Nominalized Clauses in the Grammar Matrix

Kristen Howell, Olga Zamaraeva and Emily M. Bender

Department of Linguistics, University of Washington, Seattle, WA, U.S.A.

{kphowell, olzama, ebender}@uw.edu

Abstract

We present an analysis of clausal nominalization, developed for the LinGO Grammar Matrix. In particular we examine the typological variation of nominalized clausal complements and adverbial clausal modifiers, and propose three analyses of nominalization based on the level at which nominalization occurs in the clause. We introduce a new feature NMZ and a combination of lexical and unary rules to account for nominalization at the V, VP and S levels.

1 Introduction

The LinGO Grammar Matrix (Bender *et al.*, 2002, 2010) is a starter kit for creating broad-coverage implemented precision grammars. It includes a shared core grammar as well as a series of libraries extending that core with analyses for cross-linguistically variable phenomena. It can also be seen as a means of cross-linguistic hypothesis testing, where Matrix library developers seek to develop analyses of particular phenomena that work across the full range of the world's languages (as understood from the typological literature) and then test those analyses by using the resulting system to build grammar fragments for languages not directly considered during development. The present work is motivated by our efforts to add libraries for embedded clauses, as complements to certain types of verbs, on the one hand, or as modifiers of verbal projections on the other. Typological surveys of these phenomena including Noonan (2007) and Thompson *et al.* (2007) show that clausal nominalization is a common strategy for embedded clauses.

In this abstract, we present an overview of our analysis of nominalized subordinate clauses and its implementation in the Grammar Matrix. This analysis is intended to cover the range of variation we find in the world's languages[1] while also providing a certain degree of analytical freedom to users of the Grammar Matrix. In particular, we allow for nominalization at different levels in the parse tree: low (where a lexical V becomes a nominal constituent), mid (where a VP does), and high (where the nominalized constituent is an S). We also provide options on the semantic side, using Minimal Recursion Semantics (MRS) (Copestake *et al.*, 2005), allowing high nominalization to be either a strictly syntactic phenomenon or one with semantic effects. A linguist using the customization system can test alternative analyses against text from their language to explore which fits best.

2 Nominalized Subordinate Clauses

Malouf (2000) presents a hierarchy of nominalization types for increasingly nominal properties of the phrase's internal distribution, as demonstrated in (1)–(5).

---

[1]Note that we differentiate between nominalized clauses and deverbal nouns, leaving the latter for future work.

(1) The DA was shocked that Pat illegally destroyed the evidence.

(2) The DA was shocked by Pat having illegally destroyed the evidence.

(3) The DA was shocked by Pat's having illegally destroyed the evidence.

(4) The DA was shocked by Pat's illegal destroying of the evidence.

(5) The DA was shocked by Pat's illegal destruction of the evidence

Malouf notes that (1) has no internal properties of an NP, as it has an adverbial modifier and a direct case subject (one that has not been changed), while (5) on the other hand has all the internal properties of an NP, including an adjectival modifier and both a subject and object whose case differs from those in verbal projections. Malouf argues against "change-over" approaches (e.g. that of Lapointe 1993), inter alia, because they don't constrain what kinds of change-overs are available. His mixed-category approach, combined with language-specific versions of the Head-Specifier and Head-Subject rules, elegantly accounts the mixed behavior of verbal gerunds. However, given the goal of grammar customization and the context of the Grammar Matrix code base, we take a change-over approach as it integrates more easily with the other libraries providing the phrase structure rules. On our analysis, the change-over can happen at the S, VP or V levels. High nominalization (at S, as in (2)) allows adverbial modifiers and does not allow case change on subjects; mid nominalization (at VP, as in (3)) allows adverbial modifiers and only allows case change on subjects; and low nominalization (at V, as in (4)) allows adjectival modifiers and case change on both subjects and objects.[2]

3 Analysis

In this section we present three distinct analyses for nominalization to account for the data in (§2). We begin with the addition of an NMZ feature in (§3.1). This is followed by a description of the three analyses high (§3.2), mid (§3.3) and low (§3.4). We discuss the additional work necessary to accommodate case frame changes in (§3.5) and propose two possible semantic representations of nominalized clauses in (§3.6).

