Typology of Possessive Constructions: Comparison of Head-Marking and Dependent-Marking Patterns

Anton Buzanov

MA Thesis. Linguistic Theory and Language Description

June 13

Possessive Constructions

Possessive constructions are defined as comprising two noun phrases: the possessor and the possessum. Their primary function is to convey possessive relationships, which are broadly understood.

(1) Russian (Indo-European)

dom-ø   otc-a
house-NOM.SG father-GEN.SG

'father's house'

(2) Even (Tungusic)

etiken   žu-n
old.man house-POSS.3SG

'old man's house'

Locus of Marking Typology

Based on the specific locus of marking within possessive noun phrases, languages can be distinctly categorized into several groups (Nichols and Bickel, 2013; Van Rijn, 2016a; Van Rijn, 2016b).

1. Head marking (78)
2. Dependent marking (98)
3. Double marking (22)
4. No marking (32)
5. Other (6)

Numbers are from Nichols and Bickel, 2013 (WALS sample).

Locus of Marking Typology: Key Principle

Nichols and Bickel, 2013 stick to the following principles:

1. Selecting one construction per language (exemplar-based survey)
2. Excluding constructions that do not allow overtly expressed possessors
3. Excluding consideration of personal pronouns if their behaviour is different from nouns

Locus of Marking Typology: Potential Problems

Mongolian is categorized as dependent marking in terms of Nichols and Bickel, 2013.

(3)   Mongolian (Mongolic) from Janhunen, 2012

    a.   (*min-ii) duu-men'
       1SG-GEN   younger.brother-POSS.1

       'my younger brother'

    b.   min-ii   eej(-*men')
       1SG-GEN   mother-POSS.1

       'my mother'

Locus of Marking Typology: Potential Problems

This approach cannot capture facts provided by different constructions within the same language.

Moreover, it cannot capture the variation within one construction.

(4) Even

    munnukan təwtə-*(n)
    hare   berry-POSS.3SG

    'cranberry' (lit. 'hare's berry')

(5) Kildin Saami

    mun kn'iga-(n)
    I.GEN book-POSS1

    'my book'

My Approach

The aim of my thesis was to create a metalanguage for a holistic description of possessive constructions in the languages of the world.
(Thanks to Mikhail Aleksandrovich for a nice wording.)

My Approach

I did not limit myself to choosing the single construction per language.

I employ a bottom-up strategy, initially categorizing markers as either C- or D-marking (Lander and Nichols, 2020), and subsequently classifying a language as a collection of pairs of these markers.

- ▶ D-markers = dependent markers
- ▶ C-markers = anywhere-but-dependent markers (Markers may occupy different places within the whole construction)

Russian would be characterized as a language featuring several D-markers, such as GEN and possessive forms of (pro)nouns, while lacking any C-markers.

My Approach

The overall list of questions is as follows. The answer to each question is "yes" if any construction in the language allows it.

1. Is juxtaposition possible?
2. Can C-marking be used without an overt possessor?
3. Can C-marking be used with an unmarked possessor?
4. Can D-marking be used without C-marking?
5. Can D-marking be used with an unmarked possessum?

Applying the Approach

Constructing a **variety sample** is essential.

Applying the Approach

Constructing a **variety sample** is essential.

The aim of a variety sample is to capture a broad spectrum of linguistic patterns with minimal effort, as outlined by Miestamo et al., 2016.

Applying the Approach

Constructing a **variety sample** is essential.

The aim of a variety sample is to capture a broad spectrum of linguistic patterns with minimal effort, as outlined by Miestamo et al., 2016.

Important terms:

- ▶ A **genus** refers to a grouping of languages with a time depth not exceeding 3,500 to 4,000 years (Dryer, 1989).
- ▶ **Macroareas** are continent-size linguistic areas which are independent of each other, but within which languages are to some extent typologically similar (Miestamo et al. 2016 after Dryer 1989).

Applying the Approach

Miestamo et al., 2016 suggested the Genus-Macroarea approach for variety sampling.

- ▶ Count genera in each macroarea
- ▶ Take as many languages as you need for your sample keeping the overall Genus-Macroarea proportion

Applying the Approach

Problems:

- ▶ Genera are systematically assigned only in WALS (Dryer and Haspelmath, 2013).
- ▶ Merging WALS data with Glottolog data results in coverage of only 28.1% with genera.

I augmented the number of languages with known genera, so the total is **75.8**% (6,506 languages).

Sample

I modified the algorithm of automatic sample generation proposed by Cheveleva, 2023 and came up with the sample of 23 languages.

- **Papunesia**: Marind (Marind-Yaqay), To'abaita (Oceanic), Lundayeh (North Borneo), Kobon (Kalam-Kobon), West Coast Bajau (Sama-Bajaw), Iloko (Northern Luzon)
- **South America**: Chácobo (Panoan), Hup (Nadahup), Yucuna (Japura-Colombia), Kwaza (isolate)
- **Eurasia**: Even (Tungusic), Kildin Saami (Saami), Abaza (Northwest Caucasian), Russian (Slavic), Mongolian (Mongolic)
- **North America**: Poqomam (Mayan), Central Alaskan Yupik (Eskimo), Haida (Haida)
- **Africa**: Ewe (Gbe), Ruund (Bantu), Paku (Barito), Lamang (Biu-Mandara)
- **Australia**: Ngardi (Western Pama-Nyungan)

