

# Supplement S1: Identifying and classifying adnominal possessive constructions in Universal Dependencies treebanks

Viljami Haakana and Kaius Sinnemäki

General linguistics  
University of Helsinki  
Contact: [kaius.sinnemaki@helsinki.fi](mailto:kaius.sinnemaki@helsinki.fi)

22 October 2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Analyses</b>	<b>3</b>
2.1	Afrikaans	3
2.2	Ancient Greek	3
2.3	Arabic	4
2.4	Basque	4
2.5	Bulgarian	4
2.6	Catalan	4
2.7	Chinese	4
2.8	Croatian	5
2.9	Czech	5
2.10	Danish	5
2.11	Dutch	5
2.12	English	6
2.13	Estonian	6
2.14	Finnish	6
2.15	French	8
2.16	German	8
2.17	Gothic	8
2.18	Greek	8
2.19	Hebrew	9
2.20	Hindi	9
2.21	Hungarian	9
2.22	Indonesian	9
2.23	Italian	9
2.24	Japanese	10
2.25	Korean	10
2.26	Latin	10
2.27	Latvian	10
2.28	Norwegian	11
2.29	Old Church Slavonic	11
2.30	Persian	11
2.31	Polish	11
2.32	Portuguese	12
2.33	Romanian	12
2.34	Russian	13
2.35	Serbian	13
2.36	Slovak	14

2.37	Slovenian . . . . .	14
2.38	Spanish . . . . .	14
2.39	Swedish . . . . .	14
2.40	Turkish . . . . .	15
2.41	Ukrainian . . . . .	15
2.42	Urdu . . . . .	15
2.43	Uyghur . . . . .	15
2.44	Vietnamese . . . . .	16
<b>3</b>	<b>Some further methodological issues</b>	<b>16</b>
3.1	Algorithm to identify conjoined phrases . . . . .	16
3.2	On the definition of dependency length . . . . .	17

# 1 Introduction

In this supplement we provide a detailed description of the algorithm that we used for identifying adnominal possessive constructions from the Universal Dependencies (UD) treebanks. These constructions are here also called adpossessive constructions or possessive noun phrases (NP). The key definitions, the sample, and our analyses are described in the main text of the article. In this section we repeat some of the main definitions for the sake of clarity, and describe at a general level what information we used from the UD annotations. In Section 2 we provide the descriptions for each individual sample language, followed by some further methodological issues in Section 3, such as addressing conjoined phrases and measuring dependency length.

An adnominal possessive construction is syntactically a noun phrase whose head is a noun and that has a noun or a pronoun as a dependent modifier. Semantically the relation between the head and the dependent in these constructions expresses typically (alienable) possession, such as *my car*, part-whole relationships (including inalienable possession), such as *my hand*, and kinship relationships, such as *my daughter*. Syntactic adnominal possessive constructions can also be used for various other functions depending on each language (Ortmann, 2018; Haspelmath, 2017). In adnominal possessive constructions the syntactic **dependent** is semantically the possessor and the syntactic **head** is semantically the possessee. For example, in the construction *my daughter* the syntactic head is the possessee *daughter* and the dependent is the possessor *my*.

The syntactic relationship between the head and the dependent in adpossessive constructions can be marked morphologically on the head (i.e., head marking), on the dependent (i.e., dependent marking), on both the head and the dependent (i.e. double marking), or on neither the head nor the dependent (i.e. zero marking), following the typology on locus of marking by Nichols (1992). This variety of morphological marking means each adpossessive construction with a different locus of marking is annotated in the current UD in a slightly different way. As a result, we classified the locus of marking for each adpossessive construction as well.

In our analyses the head of the adpossessive construction is always a noun or a proper noun. In terms of the universal morphosyntactic annotation of the UD, the parts-of-speech (POS) tag of the head is either NOUN or PROPN. This is not repeated in the particular languages' description below. In our analyses we regularly use the following information from the UD annotation: parts-of-speech (4th column), lexical and grammatical features associated with the word form (6th column), and syntactic function (8th column). We also compare the lemma of the possessor (3rd column) to its surface form (2nd column) when analyzing the morphological marking in a number of languages. In two very specific instances we used information about language-specific analyses, namely, language specific POS annotation (5th column) for Swedish, and language-specific syntactic function (9th column) for Latvian.

## 2 Analyses

### 2.1 Afrikaans

Afrikaans (Donaldson, 1993) has three ways of indicating adnominal possession: possessive pronouns, using the preposition *van* to mark the dependent as the possessor, and using the particle *se* to mark the dependent as the possessor.

In adpossessive constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `det`.

In adpossessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPN. In addition, one of the following sets of conditions must be met:

- 1) When the preposition *van* is used for marking the possessor as a dependent, the POS of the preposition must be ADP, its syntactic function must be `case`, and the syntactic function of the possessor must be `nmod`.
- 2) When the particle *se* is used for marking the possessor as a dependent, the POS of the particle must be PART and the syntactic function of the possessor must be `nmod`.

For most adpossessive constructions in Afrikaans the syntactic relation between the head and the dependent is marked on the dependent. The only exception is formed by a group of pronoun possessors whose lemma and surface form are identical, that is, they do not have separate forms that appear in possessive constructions. These pronouns are *u*, *ons*, *julle*, and *hulle*.

### 2.2 Ancient Greek

Adpossessive constructions in Ancient Greek (Benvenuto and Pompeo, 2015; Goodwin, 1900) were identified in the following way.

In adpossession constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation contains `Person=`, and its syntactic function is `nmod`. All these constructions are dependent marked. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, its syntactic function is `nmod`, and it is modified by a word whose lemma is *ó* (POS DET, syntactic function `det`). All these constructions are also classified as dependent marked.

## 2.3 Arabic

In Arabic (Gadalla and Abdel-Hamid, 2000; Ryding, 2005) adpossession constructions with a pronoun possessor were identified in the following way: the POS of the possessor is PRON, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `nmod`. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`. A further criterion for both pronoun and noun possessors is that they are not modified by prepositions (POS ADP, syntactic function `case`). All adpossession constructions in Arabic are dependent marked.

## 2.4 Basque

In Basque (Saltarelli, 1988), adpossession constructions with a pronoun possessor are identified so that the POS of a pronoun possessor is PRON, its morphological annotation contains `PronType=Prs`, and its syntactic function is `nmod`. If the lemma of the possessor is identical with its surface form, then the adpossession construction is classified as zero marked, otherwise as dependent marked. All word-initial asterisks are removed before the lemma comparison.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`. In addition, the noun possessor must not be modified by an adposition (POS ADP, syntactic function `case`). If the lemma of the noun possessor is identical with its surface form, the adpossession construction is classified as zero marked, otherwise as dependent marked.

## 2.5 Bulgarian

Adpossession constructions in Bulgarian (Mitkovska, 2009; Scatton, 1993; Antova et al., 2002) are formed in two ways. First, in adpossession constructions with a pronoun possessor, the POS of the possessor is PRON or DET, its morphological annotation contains both `Poss=Yes` and `PronType=Prs`, and its syntactic function is `det`. In addition, the morphological annotation must not contain `Reflex=Yes`, in order to exclude reflexive pronouns from the analysis. Second, in adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its syntactic function is `nmod`, and it must be modified by the adposition *на* (syntactic function `case`). All adpossession constructions in Bulgarian are dependent marked.

## 2.6 Catalan

Adpossession constructions in Catalan (Hualde, 1992) are identified in the following way. In adpossession constructions with a pronoun possessor, the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `det`. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its syntactic function is `nmod`, and it must be modified by the preposition *de* (whose lemma can also be *del*). Overall, adpossession constructions are identified in a very similar way in both Catalan and Spanish.

