM-DET :DEF T))
:VERB-SUBC ((NOM-NP :OBJECT ((PP :PVAL ("against"))))
(NOM-NP-PP :OBJECT ((PP :PVAL ("against"))
:PVAL ("of")))
(NOM-NP-P-ING-OC :OBJECT ((PP :PVAL ("against"))
:PVAL ("of"))))

(SVERB9 :ORTH "admit"
:LEX-FUNC ((LABOR2-0 :PVAL ("to")
:FACT2 T))
:REQUIRED ((NOM-DET :DEF T))

(SVERB10 :ORTH "smart from"
:LEX-FUNC ((OPER2 :REAL T))
:REQUIRED ((NOM-DET :DEF T))

(SVERB11 :ORTH "withdraw, drop, recant"
:LEX-FUNC ((OPER1 :ANTIFACT1 T))
:REQUIRED ((NOM-DET :DEF T))

(SVERB12 :ORTH "avoid, escape"
:LEX-FUNC ((OPER2))
:REQUIRED ((NOM-DET :DEF T))

(SVERB13 :ORTH "come from"
:LEX-FUNC ((FUNC1 :INCEP T))
:REQUIRED ((NOM-DET :DEF T))

(SVERB14 :ORTH "exchange, trade"
:LEX-FUNC ((RECIP :TRANSP T))

(SVERB15 :ORTH "clear"
:LEX-FUNC ((LABOR0-2 :PVAL ("of")
:ANTIREAL T))
:REQUIRED ((NOM-DET :DEF T))

(SVERB16 :ORTH "arraign"
:LEX-FUNC ((LABOR0-2 :PVAL ("on")))
:REQUIRED ((NOM-DET))))

```

Figure 2: NOMLEX2 Entry

<b>Notation</b>	<b>Function</b>	<b>Examples</b>
Fin	stop	<i>They completed the acquisition.</i>
Liqu	end completely	<i>They finalized their divorce.</i>
Real	expected outcome realized	<i>They approved his application.</i>
Antireal	expected outcome not realized	<i>They turned down his application. He withdrew his application.</i>
Cont	continuous	<i>They kept up their attack.</i>
Incep	beginning	<i>They started their attack at dawn.</i>
Orig	originated “came from”	<i>The accusation came from his firm.</i>
Not	negative (didn't happen)	<i>They postponed the attack.</i>
Antifact1	subject declares contents untrue	<i>He withdrew the accusation.</i>
Antifact2	object declares untrue	<i>He denies the charges.</i>
Antifact0	nom is untrue	<i>He fabricated the accusation.</i>

Table 2: Mel'çuk Functions