

# Modeling Adnominal Possession in Multilingual Grammar Engineering

Elizabeth Nielsen                  Emily M. Bender  
ekayen@uw.edu                  ebender@uw.edu  
*University of Washington*

## Introduction

The LinGO Grammar Matrix (Bender et al., 2002) is an open-source resource that allows user-linguists to jump-start the creation of implemented HPSG grammars. It consists of a web interface that elicits typological information from the user-linguist via a questionnaire (Bender et al., 2010), and Python-based back-end code that produces the implemented grammar. Since the Grammar Matrix project has always had the goal of covering the full attested range of the phenomena it includes, it functions not only as a tool for grammar engineers, but as a set of typological generalizations and predictions, in a testable and internally consistent format (Bender et al., 2002).

We extended the current Grammar Matrix customization system by adding a library to model adnominal possession. This abstract relates the typological and theoretical generalizations that were arrived at in the process of developing this extension to the Grammar Matrix. We begin by giving some background on the typological space the library was intended to cover and the way in which we broke down this typological space. Second, we discuss a typological generalization we arrived at in the process of library creation, namely the suitability of using only major phrase types already in existence in the Grammar Matrix (head-specifier, head-complement, head-modifier) to model possessive phrases. We demonstrate that all marked possessive constructions can be modeled without requiring specific additional binary phrasal constructions in any language. Lastly, we discuss another discovery, namely the implications of the decision made within the Matrix core grammar to bundle person, number, and gender features under a feature called PNG. This bundling of features turns out to be very beneficial in the context of multilingual grammar engineering, since it allows a consistent way of dealing with these features in languages with disparate ways of dealing with person, number, and gender.

## 1 Modeling the typological space

The goal in constructing this library was to model all cases of adnominal possession—that is, constructions involving two noun phrases whose referents participate in a possessive relation. The possessor may take the form of a noun or pronoun, which in the latter case may either be a separate word or an affix on the possessum. The majority of typological variation in possessive phrases can be captured by the following features:

1. Relative order of possessor and possessum
2. Presence/absence and morphotactic status of markers (affix, clitic, or word)
3. Location of possession-marking morphemes
4. Syntactic relation between possessor and possessum
5. Presence or lack of agreement between possessor and possessum

Items (1) and (2) above are fairly self-explanatory. Moving to item (3), languages either have no overt morphological markers of possession, or markers appear on the possessor, the possessum, or in both locations. For example, in Yoruba, possession is marked by inflection on the possessum:

- (1) owó            Dàda  
 money.POSS Dada  
 ‘Dada’s money’ (cf. owó ‘money’) [yor] (Haspelmath, 2005)

Moving to item (4), in the typological literature on possessive phrases, a distinction is often drawn between specifier-like possessors and modifier-like possessors. The English *’s*-genitive is a classic example of a construction with a specifier-like possessor, since the possessor fills the same slot as a specifier, blocking the possessum from taking a determiner. By contrast, Ancient Greek possessive pronouns are more like modifiers, in that they occur alongside the possessum’s determiner:

- (2) he:            to            patròs            oikía  
 the.F.SG.NOM the.M.SG.GEN father(M).SG.GEN house(F)SG.NOM  
 ‘the father’s house’ [gre] (Goodwin, 1894)

Lastly, there are languages in which the possessor agrees in person, number, and/or gender with the possessum, as well also languages (such as Finnish, illustrated in (3)) where the possessum agrees with the possessor:

- (3) heidän ystävä-nsä  
 their friend-3POSS  
 ‘their friend’ [fin] (Toivonen, 2000, 585)

This brief summary of the typological space under consideration provides the background for our analysis of possessives crosslinguistically. Next, we’ll see how, despite their differences from the sorts of phrases usually modeled as head-specifier or head-modifier phrases, possessive phrases are able to be modeled fully without adding any major phrase types to the Grammar Matrix.

## 2 Possessive phrase types

In the literature on adnominal possession, both within the HPSG framework and beyond, it is common to discuss possessive phrases as being one manifestation of highly general phrase types. In one classic example, Lyons (1986) draws a distinction between ‘adjective-genitives’ and ‘determiner-genitives’, suggesting that, modulo some inflectional morphology, possessors are essentially just another kind of specifier or modifier, no different from any other. Within the HPSG literature, there are many examples of analyses of possessive phrases being described as instances of head-modifier phrases (e.g. Beerman and Ephrem, 2007) or head-specifier phrases (e.g. Kolliakou, 1995). Though there are challenges in modeling possessive phrases in terms of these major phrase types, we demonstrate that it is possible to do so across the full typological space we explore. This serves to validate the practice of referring to possessive phrases as subtypes of these major phrase types.