## 2.7 Chinese

In Mandarin Chinese (Luo, 2012), adpossession constructions with a pronoun possessor are identified in the following way: the POS of the possessor is PRON and its morphological annotation contains `Person=` to identify it as a personal pronoun. If the syntactic function of the pronoun possessor is `det`, then the construction is classified as zero marked, but if its syntactic function is `nmod`, it is classified as dependent marked.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its syntactic function is `det`, and it must be modified by a particle (POS is PART, morphological annotation contains `Case=Gen`, and syntactic function is `case:dec`). Accordingly, adpossession constructions with noun possessors in Chinese are analyzed as dependent marked. A separate Python script has been written to identify adpossession constructions in Chinese (`Chinese.py`).

## 2.8 Croatian

In Croatian (Nomachi, 2016; Alexander, 2006), adpossession constructions with a pronoun possessor are identified in the following way: the POS of the possessor is DET, its morphological annotation contains Poss=Yes and PronType=Prs but not Reflex=Yes, and its syntactic function is either det or amod (the last criterion to identify possessive adjectives).

Some words in Croatian, such as *plaća* ‘pay’, have identical forms for genitive plural and nominative singular but different from the nominative plural. For this reason, all adpossession constructions with a plural possessor noun were classified as dependent marked regardless of the lemma and the surface form being identical. The criteria for identifying other adpossession constructions are the following:

- 1) If a plural noun possessor (the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen and Number=Plur, its syntactic function is nmod) is not modified by an adposition (POS ADP and syntactic function case), the adpossession construction is classified as dependent marked.
- 2) If a singular noun possessor (the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen and Number=Sing, and its syntactic function is nmod) is not modified by an adposition (POS ADP, syntactic function case), the adpossession construction was classified as zero marked if the lemma and the surface form of the possessor are identical, otherwise they were classified as dependent marked.
- 3) If the possessor is a possessive adjective (POS ADJ, morphological annotation containing Poss=Yes, and syntactic function amod), the adpossession construction is classified as dependent marked.

## 2.9 Czech

For Czech language (Křivan, 2013; Harkins, 1953; Naughton, 2005) we used three different treebanks and the following criteria work for each.

In adpossession constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation contains Poss=Yes and PronType=Prs, and its syntactic function is det. In addition, the morphological annotation must not contain Reflex=Yes to exclude reflexive pronouns. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological analyses contains Case=Gen, and its syntactic function is nmod. In addition, the adnominal possessor must not be modified by an adposition (identified with POS being ADP and syntactic function being case). Adpossession constructions in Czech are generally dependent marked. A handful of noun possessors have identical lemma and surface forms, and these instances are analyzed as zero marked.

As a Slavic language, Czech expresses adnominal possession also by using possessive adjectives. Adpossession constructions with possessive adjectives are identified in the following way: the POS of the possessor is ADJ, its morphological annotation contains both [psor] and Poss=Yes, and its syntactic function is amod. These constructions are classified as dependent marked.

## 2.10 Danish

Adnominal possession in Danish (Allan and Lundskaer-Nielsen, 2000) is marked by using possessive pronouns, genitive case, or the preposition *af* ‘of’.

In adpossession constructions with a pronoun possessor, the POS of the possessor is DET or PRON, its morphological annotation contains Poss=Yes and PronType=Prs but not Reflex=Yes, and its syntactic function is either det or nmod:poss. If the POS of the pronoun is DET, its syntactic function is det and if the POS is PRON, then its syntactic function is nmod:poss. These constructions are classified as dependent marked.

Adpossession constructions with a noun possessor are identified in two ways. First, the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen, and its syntactic function is nmod:poss. These genitive constructions are classified as zero marked if the lemma and the surface form of the possessor are identical, otherwise it is classified as dependent marked. Second, noun possessors (POS is NOUN or PROPN; syntactic function is nmod) can be modified by a possessive preposition *af* (its POS is ADP and syntactic function is case). These latter constructions are all classified as dependent marked.

## 2.11 Dutch

In Dutch (Donaldson, 2017) adpossession constructions with a pronoun possessor are identified in the following way (with a separate Python script *Dutch.py*): the POS of the possessor is PRON, its morphological annotation

contains `PronType=Prs`, and its syntactic function is `nmod:poss`. These constructions are classified as dependent marked.

Adpossession constructions with noun possessors are identified in two ways. First, there are the constructions with the adposition *van* ‘of’. In these adpossession constructions the POS of the noun possessor is NOUN or PROP, its syntactic function is `nmod`, and it is modified by the preposition whose lemma is *van* ‘of’ (POS ADP, syntactic function `case`). All these constructions are classified as dependent marked. Second, adpossession constructions with a noun possessor may be dependent marked with *de*. In these constructions the POS of the noun possessor is NOUN or PROP, its syntactic function is `nmod`, it is modified by the surface form *der* (its lemma is *de* and its syntactic function is `det`), and it is not modified by *van* or other adpositions. These adpossession constructions are classified as dependent marked.

## 2.12 English

There are three ways to mark possession in adpossession constructions in English. English uses possessive pronouns such as *my*, *her*, and *their*, the possessive clitic *’s*, and the preposition *of*. These are identified in slightly different ways in the UD treebanks available for English; but regardless of their identification, all adpossession constructions in English are classified as dependent marked.

In adpossession constructions with a pronoun possessor the POS of the possessor is DET in the English ParTUT treebank but PRON in the other English treebanks. In all English treebanks the morphological annotation of the possessor contains `Poss=Yes` and `PronType=Prs` and its syntactic function is `nmod:poss`.

Adpossession constructions that use the *of* preposition were identified in the following way for all English treebanks. The POS of the possessor is NOUN or PROP, its syntactic function is `nmod`, and it is modified by the preposition *of* (POS ADP, syntactic function `case`). Adpossession constructions that use the possessive clitic *’s* were identified in the following way. The POS of the possessor is NOUN or PROP and it is modified by a particle *’s* (POS PART, syntactic function `case`). The syntactic function of the possessor noun is `nmod:poss` in the EWT and GUM treebanks and `nmod` in the ParTUT treebank; in the LinES treebank its syntactic function can be either `nmod:poss`, `amod` or `nmod`.

## 2.13 Estonian

In Estonian adpossession constructions with a pronoun possessor are identified in the following way: the POS of the possessor is PRON, its morphological annotation contains `PronType=Prs`, `Case=Gen`, and `Person=`, and its syntactic function is `nmod`. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROP, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`.

The morphological marking of adpossession constructions was determined via comparing the lemma of the dependent to its surface form. If the lemma and the surface form of a word were identical, the adpossession construction was classified as zero marked, otherwise as dependent marked. The algorithm identifies the lemma and the word form as identical if, after removing underscores from the lowercase word form, one of the following is true:

- 1) the word form is either *mina*, *ma*, *me* or *meie* and the lemma is *mina*,
- 2) the word form is either *sina*, *sa*, *te* or *teie* and the lemma is *sina*,
- 3) the word form is either *nemad*, *nad*, *ta* or *tema* and the lemma is *tema* or
- 4) the word form is identical to the lowercase lemma after removing underscores and equals signs.

## 2.14 Finnish

Adpossession constructions in Finnish appeared the most difficult to analyze and classify. Their identification in the two Finnish treebanks became so complicated that we advise the reader who wants to understand it in detail to leaf through the source code in the script `Finnish.py`. In this subsection we summarize only the most important criteria.

We first identified a group of potential adpossession constructions with the following criteria:

- 1) The POS of the possessor is PRON, its morphological annotation contains `PronType=Prs`, and its syntactic function is `nmod:poss`.
- 2) The POS of the possessor is PRON, its morphological annotation contains `Case=Gen` and `PronType=Prs`, and its syntactic function contains `nmod`.