3.1 The feature NMZ

Our analysis allows for the disassociation of the nominalizing morphology from the actual change of the HEAD value from *verb* to *noun*. To facilitate this, we propose a boolean HEAD feature NMZ, which we use to distinguish verbs inflected with a nominalization morpheme (but not yet nominalized) from other verbs. We also use it to differentiate between nominal constituents built from nominalized verbs and other (lexical) nouns. Nouns and verbs in the lexicon are constrained to be [NMZ −] and this constraint is changed to [NMZ +] only by nominalization lexical rules. The low nominalization analysis changes the verb to a noun in the lexical rule. However, the mid and high nominalization analyses employ a unary rule to change the HEAD value from *verb* to *noun* and that unary rule has [NMZ +] on both the daughter and mother. Complementizers, subordinators, and clausal verbs that require nominalized clausal complements constrain their complement to be both [NMZ +] to prevent selection of a lexical noun and [HEAD *noun*] to prevent selection of a verb that has gone through the lexical rule, but not the corresponding unary rule.

3.2 High Nominalization

Our first nominalization analysis involves the nominalization of an S whose valence features are fully saturated. This is accomplished by a lexical rule that puts a morpho-syntactic marker on the verb and a unary rule that changes the verbal constituent into a nominalized clause. The lexical rule, defined in (6),

---

[2]Neither (1) nor (5) involve nominalization of the type we are concerned with; the former because the constituent is verbal at all levels, and the latter because the clause has no verbal properties.

adds [NMZ +] to the mother and identifies the remaining features between the daughter and mother.[3]

(6)
$$
\begin{bmatrix}
\textit{high-or-mid-nominalization-lex-rule} \\[4pt]
\text{SYNSEM | LOCAL} \quad
\begin{bmatrix}
\text{CAT | HEAD} & \begin{bmatrix} \textit{verb} \\ \text{NMZ} & + \end{bmatrix} \\[10pt]
\text{CONT | HOOK | INDEX} & \textit{event}
\end{bmatrix} \\[18pt]
\text{DTR | SYNSEM | LOCAL} \quad \begin{bmatrix} \text{CAT | HEAD} & [\textit{verb}] \end{bmatrix}
\end{bmatrix}
$$

Once the morpho-syntactic marker has been added to the verb and its valence requirements have been satisfied, it can be selected by the *nominalized-clause-phrase* unary rule, defined in (7). Here we constrain the SUBJ and COMPS lists to be empty, ensuring that only a sentence can go through this rule.[4]

(7)
$$
\begin{bmatrix}
\textit{nominalized-clause-phrase} \\[4pt]
\text{SYNSEM | LOCAL} \quad \begin{bmatrix} \text{CAT | HEAD} & \begin{bmatrix} \textit{noun} \\ \text{NMZ} & + \end{bmatrix} \end{bmatrix} \\[20pt]
\text{ARGS} \quad \left\langle
\begin{bmatrix}
\text{SYNSEM | LOCAL} &
\begin{bmatrix}
\text{CAT} &
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{NMZ +} \end{bmatrix} \\[8pt]
\text{VAL} & \begin{bmatrix} \text{COMPS} & \langle\,\rangle \\ \text{SUBJ} & \langle\,\rangle \end{bmatrix}
\end{bmatrix} \\[20pt]
\text{CONT} & \begin{bmatrix} \text{HOOK | LTOP} & \boxed{0} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\right\rangle \\[40pt]
\text{C-CONT} \quad
\begin{bmatrix}
\text{RELS} & \left\langle ! \begin{bmatrix}
\text{PRED} & \text{nominalization\_rel} \\
\text{LBL} & \boxed{1} \\
\text{ARG0} & \boxed{1} \\
\text{ARG1} & \boxed{0}
\end{bmatrix} ! \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

In addition to changing the verb to a noun syntactically, the non-branching rule also adds nominalization_rel to the MRS and identifies its first argument with the daughter, creating an identity between the verb and the nominalized clause.

3.3 Mid Nominalization

The analysis for mid nominalization is very similar to that for high nominalization in that a lexical rule adds a morpho-syntactic marker to the verb and a unary rule nominalizes the verbal constituent higher in the tree. If the case of the subject is not changed, we make use of the same lexical rule as in high nominalization: *high-or-mid-nominalization-lexical-rule* (6). The subject and object are already identified between the mother and daughter of that rule, so we simply create a subtype whose subject list is also [SUBJ ⟨[ ]⟩]. If on the other hand, case change on the subject is possible, we define a separate *mid-nominalization-lex-rule* which is similar to the *high-or-mid-nominalization-lexical-rule* except that the SUBJ list is constrained to be non-empty but only the subject's INDEX and not the whole SUBJ list is identified.

---

[3]The AVMs shown in this abstract are abbreviated in order to focus on features of interest. The lexical rules produced by the Grammar Matrix customization system also have many constraints that serve to copy information from daughter to mother. In this abstract the reader can assume that all features are copied from daughter to mother unless otherwise specified.

[4]The implemented analysis includes qeq identities in HCONS for (7) and (8), in order to produce our intended target representations in Minimal Recursion Semantics. This has been suppressed in the abstract, as it is not essential to the presentation of the analysis.