Using major phrase types to model possessive phrases does present several challenges. For the purposes of this abstract, we discuss the use of the head-specifier construction to model all phrases with specifier-like possessors, though similar difficulties also exist with the head-modifier construction. As constituted in the Grammar Matrix with its implementation of Minimal Recursion Semantics (MRS; Copestake et al., 2005), the head-specifier rule is non-head-compositional—that is, semantic information (HOOK) from the non-head daughter is identified with semantic information on the parent (see (4) below). Given the nature and goals of the Grammar Matrix, this formulation of the head-specifier rule is not merely a convenient implementation choice, but a cross-linguistic analytical claim (Bender et al., 2002).

$$(4) \left[ \begin{array}{l} \textit{basic-head-spec-phrase} \\ \text{NON-HEAD-DTR.SYNSEM.LOCAL.CONT.HOOK } \boxed{1} \\ \text{C-CONT.HOOK } \boxed{1} \end{array} \right]$$

In a typical head-specifier construction, such as a noun phrase consisting of a determiner and a noun, the determiner identifies its own INDEX with the INDEX of the noun (through its SPEC value), so the INDEX of the head-specifier phrase is still identified with the INDEX of the noun. However, this approach does not work for modeling specifier-like possessive constructions. Take for example the scenario where the possessor is marked with a possessive affix. If we were to attempt a similar approach, the lexical rule for the possessive affix would look something like (5) (in abbreviated form), where the overall index of the lexical rule ( $\boxed{1}$ ) is identified with the index of the possessum noun, much like in the lexical type for determiners. Note that the possession relation is encoded in our MRS representation by the elementary predication with [PRED *poss\_rel*].

$$(5) \left[ \begin{array}{l} \textit{possessor-lex-rule} \\ \text{SYNSEM.LOCAL.CAT.VAL.SPEC } \left\langle \left[ \text{LOCAL.CONT.HOOK.INDEX } \boxed{1} \right] \right\rangle \\ \text{DTR.SYNSEM.LOCAL.CONT.HOOK.INDEX } \boxed{2} \\ \text{C-CONT } \left[ \begin{array}{l} \text{HOOK.INDEX } \boxed{1} \\ \text{RELS } \left\langle ! \left[ \begin{array}{l} \text{PRED } \textit{poss\_rel} \\ \text{ARG1 } \boxed{1} \\ \text{ARG2 } \boxed{2} \end{array} \right] ! \right\rangle \end{array} \right] \end{array} \right]$$

Problems arise with this analysis because the possessor noun must still participate in some constructions as a typical noun would, but it has partially adopted the semantics of the possessum. For example, when a determiner attaches to this possessor noun, it will quantify the possessum noun, rather than the possessor noun.

We solve this in our library in the following way: the *possessor-lex-rule* is pared down to a rule that simply adds a HEAD feature [ POSSESSOR + ]. A unary phrase rule then takes the NP consisting of the possessor (and any determiner and/or modifiers it may take) as its daughter, introduces the possessive predication (*poss\_rel* in (5) above), and produces a constituent whose INDEX is identified with the possessum. This allows the possessor to be a semantically typical noun within its own NP, and then to take on the necessary semantically anomalous behavior when interacting with the rest of the possessive phrase. This analysis is used for all specifier-like possessive constructions.

Though possessive phrases are challenging for the established major phrase types in the Grammar Matrix, it is ultimately still possible to assimilate them to the existing major phrase types (though at the cost of adding minor phrase types that are specific to possessives). This analysis supports the widespread claim that possessive phrases are essentially subtypes of head-specifier, head-modifier, and head-complement phrases, modulo inflectional morphology.

### 3 Feature bundling

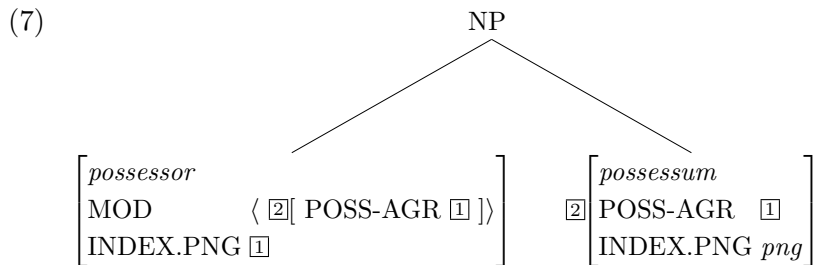
In this section, we discuss the analysis developed for agreement between possessor and possessum, focusing on how bundling together certain features is particularly effective in multilingual grammar engineering. The phenomenon of possessum agreeing with possessor (or vice versa) is observed in many languages, as shown in (3) above. When the possessor is specifier-like, this is easy enough

to account for: the possessum constrains the relevant person, number, and gender features of the possessor, which appears on its SPR list. Since the head and non-head daughters both select for each other in the head-specifier schema, this analysis works equally well in the case where the possessor agrees with the possessum. However, when the possessor is modifier-like, possessor and possessum are joined by a head-modifier rule which has no such mutual selection. The possessor can constrain the features of the possessum, which appears on its MOD list, but the possessum has no access to its possessor’s features. In order to fully cover the possible typological space, agreement in both directions should be possible whether the possessor is modifier-like or specifier-like.