- 3) The POS of the possessor is DET and its syntactic function det.
- 4) The POS of the possessor is NOUN or PROPN and its syntactic function is nmod:poss.
- 5) The POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen, and its syntactic function contains nmod.
- 6) The POS of the **possessee**, that is the head of the adpossession construction, is NOUN or PROPN, its morphological annotation contains [psor], and it is modified neither by a dependent possessor whose POS is NOUN, PROPN or PRON and whose syntactic function contains nmod nor by a dependent possessor whose POS is DET and whose morphological annotation contains Case=Gen or Poss=Yes. These adpossession constructions are classified as head\_exist constructions, that is, the possessor is only marked on the head and not as a separate dependent word.

Having collected these potential **dependents** we removed typos<sup>1</sup> as well as possessors that were modified by any adposition (POS ADP). This preprocessing resulted in a list of adpossession constructions in Finnish. Their morphological marking was classified in the following way.

- 1) If the morphological annotation of the possessor contains Case=Gen or Poss=Yes and the morphological annotation of the possessee (the head of the construction) contains [psor], the adpossession construction is classified as double marked.
- 2) If the morphological annotation of the possessor contains Case=Gen or Poss=Yes and the morphological annotation of the possessee (the head of the construction) does not contain [psor], the adpossession construction is classified as dependent marked.

There are some qualifications, however.

- 1) If the morphological annotation of the possessor contains Case=Gen and the morphological annotation of the possessee (the head of the construction) does not contain [psor], the adpossession construction is classified as zero marked provided that the surface form of the possessor ends in a vowel and the lemma of the possessor is identical with its surface form.<sup>2</sup> If the possessor does not end in a vowel and the lemma of the possessor is not identical with its surface form, the adpossession construction is classified as dependent marked.
- 2) If the morphological annotation of the possessor contains psor and Number=Sing, the adpossession construction is classified as zero marked.<sup>3</sup> The reason for this classification is that all Finnish words with possessive suffixes are identical in nominative singular, genitive singular and nominative plural. For example, the expression *lapseni* is ambiguous in three ways, meaning ‘my child’, ‘my child’s’ and ‘my children’. In our analyses this ambiguity results in zero-marked adpossession constructions, but only with singular possessors. Thus the adpossession construction *lapseni paras ystävä* ‘my child’s best friend’ is classified as zero marked. This phenomenon, as stated before, is true for (almost) all Finnish nouns.<sup>4</sup>

In rare instances the possessive suffix seemed to be misanalyzed in the treebank if the expression was in a colloquial form. For instance, the colloquial expression *Meidän suhde* ‘our relationship’ (from sentence b204.33 in the Finnish-TDT treebank) was analyzed so that the morphological annotation for the head (*suhde*) contained a possessive suffix even though there is no possessive suffix in the example. The morphological annotation for the head contains Style=Coll to mark the expression as colloquial language. For this reason we double-checked whether the morphological annotation of the possessee matched with the actual surface form.<sup>5</sup>

<sup>1</sup>Typos included, for instance, the misspelled genitive *kesä*, whose correct spelling would have been *kesän* in sentence b302.17 of the treebank TDT. This construction would have been classified as zero marked had it not been annotated with Typo=Yes in its morphological annotation.

<sup>2</sup>The reason why we need to check whether the word ends in a vowel or not is because the lemma of *Helsingin* in *Helsingin Sanomat* (a large Finnish newspaper) in FTB’s sentence c1p9c-1701 has been annotated as *helsingin* instead of *helsinki* even though *Helsingin* is the genitive of *Helsinki*. The name of this newspaper has been annotated correctly in some other sentences in the same treebank, e.g. in o6ovb-9357. Analyzing the vowels enables also identifying dialectal genitive constructions where the genitive case -n has been dropped, resulting in a word form that is identical with the nominative. Both treebanks contain one such example (sentence w8k7r-14723.2 in the FTB treebank and h1079.12 in TDT).

<sup>3</sup>A list of possessive suffixes, including the most common colloquial/dialectal ones, and their associated person and number pairings can be found in the source code in the variable `possessive_suffixes_dict`.

<sup>4</sup>This ambiguity does not exist as such for compound words whose initial part also declines, such as *nuori pari* ‘newlywed couple’ (*nuori* ‘young’, *pari* ‘couple, pair’). Because these compounds are infrequent and difficult to identify, they were not addressed in the algorithm.

<sup>5</sup>The existence of Style=Coll and [psor] in the morphological annotation do not yet imply absence of a possessive suffix. For example, the second person singular possessive suffix -si is often shortened to -s in speech (instead of dropping it entirely), so the word *juttusi* in Finnish-TDT’s sentence fF02.42 contains Style=Coll as it would be *juttuasi* in Standard Finnish. This adpossession construction was classified as double marked, because the possessive suffix -s exists in the surface form, although in colloquial form.

## 2.15 French

In French (Batchelor and Chebli-Saadi, 2011) adpossession constructions were identified in two ways in the UD treebanks. In adpossession constructions with pronoun possessors, the POS of the possessor is DET or PRON, its morphological annotation contains `PronType=Prs`, and its syntactic function is `det`. These constructions were classified as dependent marked. In adpossession constructions with noun possessors, the POS of the possessor is NOUN or PROPN, its syntactic function is `nmod`, and it is modified by the preposition *de* (its POS is ADP and syntactic function is `case`). These constructions are classified as dependent marked as well.<sup>6</sup>

## 2.16 German

Adpossession constructions in German were analyzed with a separate Python script (`German.py`). Adpossession constructions with a pronoun possessor were identified in two ways. First, the POS of the possessor is DET, its morphological annotation contains either `Poss=Yes` or `PronType=Prs`, and the syntactic function is `det:poss`. Second, the POS of the possessor is PRON, its morphological annotation contains either `Poss=Yes`, and the syntactic function is `det:poss`. Both adpossession constructions were classified as dependent marked.

In adpossession constructions with a noun possessor, some constructions contain the preposition *von*. In these constructions the POS of the noun possessor is NOUN or PROPN, its morphological annotation contains `Case=Dat`, its syntactic function is `nmod`, and the possessor is modified by the adposition *von* (POS ADP, syntactic function `case`). These constructions were classified as dependent marked. Some adpossession constructions with a noun possessor are identified based on genitive case marking. In these constructions the POS of the noun possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, its syntactic function is `nmod`, and it is not modified by any word whose POS is DET and whose syntactic function is `det`. If the lemma and the surface form of the possessor are identical, the adpossession construction is zero marked, otherwise it is dependent marked. For every conjoined possessive NP (their syntactic function is `conj`), repeat these steps for every `conj` in the sentence.

## 2.17 Gothic

Adpossession constructions in Gothic were analyzed using information from Harbert (1978) as well as using Wiktionary for identifying Gothic words from Bible passages and comparing them with other translations.

To identify adpossession constructions, one of the following conditions must be met:

- 1) The POS of the possessor is PRON, its morphological annotation contains `PronType=Prs`, and its syntactic function is `det`.
- 2) The POS of the possessor is ADJ, its morphological annotation contains `Poss=Yes` and no `Reflex=Yes`, and its syntactic function is `nmod`.
- 3) The POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`.

All these constructions were classified as dependent marked.

## 2.18 Greek

In Modern Greek (Joseph and Philippaki-Warbuton, 1987) adpossession constructions with a pronoun possessor were identified in the following way: the POS of the possessor is PRON, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `nmod`. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`. Accordingly, all adpossession constructions in Greek were classified as dependent marked.

---

<sup>6</sup>In French, the word “des” is both a contraction of the preposition *de* and the definite plural article *les*, as well as the indefinite plural article on its own. Therefore, expressions such as “des noms” can mean both ‘names’ and ‘of the names’. One might suppose that these instances with *des* were zero marked, but because one of them is definite and the other is indefinite, this homonymy does not result in zero marking. If *de* and the indefinite plural article *des* were to be contracted to *des*, that would create zero marked possessive NPs into the French language. As stated before, “des noms” means both ‘names’ and ‘of the names’, but not ‘the names’ nor ‘of (some) names’, and it would be required to have at least one of those meanings as well in order to lead to zero marking.