3.4 Low Nominalization

Finally under our low nominalization analysis, the verb is nominalized in the lexical rule that licenses the nominalization morpheme. The low nominalization lexical rule, defined in (8), changes the HEAD to type *noun* and adds the necessary semantic constraints for the nominalized verb to be represented as a noun. This is accomplished by adding nominalization_rel to the MRS, and identifying that predication to make the verb a semantic argument of the nominalized clause. The resulting MRS representation will be discussed in more detail in (§3.6).

(8)
$$
\begin{bmatrix}
\textit{low-nominalization-lex-rule} \\[4pt]
\text{SYNSEM | LOCAL} \quad
\begin{bmatrix}
\text{CAT} \quad
\begin{bmatrix}
\text{HEAD} \begin{bmatrix} \textit{noun} \\ \text{NMZ} \quad + \end{bmatrix} \\
\text{VAL | SUBJ | INDEX} \quad \boxed{0}
\end{bmatrix}
\end{bmatrix} \\[20pt]
\text{DTR | SYNSEM | LOCAL} \quad
\begin{bmatrix}
\text{CAT} \quad \begin{bmatrix} \text{VAL | SUBJ} \langle\, \text{INDEX} \quad \boxed{0} \,\rangle \end{bmatrix} \\
\text{CONT} \quad \begin{bmatrix} \text{HOOK | LTOP} \quad \boxed{1} \end{bmatrix}
\end{bmatrix} \\[20pt]
\text{C-CONT} \quad
\begin{bmatrix}
\text{RELS} \quad \left\langle\, ! \,
\begin{bmatrix}
\text{PRED} \quad \text{nominalization\_rel} \\
\text{LBL} \quad \boxed{2} \\
\text{ARG0} \quad \boxed{2} \\
\text{ARG1} \quad \boxed{1}
\end{bmatrix}
\, ! \,\right\rangle
\end{bmatrix}
\end{bmatrix}
$$

The rule in (8) identifies the INDEX of the mother's subject with the index of the daughter's. If the case on the subject changes upon nominalization, this constraint is sufficient (in combination with constraints on case discussed in §3.5 below). However, if case change does not occur, the entire subject must be identified. Similarly, if the case on the object changes, a constraint to identify its INDEX is added, whereas if the object's case does not change or the verb is intransitive, the entire complements list is identified.

3.5 Accommodating Case Frame Changes

In §3.3 and §3.4 we describe separate rules for low and mid nominalization when the case frame is changed that do not identify an entire subject or object, but merely the INDEX so that changes can be made to HEAD. The case associated with a particular lexical rule can then be added as a CASE constraint to either the subject or the object as well as the HEAD type corresponding to that case.

3.6 Semantics of Nominalization

For an example like the Turkish sentence in (9) with a nominalized clausal complement, the analyses described above result in an MRS semantic representation in (10):

(9) senin sinema-ya gel-me-n-i isti-yor-um
  2SG.GEN cinema-DAT come-NMZ-2SG-ACC want-PROG-1SG
  "I want you to come to the movies." [tur] adapted from Kornfilt (1997, p. 48)

(10)
$$
\begin{bmatrix}
\textit{mrs} \\
\text{LTOP} \quad h1 \\
\text{INDEX} \quad e2 \\[6pt]
\text{RELS} \quad \left\langle
\begin{bmatrix} \textit{\_you\_n\_rel} \\ \text{LBL} \quad h3 \\ \text{ARG0} \quad x4 \end{bmatrix},
\begin{bmatrix} \textit{exist\_q\_rel} \\ \text{LBL} \quad h5 \\ \text{ARG0} \quad x4 \\ \text{RSTR} \quad h6 \\ \text{BODY} \quad h7 \end{bmatrix},
\begin{bmatrix} \textit{\_cinema\_n\_rel} \\ \text{LBL} \quad h8 \\ \text{ARG0} \quad x9 \end{bmatrix},
\begin{bmatrix} \textit{exist\_q\_rel} \\ \text{LBL} \quad h10 \\ \text{ARG0} \quad x9 \\ \text{RSTR} \quad h11 \\ \text{BODY} \quad h12 \end{bmatrix},
\begin{bmatrix} \textit{\_come\_v\_rel} \\ \text{LBL} \quad h13 \\ \text{ARG0} \quad e14 \\ \text{ARG1} \quad x4 \\ \text{ARG2} \quad x9 \end{bmatrix},
\begin{bmatrix} \textit{nominalization\_rel} \\ \text{LBL} \quad h15 \\ \text{ARG0} \quad x17 \\ \text{ARG1} \quad h16 \end{bmatrix},
\begin{bmatrix} \textit{exist\_q\_rel} \\ \text{LBL} \quad h18 \\ \text{ARG0} \quad x17 \\ \text{RSTR} \quad h19 \\ \text{BODY} \quad h20 \end{bmatrix},
\begin{bmatrix} \textit{\_want\_v\_rel} \\ \text{LBL} \quad h21 \\ \text{ARG0} \quad e2 \\ \text{ARG1} \quad x22 \\ \text{ARG2} \quad x17 \end{bmatrix}
\right\rangle \\[6pt]
\text{HCONS} \quad \left\langle
\begin{bmatrix} \textit{qeq} \\ \text{HARG} \quad h6 \\ \text{LARG} \quad h3 \end{bmatrix},
\begin{bmatrix} \textit{qeq} \\ \text{HARG} \quad h11 \\ \text{LARG} \quad h8 \end{bmatrix},
\begin{bmatrix} \textit{qeq} \\ \text{HARG} \quad h16 \\ \text{LARG} \quad h13 \end{bmatrix},
\begin{bmatrix} \textit{qeq} \\ \text{HARG} \quad h19 \\ \text{LARG} \quad h15 \end{bmatrix}
\right\rangle
\end{bmatrix}
$$