For example, take the scenario where the possessor is the modifier of the possessum, but we still see agreement markers on the possessum. This agreement pattern can be seen in a construction like the following from Hungarian, where the possessum agrees with a modifier-like possessor:

- (6) az én kalap-ja-i-m  
 the I hat-POSS-PL-1SG  
 ‘my hats’ [hun] Laczko (2007)

Since the possessum cannot select its modifier, instead the possessum must somehow ‘publish’ the person, number or gender features it agrees with, so that the possessor can select for a possessum with the correct agreement features. This means it is necessary for the possessum to carry two sets of agreement features: the inherent person, number, and gender features it has as a noun; and the person, number, and gender features that it agrees with. The former are found (as usual) at INDEX.PNG, while the latter are in the new head feature we posit, called POSS-AGR. The possessum can then do the work of identifying the possessor’s agreement features with its own features, as sketched in the tree in (7):



Adding this second set of agreement features has the potential to be difficult in the multilingual grammar engineering context. While some languages have separate person, number, and gender features, others lack one of these three, or are better analyzed as having a combined PERNUM feature (Drellishak, 2009). Our library needs to be interoperable with all of these options. Given just three possible features, which may or may not appear, or which may be combined, there are a dozen possible features sets available. Creating different variants of the possessor-possessum agreement constraints in each of these cases would amount to redundantly reproducing the work of Drellishak’s PNG library.

Fortunately, and for independent reasons, Drellishak bundled all person, number, and gender features as features of the type *png*. This turns out to be very beneficial for us: We simply reuse the type *png* as the value of our new feature POSS-AGR. This allows our library to abstract away from the specifics of how person, number, and gender work. Thus we see that in addition to providing

efficiency as a monolingual level (Flickinger, 2000), types also add efficiency to cross-linguistic grammar engineering.

## 4 Conclusion

The process of implementing an analysis for any phenomenon inevitably leads to theoretical insights or analytical refinements. In the context of multilingual grammar engineering, the added constraint of harmonizing analyses for hundreds of possible variations on the phenomenon crosslinguistically provides all the more opportunity for finding such insights and refinements. In this abstract, we have detailed two ways in which the crosslinguistic perspective on modeling adnominal possession is beneficial, namely the confirmation of the applicability of major phrase types to modeling possessive phrases, and the advantages of bundling person, number, and gender features under a single type. This analysis has been tested by constructing testsuites for ten typologically and genetically diverse languages, half of which weren't considered during library development and then creating grammars using the augmented customization system to evaluate against those testsuites.

## References

- Dorothee Beerman and Binyam Ephrem. The definite article and possessive markings in Amharic. In Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu, and Jessica White, editors, *Proceedings of the Texas Linguistics Society IX Conference*, pages 21–32. CSLI Publications, 2007.
- Emily M. Bender, Dan Flickinger, and Stephan Open. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In John Carroll, Nelleke Oostdijk, and Richard Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan, 2002.
- Emily M. Bender, Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyah Saleem. Grammar customization. *Research on Language and Computation*, 8(1):23–72, 2010.
- Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan Sag. Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2):281–332, July 2005.
- Scott Drellishak. Widespread but not universal: Improving the typological coverage of the grammar matrix, January 2009. URL <http://search.proquest.com/docview/305015786/>.
- Dan Flickinger. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6 (1) (Special Issue on Efficient Processing with HPSG):15–28, 2000.
- William W. Goodwin. *A Greek Grammar*. Macmillan & Co., London, 1894.
- Martin Haspelmath. *The world atlas of language structures*. Oxford University Press, Oxford, 2005.
- Dimitra Kolliakou. *Definites and possessives in Modern Greek*. PhD thesis, University of Edinburgh, 1995.
- Tibor Laczko. Revisiting possessors in Hungarian DPs: a new perspective. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG07 Conference*. CSLI Publications, 2007.
- Ida Toivonen. The morphosyntax of Finnish possessives. *Natural Language and Linguistic Theory*, 18(3):579–609, 2000.