## 2.19 Hebrew

In Hebrew (Glinert, 1989) adposessive constructions with a pronoun possessor were identified in the following way. The POS of the possessor is PRON, its morphological annotation contains `PronType=Prs`, and its syntactic function is `nmod:poss`. In adposessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPON, its syntactic function is `nmod:poss`, and it is modified by an adposition whose syntactic function is `case:gen` (POS ADP). All the identified adposessive constructions are classified as dependent marked.

## 2.20 Hindi

In Hindi (Kachru, 2006) adposessive constructions with a pronoun possessor were identified in the following way. The morphological annotation of the possessor contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `nmod`. In adposessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPON, its morphological annotation contains `Case=Acc`, its syntactic function is `nmod`, and it is modified by the adposition क़ी (POS ADP; syntactic function `case`). All the identified adposessive constructions in Hindi are classified as dependent marked.

## 2.21 Hungarian

In Hungarian (Kenesei et al., 1998) adposessive constructions with a pronoun possessor were identified in the following way. The morphological annotation of the possessor contains `PronType=Prs` and its syntactic annotation is `nmod:att`. These constructions are classified as head marked. In adposessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPON, its syntactic function is `nmod:att`, and the morphological annotation of the possessee (the head noun) contains `[psor]=`. If the morphological annotation of the possessor contains `Case=Nom`, the construction is classified as head marked. If the morphological annotation of the possessor contains `Case=Dat`, the construction is classified as double marked. If the construction does not have a possessor, the construction is classified as `head_exist`, that is, as head marked construction with no separate possessor as a dependent.

## 2.22 Indonesian

In Indonesian (Sneddon, 1996; Sneddon et al., 2010) adposessive constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation contains `PronType=Prs` and its syntactic function is `det`. If the morphological annotation of the possessee (the head of the construction) contains `[psor]`, the construction is head marked, otherwise it is zero marked.

In adposessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPON, its syntactic function is `nmod`, and it is not modified by an adposition (POS ADP, syntactic function `case`). If the morphological annotation of the possessee (the head of the construction) contains `[psor]`, the construction is head marked, otherwise it is zero marked. If the construction does not have a possessor (dependents with syntactic function `nmod` or `det`) but the morphological annotation of the possessee contains `[psor]`, the construction is classified as `head_exist`, that is, as head marked construction with no separate possessor as a dependent.

## 2.23 Italian

In Italian (Maiden and Robustelli, 2007) there are several different ways for identifying adposessive constructions in the three available treebanks. In adposessive constructions with a pronoun possessor, for three ways how pronoun possessors can be identified:

- 1) the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `det:poss`, or
- 2) the POS of the possessor is PRON, its morphological annotation contains `PronType=Prs`, its syntactic function is `nmod`, and it is modified by the preposition *di* (POS ADP; syntactic function `case`), or alternatively
- 3) the POS of the possessor is PRON, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, its syntactic function is `nmod`, and it is not modified by a preposition (POS ADP; syntactic function `case`).

Each of these different types of adposessive constructions are classified as dependent marked. In 1), there is some potential for zero marking. The word *loro* can mean both ‘they’ and ‘their’, so cases where *loro* is used in the meaning ‘their’ without an article are classified as zero marked (normally pronoun possessors are indicated in

Italian as “definite article + possessive pronoun + possessee”, for example *il mio paese* ‘my country’, literally “the my country” – but in some cases the article is omitted, and if the article is omitted and the pronoun is *loro*, then (and only then) the construction is classified as zero marked.<sup>7</sup>

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its syntactic function is *nmod*, and it is modified by the preposition *di* (POS ADP; syntactic function is *case*).<sup>8</sup> These adpossession constructions are classified as dependent marked.

## 2.24 Japanese

Our algorithm recognizes only one type of adpossession construction in Japanese (Hinds, 1986), all of which are classified as dependent marked. In adpossession constructions, the POS of the possessor is NOUN, PROPN or PRON, its syntactic function is *nmod*, and it is modified by a possessive postposition *no*.

## 2.25 Korean

The UD contains two treebanks for Korean and adpossession constructions are identified and classified in slightly different ways in them. For this reason we wrote two Python scripts to analyze the Korean treebanks (`KoreanHYPHENGSD.py` and `KoreanHYPHENKaist.py`). Our main reference source for Korean was Chang (1996).

In the Korean-GSD treebank, adpossession constructions are identified in the following way: the POS of the possessor is PRON, NOUN, or PROPN and its syntactic function is *det : poss*. In Korean-Kaist, the syntactic function of the possessor is *nmod* and its lemma (not actually a lemma but a morpheme segmentation) must end in *의*. All identified constructions are classified as dependent marked.<sup>9</sup>

## 2.26 Latin

In Latin (Bennett, 1908) adpossession constructions are identified in slightly different ways in the available treebanks.

In some adpossession constructions the POS of the possessor is ADJ, its morphological annotation contains *Poss=Yes*, and its syntactic function is *amod* or *nmod*. These adpossession constructions were classified as dependent marked. In others, the POS of the possessor is NOUN or PROPN, its morphological annotation contains *Case=Gen*, and its syntactic function is *nmod*. If the lemma and the surface form of the possessor are identical, the adpossession construction was classified as zero marked, otherwise as dependent marked.

In the PROIEL treebank, there is one additional way of identifying adpossession constructions. In adpossession constructions with a pronoun possessor, the POS is PRON, its morphological annotation contains *Case=Gen* and *PronType=Prs*, and its syntactic function is *det*. These adpossession constructions were classified as dependent marked.

## 2.27 Latvian

Unlike in treebanks for other languages, in Latvian adpossession constructions were identified using the language-specific syntactic functions. For this reason we wrote a separate Python script (`Latvian.py`) to analyze its treebank. Our main reference source was Fennell and Gelsen (1980).

Adpossession constructions with a pronoun possessor were identified and classified in the following way. The POS of the possessor is DET, its morphological annotation contains *Poss=Yes* and *PronType=Prs*, and its syntactic function is *det*. These adpossession constructions were classified as dependent marked.

Adpossession constructions with a noun possessor were identified and classified in the following way. The POS of the possessor is NOUN or PROPN and its language-specific syntactic function is *nmod : gen*. If the lemma and the surface form were identical, the adpossession construction was classified as zero marked, otherwise as dependent marked.

As for identifying conjoined possessive constructions involving pronoun possessors, conjoined possessive constructions (syntactic function is *conj*) do not seem to exist in the Latvian treebank. For possessors whose POS is

<sup>7</sup>The head of the article is the head of the entire NP, which is why the classification instruction in the tag chart is `i1-DET~~clemma.zero;dep`. That means: ‘if the head has no dependent whose lemma is *il* and whose POS is DET, then do a lemma comparison for the actual dependent of the NP, and if it returns the value `True`, then classify it as a zero marked possessive NP, otherwise classify it as a dependent marked one’. The lemma comparison in Italian returns `True` only if the lowercase word form and the lemma are both equal to *loro*.

<sup>8</sup>Note that the lemma of *di* can also be capitalized (*Di*) in the Italian-PoSTWITA treebank.

<sup>9</sup>We are grateful to N. N. for consulting us with the Korean treebanks. In addition to the abovementioned adpossession constructions in Korean, (colloquial) Korean has also other adpossession constructions, such as those marked with *의*. However, we were unable to identify these adpossession constructions reliably, so they are not included in our analysis.

DET, the syntactic function (including the language-specific syntactic function) is never `conj`. As for identifying conjoined possessive constructions involving noun possessors, conjoined possessors are identified as words whose head (possee) is a dependent of a possessive noun phrase and whose language-specific syntactic function is `conj`. The lemma comparison is also done to potential conjoined noun possessors (syntactic function `conj`).