This semantic structure contains a relation with the predicate nominalization_rel and the verb is the first argument of this predication.[5] The intrinsic argument (ARG0) of the nominalization_rel is of type x for individual, rather than e for event, because any adjectival modifiers or quantifiers must be predicated of an individual. Because the low and mid analysis allow for the attachment of adjectives and quantifiers syntactically, this must be accounted for in the semantics as well. However, if adjectives and quantifiers do not attach to nominalized clauses in the language, a nominalization_rel is not not necessary under the high nominalization analysis. To provide the user with analytical freedom regarding the semantic structure, we provide the option for nominalization that is purely syntactic. In this case the unary rule changes the HEAD value to noun and creates a direct semantic identity between the mother and daughter without adding nominalization_rel.

4   Implementation in the Grammar Matrix

We implemented these analyses in the Grammar Matrix, such that the user can define multiple nominalization strategies that can be used in clausal complements or adverbial clausal modifiers. We evaluated the performance of this application by creating small grammars that include nominalization as it is used in a real language. We have tested nominalization in clausal complements with implemented grammars of Russian [rus] and Turkish [tur] and adverbial clausal modifiers in Rukai [dru]. As a standard practice we also test these phenomena on languages that were not considered in the development of the implementation, which we term "held out languages", in order to test the typological robustness of our analysis. The results of held out language evaluation will be included in the full paper.

5   Conclusion

In this paper we describe the range of nominalization in clausal complements and adverbial clausal modifiers and present three analyses to handle this variation. We take this analysis a step further by implementing it in the Grammar Matrix and testing it on data from three languages. We will extend this testing to held out languages and make the data and implemented grammars publicly available. By implementing our analyses in an open-source platform, we make them directly available for others to build on, either in the process of creating individual language grammars or for future research on nominalization cross-linguistically.

References

Bender, E. M., Flickinger, D., and Oepen, S. (2002). The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In J. Carroll, N. Oostdijk, and R. Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan.

Bender, E. M., Drellishak, S., Fokkens, A., Poulson, L., and Saleem, S. (2010). Grammar customization. *Research on Language & Computation*, **8**(1), 23–72. 10.1007/s11168-010-9070-1.

Copestake, A., Flickinger, D., Pollard, C., and Sag, I. A. (2005). Minimal recursion semantics: An introduction. *Research on Language and Computation*, **3**(2-3), 281–332.

Kornfilt, J. (1997). Turkish. descriptive grammars. *London and New York*.

Lapointe, S. (1993). Dual lexical categories and the syntax of mixed category phrases. In *Proceedings of the Eastern States Conference of Linguistics*, volume 199, page 210. CLC Publications Ithaca, NY.

Malouf, R. P. (2000). *Mixed Categories in the Hierarchical Lexicon*. CSLI Publications.

Noonan, M. (2007). Complementation. Language typology and syntactic description. Volume 2: Complex constructions. ed. by Timothy Shopen. 52-150. Cambridge: Cambridge University Press.

Thompson, S. A., Longacre, R. E., and Hwang, S. J. J. (2007). Adverbial clauses. Language typology and syntactic description. Volume 2: Complex constructions. ed. by Timothy Shopen. 237-269. Cambridge: Cambridge University Press.

---

[5]This relationship is mediated by a so-called *qeq* constraint, to accommodate the MRS analysis of quantifier scope ambiguities.