## 2.28 Norwegian

Adnominal possession in Norwegian is very similar to that in Danish, being marked by possessive pronouns, genitive case, or the preposition *av* ‘of’.

In adpossession constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `nmod`.

Adpossession constructions with a noun possessor are identified in two ways. First, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod:poss`. These genitive constructions are classified as zero marked if the lemma and the surface form of the possessor are identical, otherwise it is classified as dependent marked. Second, noun possessors (POS is NOUN or PROPN; syntactic function is `nmod`) can be modified by a possessive preposition *av* (its POS is ADP and syntactic function is `case`). These latter constructions are all classified as dependent marked.

## 2.29 Old Church Slavonic

In Old Church Slavonic (Lunt, 2001) adpossession constructions with pronoun possessors are identified in the following way. The POS of the possessor is PRON, its morphological annotation contains `Case=Gen` and `PronType=Prs` but no `Reflex=Yes`, and its syntactic function is either `det` or `nmod`.

Adpossession constructions with possessive adjectives are identified in the following way: the POS of the possessor is ADJ, its morphological annotation contains `Poss=Yes` or `Case=Gen` but no `Reflex=Yes`, and its syntactic function is `nmod`.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`. In addition, the possessor must not be modified by an adposition (POS ADP; syntactic function `case`) and they must occur in the adpossession construction after the possee (the head noun). In all identified instances the adpossession construction is dependent marked.

## 2.30 Persian

In Persian (Lazard, 1992) adpossession constructions are identified in the following way. In adpossession constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation contains `PronType=Prs` but not `Reflex=Yes`, and its syntactic function is `nmod:poss`. In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN and its syntactic function is `nmod:poss`.

The syntactic relation between the head and the dependent is morphologically marked on the head in Persian. The marking, called *ezāfe*, is usually not marked in the written language, and because the UD treebank for Persian contains written language, the *ezāfe* is not marked in the treebank. We simply assume that all adpossession constructions in Persian are consistently head marked.

## 2.31 Polish

The UD contains two treebanks for Polish and adpossession constructions are identified in significantly different ways in them. For this reason we wrote two Python scripts to analyze these treebanks (`PolishHYPHENLFG.py` and `PolishHYPHENSZ.py`). Our main reference source for Polish was Brooks (1975).

In the Polish-LFG treebank, adpossession constructions with a pronoun possessor are identified in three ways. First, the POS of the possessor is PRON and its syntactic function is `nmod:poss`. Second, the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `nmod:poss`. Third, the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs` and no `Reflex=Yes`, and its syntactic function is `det`. All these adpossession constructions are classified as dependent marked.

In the Polish-LFG treebank, adpossession constructions with a noun possessor are identified in the following way. The POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod:poss`. If the lemma and the surface form of the possessor are identical, the adpossession construction is classified as zero marked, otherwise as dependent marked.

In the Polish-SZ treebank, adpossession constructions with a pronoun possessor are also identified in three ways. First, the POS of the possessor is PRON, its morphological annotation contains `Case=Gen`, `PronType=Prs`

and `Person=`, and its syntactic function is `nmod`. Second, the POS of the possessor is `DET`, the morphological annotation contains `Case=Gen`, `Poss=Yes` and `PronType=Prs` but no `Reflex=Yes`, and its syntactic function is `nmod`. Third, the POS of the possessor is `DET`, the morphological annotation contains `Case=Gen`, `Poss=Yes` and `PronType=Prs` but no `Reflex=Yes`, and its syntactic function is `det`. All these adpossession constructions are classified as dependent marked.

In the Polish-SZ treebank, adpossession constructions with a noun possessor are identified in the following way. The POS of the possessor is `NOUN` or `PROPN`, its morphological annotation contains `Case=Gen`, its syntactic function is `nmod`, and it should not be modified by a preposition (POS `ADP`; syntactic function `case`).<sup>10</sup> If the lemma and the surface form of the possessor are identical, the adpossession construction is classified as zero marked, otherwise as dependent marked.

## 2.32 Portuguese

There are three ways for identifying adpossession constructions in Portuguese, described in the following. Our main reference source was Hutchinson and Lloyd (2003).

Adpossession constructions with a pronoun possessor are identified in two ways. First, the POS of the possessor is `PRON`, its morphological annotation contains `PronType=Prs`, its syntactic function is `nmod`, and it is modified by a word whose lemma is *de* (syntactic function `case`). Second, the POS of the possessor is `DET`, its morphological annotation contains `PronType=Prs`, its syntactic function is `nmod`, and it is not modified by a word whose lemma is *de* (syntactic function `case`). These two adpossession constructions are classified as dependent marked.

In adpossession constructions with a noun possessor, the POS of the possessor is `NOUN` or `PROPN`, its syntactic function is `nmod`, and it is modified by a word whose lemma is *de* (syntactic function `case`). These adpossession constructions are classified as dependent marked.

## 2.33 Romanian

There are two treebanks for Romanian in the UD and they differ significantly from one another. The RRT treebank is for standard Romanian and the Nonstandard treebank is for a non-standard variety of Romanian. For this reason we have written two separate scripts (`RomanianHYPHENNonstandard.py`, `RomanianHYPHENRRT.py`) to identify and classify adpossession constructions in these treebanks. Our reference source for standard Romanian was Mallinson (1986), and for non-standard Romanian we consulted Andrei Dumitrescu.

In the Romanian-Nonstandard treebank, adpossession constructions were identified in the following way, and all of them were classified as dependent marked:

- 1) The POS of the possessor is `PRON`, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is either `det` or `nmod`.
- 2) The POS of the possessor is `PRON`, its morphological annotation contains `PronType=Prs`, no `Reflex=Yes`, and either `Case=Gen` or `Case=Dat, Gen`, and its syntactic function is `nmod`.
- 3) The POS of the possessor is `DET`, its morphological annotation contains `[psor]` and `PronType=Prs` but no `PronType=Prs`, and its syntactic function is `det`.
- 4) The POS of the possessor is `NOUN` or `PROPN`, its morphological annotation contains `Case=Dat, Gen`, its morphological annotation is `nmod`, and it is not modified by a determiner whose lemma is *al* (POS `DET`; morphological annotation contains `Poss=Yes` and `PronType=Art`; syntactic function is `det`).
- 5) The POS of the possessor is `NOUN` or `PROPN`, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`.
- 6) The POS of the possessor is `NOUN` or `PROPN`, and it is modified by a determiner whose lemma is *-ul* or *lui* (POS `DET`; morphological annotation contains `Case=Dat, Gen` and `PronType=Art`; syntactic function `det`). In addition, the possessor should not be modified by a dependent whose lemma is *al* (POS `DET`; morphological annotation contains `Poss=Yes` and `PronType=Art`; syntactic function `det`).

In the Romanian-RRT treebank, adpossession constructions were identified in the following way, and all of them were classified as dependent marked:

<sup>10</sup>The criterion concerning prepositions is not verified in the LFG treebank. The reason for this is that in the LFG treebank the syntactic function of noun possessors is `nmod:poss` which identifies the construction as possessive, but in the SZ treebank the syntactic function of the possessor is `nmod` and other criteria must be used to exclude constructions which are not adpossession.

- 1) The POS of the possessor is PRON, its morphological annotation contains Case=Dat , Gen or Case=Gen, and its syntactic function is nmod.
- 2) The POS of the possessor is PRON, its morphological annotation contains Poss=Yes and PronType=Prs, and its syntactic function is either det or nmod.
- 3) The POS of the possessor is DET, its lemma is *lui*, its morphological annotation contains Poss=Yes, PronType=Prs and Person=, and its syntactic function is det.
- 4) The POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Dat , Gen or Case=Gen, its syntactic function is nmod, and it is not modified by a word whose lemma is *al* (POS DET; morphological annotation contains Poss=Yes and PronType=Prs; syntactic function is det).
- 5) The POS of the possessor is DET, its morphological annotation contains Poss=Yes and PronType=Prs, and its syntactic function is det.

In addition, the following three types of adpossession constructions in the Romanian-RRT treebank are classified as dependent marked:

- 1) The POS of the possessor is PRON, NOUN or PROPN, its syntactic function is nmod, and it is modified by an adposition whose lemma is *de* (POS ADP; morphological annotation contains Case=Acc; syntactic function case).
- 2) The POS of the possessor is NOUN or PROPN, its syntactic function is nmod, and it is modified by a determiner whose lemma is *al* (POS DET; morphological annotation contains Poss=Yes and PronType=Prs; syntactic function is det).
- 3) The POS of the possessor is NOUN or PROPN, its syntactic function is nmod, and it is modified by a determiner whose lemma is *lui* (POS DET; morphological annotation contains Case=Dat , Gen and PronType=Art; syntactic function det) and which is not modified by the determiner *al*.

## 2.34 Russian

There are two treebanks for Russian in the UD, and we have written our algorithm (`Russian.py`) so that it works for both treebanks despite some major differences between them.<sup>11</sup> Our reference source for Russian was Andrews (2004).

In adpossession constructions with a pronoun possessor, the POS of the possessor is DET and its syntactic function is det. In addition, the lemma of the pronoun must be one of the following Russian possessive pronoun lemmas: мой, твой, ее, его, наш, ваш, их, МОЙ, ТВОЙ, ЕЕ, ЕГО, НАШ, ВАШ, and ИХ. These adpossession constructions were classified as dependent marked.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen, its syntactic function is nmod, and it must not be modified by an adposition (POS ADP; syntactic function case). If the lemma and the surface form of the possessor are identical (after converting to lowercase and replacing the letter *ë* with *e*), the adpossession construction was classified as zero marked, otherwise as dependent marked. Russian has indeclinable nouns, so even a plural possessor can be zero marked.

## 2.35 Serbian

In Serbian (Alexander, 2006), adpossession constructions with a pronoun possessor are identified so that the POS of the possessor is PRON, its morphological annotation contains Poss=Yes, and its syntactic function is det. These adpossession constructions are classified as dependent marked.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen, and its syntactic function is nmod. In addition, the noun possessor should not be modified by a preposition (POS ADP; syntactic function case). If the lemma and the surface form of the possessor are identical, these adpossession constructions are classified as zero marked, otherwise as dependent marked.

In adpossession constructions with a possessive adjective, the POS of the possessor is ADJ, its morphological annotation contains Poss=Yes, and its syntactic function is amod. Adpossession constructions with possessive adjectives are classified as dependent marked.

<sup>11</sup>For instance, the lemmas in Russian-GSD are capitalized and in Russian-SynTagRus they are not. It might have been possible to encode a working algorithm for Russian-GSD into the tagchart. However, it was not possible to identify possessive pronouns from the tagging alone in the SynTagRus treebank, so instead we had to identify whether the lemma is a possessive pronoun. As implied before, we do that for both Russian treebanks despite that being necessary for just one.

## 2.36 Slovak

In Slovak (Mistrík, 1983), adpossession constructions with a pronoun possessor are identified so that the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs` and `no Reflex=Yes`, and its syntactic function is `det`. These adpossession constructions are classified as dependent marked.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROP, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`. In addition, the noun possessor should not be modified by a preposition (POS ADP; syntactic function `case`). If the lemma and the surface form of the possessor are identical, these adpossession constructions are classified as zero marked, otherwise as dependent marked.

We are aware that Slovak has possessive adjectives, but we were unable to identify them from the Slovak treebank using the UD annotation.

## 2.37 Slovenian

In Slovenian (Derbyshire, 1993), adpossession constructions with a pronoun possessor are identified so that the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs` and `no Reflex=Yes`, and its syntactic function is `det`. These adpossession constructions are classified as dependent marked.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROP, its morphological annotation contains `Case=Gen`, and its syntactic function is `nmod`. In addition, the noun possessor should not be modified by a preposition (POS ADP; syntactic function `case`). If the lemma and the surface form of the possessor are identical, these adpossession constructions are classified as zero marked, otherwise as dependent marked.

In some adpossession constructions the possessors are possessive adjectives. The POS of their possessor is ADJ, their morphological annotation contains `Poss=Yes`, and their syntactic function is `amod`. These constructions are classified as dependent marked.

## 2.38 Spanish

Adpossession constructions can be identified in two ways in Spanish. Spanish has possessive pronouns like *mi* ‘my (singular head)’ and *sus* ‘his, her, their, one’s, your (formal) (all with plural head)’. With noun possessor the preposition *de* ‘of’ is used. If *de* should be followed by the masculine singular definite article *el*, they will be contracted to *del*. The following criteria can be used for identifying adpossession constructions regardless of the treebank available for Spanish.

In adpossession constructions with a pronoun possessor, the possessor can be identified with a morphological tag that contains `Poss=Yes` and `PronType=Prs`; in principle the POS and syntactic function tags are irrelevant in their identification. In adpossession constructions with a noun possessor, the POS is NOUN or PROP, its syntactic function is `nmod`, and it must be modified by the preposition *de* (POS ADP; syntactic function `case`). The lemma of the preposition can also be *del* as this contraction is not recognized as such in the treebanks. All adpossession constructions in Spanish are classified as dependent marked.

## 2.39 Swedish

There are two treebanks available for Swedish in the UD. Although some differences exist between the treebanks, we wrote the algorithm (`Swedish.py`) so that it works for both.<sup>12</sup>

In adpossession constructions, the POS of the possessor is NOUN or PROP for noun possessors or PRON for pronoun possessors, and the syntactic function is `nmod:poss`. The morphological annotation for pronoun possessors contains `Poss=Yes` and for noun possessors `Case=Gen`. In the Swedish treebanks we identified possessive constructions also by using the language-specific morphological annotation: if it contained a block GEN, when splitting from either the character | or - (| is used in the Swedish-Talbanken treebank and - is used in LinES), then the construction was classified as an adpossession construction. The constructions were classified as zero marked if the lemma and the surface form were identical, otherwise they were classified as dependent marked.

There is also another adpossession construction frequently used in Swedish. In these constructions the possessor is marked with the preposition *av* ‘of’, similarly to Danish and Norwegian. In these adpossession constructions, the POS of the possessor is NOUN, PROP, or PRON, its syntactic function is `nmod`, and the possessor is modified by the preposition *av*. These constructions are all dependent marked.

<sup>12</sup>Note that in our analyses, Swedish and Finnish are the two languages where we were able to identify `flats` correctly; the source code in the file `Swedish.py` contains some more information about this process. For most treebanks we did not analyze `flats` systematically at all.

## 2.40 Turkish

Adposessive constructions in Turkish (Underhill, 1976) were identified in the following way. In adposessive constructions with a pronoun possessor, the POS of the possessor is PRON, its morphological annotation does not contain `Reflex=Yes`, and its syntactic function is `nmod:poss`. In addition, if the morphological annotation of the possessee (the head) contains `[psor]`, the construction was classified as double marked, otherwise as dependent marked.

In adposessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Person` and either `Case=Gen` or `Case=Nom`. The morphological marking of these constructions was classified as follows:

- 1) If the morphological annotation of the dependent (possessor) contains `Case=Gen` and that of the head (possee) contains `[psor]`, the construction was classified as double marked.
- 2) If the morphological annotation of the head does not contain `[psor]`, the construction was classified as dependent marked.
- 3) If the morphological annotation of the dependent contains `Case=Nom` and that of the head contains `[psor]`, the construction was classified as head marked.
- 4) If the morphological annotation of the head does not contain `[psor]`, the construction was classified as zero marked.

In addition, adposessive constructions in which the POS of the possessive is NOUN or PROPN, whose morphological annotation contains `[psor]`, and which does not have a dependent with syntactic function `nmod:poss` were classified as head marked. More precisely, these adposessive constructions only have the head noun, but the possessor is morphologically marked only on the head and the possessor is not marked with a separate dependent word. We label these constructions as `head_exist` constructions in the algorithm.<sup>13</sup>

## 2.41 Ukrainian

In Ukrainian (Pugh and Press, 1999), adposessive constructions with a pronoun possessor are identified as constructions in which the POS of the possessor is DET, its morphological annotation contains `Poss=Yes` and `PronType=Prs` but no `Reflex=Yes`, and its syntactic function is `det`. All adposessive constructions with pronoun possessors were classified as dependent marked.

In adposessive constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its morphological annotation contains `Case=Gen`, its syntactic function is `nmod`, and it should not be modified by an adposition (POS is ADP and syntactic function is `case`). These constructions were classified as zero marked if the lemma and the surface form were identical, otherwise they were classified as dependent marked.

## 2.42 Urdu

In Urdu (Schmidt, 1999) adposessive constructions with pronoun possessor are identified in two ways. First, if the POS of the possessor is PRON, its morphological annotation contains `Poss=Yes` and `PronType=Prs`, and its syntactic function is `nmod`, the construction was classified as dependent marked. Second, pronoun possessors (POS is PRON, morphological annotation contains `PronType=Prs`, and syntactic function is `nmod`) may be dependent marked if they are modified by an adposition whose lemma is `کا`, `کے`, or `کی`.

In adposessive constructions with noun possessors, the POS of the possessor is NOUN or PROPN, its syntactic function is `nmod`, and it is modified by an adposition whose lemma is `کا`, `کے`, or `کی`. All these constructions are classified as dependent marked.

## 2.43 Uyghur

In the following we describe how adposessive constructions were identified and classified using the Uyghur treebank. Our main reference source was with Ablahat Ibrahim (1991).

Adposessive constructions with a pronoun possessor are identified and classified in two ways.

---

<sup>13</sup>Although our algorithm can detect dependent marked and zero marked constructions in Turkish, no such constructions were identified in the treebank. According to (Underhill, 1976, 92–93) and personal communication with N. N. , it is sometimes possible in colloquial Turkish to omit the possessive suffix, thus leading to adposessive constructions with no head marking. Such patterns occur especially in constructions with proper noun heads such as “My Istanbul”. However, it is not possible to accurately identify such constructions from the Turkish treebank.

- 1) If the POS of the possessor is PRON, its morphological annotation contains Case=Gen and PronType=Prs but no Reflex=Yes, and its syntactic function is nmod:poss, the adpossession construction is classified as dependent marked.
- 2) If the POS of the possessor is PRON, its morphological annotation contains Case=Nom and PronType=Prs but no Reflex=Yes, and its syntactic function is nmod:poss, the adpossession construction is classified as zero marked.

Adpossession constructions with a noun possessor are identified and classified in the following ways.

- 1) If the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen, its syntactic function is nmod:poss, and the morphological annotation of its head (the possessee) contains [psor], the adpossession construction was classified as **double marked**.
- 2) If the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Nom, its syntactic function is nmod:poss, and the morphological annotation of its head (the possessee) contains [psor], the adpossession construction was classified as **head marked**.
- 3) If the POS of the **possessee** is NOUN or PROPN, its morphological annotation contains [psor], and it is not modified by a dependent whose syntactic function is nmod:poss, the adpossession construction was classified as **head\_exist**, that is, as an adpossession construction in which the possessor is only marked on the head and not as a separate dependent word.
- 4) If the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Gen, its syntactic function is nmod:poss, and the morphological annotation of its head (the possessee) does not contain [psor], the adpossession construction was classified as **dependent marked**.
- 5) If the POS of the possessor is NOUN or PROPN, its morphological annotation contains Case=Nom, its syntactic function is nmod:poss, and the morphological annotation of its head (the possessee) does not contain [psor], the adpossession construction was classified as **zero marked**.

## 2.44 Vietnamese

It seemed to us that when analyzing adpossession constructions with pronoun possessors in Vietnamese, we had to separately list the pronouns, since their POS was PROPN and not e.g. PRON as expected. For this reason we wrote a separate Python script (`Vietnamese.py`) to identify and classify adpossession constructions in Vietnamese treebank. Our main reference source was Thompson (1965).

In adpossession constructions with a pronoun possessor, the POS of the possessor is PROPN and its syntactic function is det. In addition, the lemma had to be one of the following: *anh, ấy, chúng, chúng ta, chúng tôi, hắn, họ, mày, mình, nó, ông, ta, tôi, tui* and *y*. If the possessor was also modified by the preposition *của* (POS ADP; syntactic function case), then the adpossession construction was classified as dependent marked, otherwise as zero marked.

In adpossession constructions with a noun possessor, the POS of the possessor is NOUN or PROPN, its syntactic function is nmod, and it is modified by the preposition *của* (POS ADP; syntactic function case). These adpossession constructions are classified as dependent marked.<sup>14</sup>

## 3 Some further methodological issues

### 3.1 Algorithm to identify conjoined phrases

Unless otherwise stated above (cf. German genitive constructions and Latvian as a whole) we identify conjoined possessors, or conjs (syntactic function conj) in the following way. If the type of possessive construction has no adposition or other function word that serves the same function as English *of*, the conjs can be identified as words that otherwise would satisfy the conditions of some non-adpositional way of identifying possessive NPs. In other words, after all non-conjs have been checked from the sentence, any conjs whose head (the conjunct in the NP that doesn't have conj as its syntactic function) has already been classified as a dependent of a possessive NP (so it doesn't matter if it has been classified as a head of a head\_exist NP), will be evaluated if they satisfy any set of conditions where the requirement of the syntactic function is not considered.

<sup>14</sup>We expect some adpossession constructions with noun possessors may be zero marked, but we were unable to identify them, since not all nouns that were modified by another noun were adpossession constructions.



For adpositional conjs, our algorithm only allows collecting conjs from one condition, leaving pronouns that are conjoined with nouns or vice versa out from most languages with a possessive adposition. If the first conjunct word is not a possessor, it can even be proven that its conjuncts that are possessors are impossible to identify. For example, the sentence 9\_PS-306 in the Italian-ISDT treebank contains the NP *la salute e la nutrizione materna e del bambino* ‘maternal and child health and nutrition’, literally word by word: “the health and the nutrition maternal and of the child”. In this construction the syntactic function of the only possessor is conj as the word *materna*, with which it is conjoined, is an adjective whose syntactic function is amod. Since the annotations do not contain information of that conj’s “actual” syntactic function, conj possessors like that are therefore impossible to identify, so even though our algorithm is decent but obviously not perfect, that example proves that there is no way it even could be perfect.

Whenever a dependent with a possessive adposition is detected, all conjs from that sentence are checked if they satisfy the same conditions (except for the syntactic function) that led to the detection of the first dependent. An important difference between adpositional possessive NPs and others is that adpositional NPs are always dependent marked (the adposition is the marker) while conjunct dependents of non-adpositional NPs can at least logically belong to any class, except (of course) to head\_exist.

Whenever a conj is identified, its head will be marked as the head of the first of the conjunct dependents, which is different from the one given in the UD. This is needed in order to calculate the dependency length correctly.

### 3.2 On the definition of dependency length

We computed dependency length in two alternative ways. First, our algorithm defines the *dependency length* of an adnominal possessive construction as the absolute difference of the position of the head in the sentence and the position of the dependent subtracting the number of tokens for PUNCT in between the head and the dependent. Dependency length is undefined for constructions that contain only the head, which we label head\_only or head\_exist. This dependency length is labelled length1 in the result file FullNPPossData.txt. Second, we provided also another way for computing dependency length, labelled length2 in the result file FullNPPossData.txt. In this method we also took into account that an adposition, particle, or another word that intervenes between the head and the dependent may also mark morphologically the syntactic relation between the head and the dependent in the adpossession construction. We subtracted such words from the total number of intervening words between the head and dependent in adpossession constructions. This method for dependency length was suggested to us as an alternative to counting all intervening words between the head and the dependent.

## References

- Ronelle Alexander. *Bosnian, Croatian, Serbian, a Grammar With Sociolinguistic Commentary*. The University of Wisconsin Press, Madison, WI, 2006.
- Philip Holmes Robin Allan and Tom Lundskær-Nielsen. *Danish: an essential grammar*. Routledge, London and New York, 2000.
- Edna Andrews. *Russian*. SEELRC, 2004.
- Evgenia Antova, Ekaterina Boytchinova, and Poly Benatova. *A short grammar of Bulgarian for English speaking learners*. ABM Komers, Sofia, second edition, 2002.
- R. E. Batchelor and M. Chebli-Saadi. *A Reference Grammar of French*. Cambridge University Press, Cambridge, 2011.
- C. E. Bennett. *A Latin Grammar*. Allyn and Bacon, Boston and Chicago, second edition, 1908.
- Maria Carmela Benvenuto and Flavia Pompeo. Verbal semantics in ancient greek possessive constructions with einai. *Journal of Greek Linguistics*, 15(1):3–33, 2015. URL [https://brill.com/view/journals/jgl/15/1/article-p3\\_2.xml](https://brill.com/view/journals/jgl/15/1/article-p3_2.xml).
- Maria Zagorska Brooks. *Polish Reference Grammar*. Mouton, The Hague, 1975.
- Suk-Jin Chang. *Korean*. John Benjamins, Philadelphia, 1996.
- William W. Derbyshire. *A Basic Reference Grammar of Slovene*. Slavica, Columbus, OH, 1993.
- Bruce Donaldson. *Dutch: A Comprehensive Grammar, 3rd edition*. Routledge, London, 2017.

- Bruce C. Donaldson. *A Grammar of Afrikaans*, volume 8 of *Mouton Grammar Library*. Mouton de Gruyter, Berlin, 1993.
- Trevor G. Fennell and Henry Gelsen. *A grammar of modern Latvian: Vol 1-3*, volume 304 of *Slavistic Printings and Reprintings*. Mouton de Gruyter, The Hague, 1980.
- Hassan A. H. Gadalla and Abdel-Nasser M. E. Abdel-Hamid. Genitive constructions in english and arabic: A contrastive study. *Bulletin of the Faculty of Arts*, 6:1–64, 2000. URL [https://www.academia.edu/208345/Genitive\\_Constructions\\_in\\_English\\_and\\_Arabic\\_A\\_Contrastive\\_Study](https://www.academia.edu/208345/Genitive_Constructions_in_English_and_Arabic_A_Contrastive_Study).
- Lewis Glinert. *The grammar of modern Hebrew*. Cambridge University Press, New York, 1989.
- William W. Goodwin. *Greek Grammar*. Ginn & Company, Boston, 1900.
- Wayne Harbert. *Gothic Syntax: A relational grammar*. PhD dissertation, Urbana-Champaign, 1978.
- William E. Harkins. *A Modern Czech Grammar*. King's Crown Press, New York, 1953.
- Martin Haspelmath. Explaining alienability contrasts in adposessive constructions: Predictability vs. iconicity. *Zeitschrift für Sprachwissenschaft*, 36(2):193–231, 2017.
- John Hinds. *Japanese*. Croom Helm Descriptive Grammars. Croom Helm, London, 1986.
- José I. Hualde. *Catalan*. Croom Helm Descriptive Grammars. Routledge, London & New York, 1992.
- Amélia P. Hutchinson and Janet Lloyd. *Portuguese: an essential grammar 2nd edn*. Routledge, London, 2003.
- Brian D. Joseph and Irene Philippaki-Warbuton. *Modern Greek*. Croom Helm Descriptive Grammars. Routledge, London, 1987.
- Yamuna Kachru. *Hindi*, volume 12 of *London Oriental and African language library*. Benjamins, Philadelphia, 2006.
- István Kenesei, Robert M. Vago, and Anna Fenyesi. *Hungarian*. Descriptive Grammars Series. Routledge, London & New York, 1998.
- Jan Krivan. The role of information structure in czech possessive constructions, 2013. URL [http://ulug.ff.cuni.cz/lingvistika/krivan/veda/Krivan\\_2013\\_olinco\\_role-inf-str.pdf](http://ulug.ff.cuni.cz/lingvistika/krivan/veda/Krivan_2013_olinco_role-inf-str.pdf).
- Gilbert Lazard. *A Grammar of Contemporary Persian*. Mazda Publishers, Costa Mesa, CA, 1992.
- Horace G. Lunt. *Old Church Slavonic grammar*. Mouton de Gruyter, Berlin, seventh edition, 2001.
- Yongxian Luo. Possessive constructions in mandarin chinese. In Alexandra Y. Aikhenvald and R. M. W. Dixon, editors, *Possession and Ownership: A Cross-Linguistic Typology*, pages 186–207. Oxford University Press, Oxford, 2012.
- Martin Maiden and Cecilia Robustelli. *A reference grammar of modern Italian*. NTC, Lincolnwood, Ill., second edition, 2007.
- Graham Mallinson. *Rumanian*. Croom Helm Descriptive Grammars. Croom Helm, London, 1986.
- Jozef Mistrik. *A Grammar of Contemporary Slovak*. Slovenské pedagogické nakladateľstvo, Bratislava, 1983.
- Liljana Mitkovska. Expressing prototype possessive relations at the noun phrase level in macedonian and bulgarian. In Mila Dimitrova-Vulchanova and Olga Mišeska Tomić, editors, *Investigations in the Bulgarian and Macedonian Nominal Expression*, pages 121–147. Tapir Academic Press, Trondheim, 2009.
- James Naughton. *Czech: an essential grammar*. Routledge, London, 2005.
- Johanna Nichols. *Linguistic Diversity in Space and Time*. University of Chicago Press, Chicago, 1992.
- Motoki Nomachi. Dative of external possession in croatian: From an areal-typological perspective. *Jezikoslovlje*, 17:1–2, 453–474, 2016. URL <https://hrcak.srce.hr/file/247526>. URL is force download.
- Albert Ortmann. Connecting the typology and semantics of nominal possession: alienability splits and the morphology–semantics interface. *Morphology*, 28(1):99–144, 2018.

- Stefan M. Pugh and Ian Press. *Ukrainian: A Comprehensive Grammar*. Routledge Grammars. Routledge, London & New York, 1999.
- Karin C. Ryding. *A reference grammar of modern standard Arabic*. Cambridge Univ. Press, New York, 2005.
- Mario Saltarelli. *Basque*. Croom Helm Descriptive Grammars. Croom Helm, New York, 1988.
- Ernest Scatton. Bulgarian. In Bernard Comrie and Greville Corbett, editors, *The Slavonic Languages*, pages 188–248. Routledge, London / New York, 1993.
- Ruth Laila Schmidt. *Urdu: an essential grammar*. Routledge, London-New York, 1999.
- James N. Sneddon. *Indonesian: A Comprehensive Grammar*. Routledge Grammars Series. Routledge, London & New York, 1996.
- James Neil Sneddon, Alexander Adelaar, Dwi Noverini Djenar, and Michael C. Ewing. *Indonesian Reference Grammar*. Allen & Unwin, second edition, 2010.
- Laurence C. Thompson. *A Vietnamese Grammar*. University of Washington Press, Seattle, 1965.
- Robert Underhill. *Turkish Grammar*. MIT Press, Cambridge, MA, 1976.
- Reinhard F. Hahn with Ablahat Ibrahim. *Spoken Uyghur*. University of Washington Press, Seattle, WA, 1991.