Results: Languages with Single Type of Marking

language	juxtaposition	C-marker without possessor	C-marker with unmarked possessor	C-marker & D-marker	D-marker with unmarked possessum
Ruund	-	NA	NA	NA	+
Russian	-	NA	NA	NA	+
Kwaza	-	NA	NA	NA	+
Paku	+	+	-	NA	NA
WC Bajau	+	+	-	NA	NA
Iloko	+	+	-	NA	NA
Lamang	-	+	+	NA	NA
Abaza	-	+	+	NA	NA
Poqomam	+	+	+	NA	NA
To'abaita	+	+	+	NA	NA
Kobon	+	+	+	NA	NA

Results: Languages with Single Type of Marking

- ► Only three languages – Russian, Ruund, and Kwaza – do not use any form of C-marking in the possessive domain. These languages do not allow juxtaposition to convey possessive meaning as well.
- ► No language exclusively allows a C-marker without an overt possessor; instead, juxtaposition and/or using the unmarked possessor are always options.
  - ► Some languages allow juxtaposition while disallowing an expressed possessor
  - ► Others disallow juxtaposition but permit an overt possessor
  - ► Some languages allow both.

Results: Languages with Two Types of Marking

language	juxtaposition	C-marker without possessor	C-marker with unmarked possessor	C-marker & D-marker	D-marker with unmarked possessum
Haida	-	+	+	-	+
Hup	-	+	+	-	+
CA Yupik	-	+	-	+	+
Chacobo	-	+	-	+	+
Kildin Saami	-	+	-	+	+
Even	-	+	-	+	-
Ngardi	-	+	-	-	+
Marind	+	+	+	+	+
Mongolian	+	+	-	-	+
Lundayeh	+	+	-	-	+
Yucuna	+	+	-	-	+
Ewe	+	-	+	-	+

Results: Languages with Two Types of Marking

- ▶ Typically, languages disallow the use of a C-marker with an unmarked possessor, given that D-marking is present in the language, with two exceptions.
    1. If a language does not allow a C-marker without an overt possessor, it allows a C-marker with an unmarked possessor.
    2. If a language disallows both juxtaposition and the expression of both C- and D-markers, it allows a C-marker with an unmarked possessor.
- ▶ Typically, languages allow using D- and C-markers to function on their own at least for some lexemes or constructions.

Results: Languages with Two Types of Marking

- ▶ Typically, languages disallow the use of a C-marker with an unmarked possessor, given that D-marking is present in the language, with two exceptions.
    1. If a language does not allow a C-marker without an overt possessor, it allows a C-marker with an unmarked possessor.
    2. If a language disallows both juxtaposition and the expression of both C- and D-markers, it allows a C-marker with an unmarked possessor.
- ▶ Typically, languages allow using D- and C-markers to function on their own at least for some lexemes or constructions.
    - ▶ Two exceptions are Even and Ewe (don't mix up the two!!).

Further Work

- ► Incorporate the data on indexing and registration at more general level.
  - ► Some patterns could be described with reference to this distinction.
  - ► However, it might be a part of some other phenomenon, e. g. the ability to identify features of a participant from the context.
- ► Investigate the juxtaposition strategy in more detail.
  - ► Juxtaposition in my sample is treated inconsistently.
  - ► Perhaps, it must be seen as an instance of head or dependent marking.
- ► Make a larger sample

Conclusions

- ▶ Primary set of criteria for systematic annotation of possessive constructions.
- ▶ Expanded Glottolog classifications using WALS data, annotating over 6,000 languages.
- ▶ Selected 23 genera for a pilot sample using the Genus-Macroarea method.
- ▶ Manually described the possessive system for one language from each genus.
- ▶ Identified spectrum of marking strategies.
- ▶ Analysis underscores the complexity and diversity of possessive marking strategies.
- ▶ Trends identified provide a foundation for future statistical analyses on larger samples.

References

[1] Anastasia Cheveleva. "Neutralization of Gender Values in the Plural". BA Thesis. HSE University, 2023.

[2] Matthew S Dryer. "Large linguistic areas and language sampling". In: *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 13.2 (1989), pp. 257–292.

[3] Matthew S Dryer and Martin Haspelmath, eds. *WALS Online (v2020.3)*. Zenodo, 2013. DOI: 10.5281/zenodo.7385533. URL: https://doi.org/10.5281/zenodo.7385533.

References

[4] Juha A Janhunen. *Mongolian*. Vol. 19. John Benjamins Publishing, 2012.

[5] Yury Lander and Johanna Nichols. "Head/dependent marking". In: *Oxford research encyclopedia of linguistics*. 2020.

[6] Matti Miestamo, Dik Bakker, and Antti Arppe. "Sampling for variety". In: *Linguistic Typology* 20.2 (2016), pp. 233–296.

References

[7]   Johanna Nichols and Balthasar Bickel. "Locus of Marking in Possessive Noun Phrases". In: *The World Atlas of Language Structures Online*. Ed. by Matthew S Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. URL: https://wals.info/chapter/24.

[8]   Anna Szabolcsi. "The possessive construction in Hungarian: A configurational category in a non-configurational language". In: *Acta Linguistica Academiae Scientiarum Hungaricae* 31.1/4 (1981), pp. 261–289.

References

[9]   Marlou Van Rijn. "Locus of marking typology in the possessive NP: A new approach". In: *Folia Linguistica* 50.1 (2016), pp. 269–327.

[10]   Marlou Van Rijn. "The grammaticalization of possessive person marking: A typological approach". In: *Transactions of the Philological Society* 114.2 (2016), pp. 233–276.

Locus of Marking Typology: Potential Problems

Hungarian is categorized as head marking in terms of Nichols and Bickel, 2013.

(4)   Hungarian adapted from Szabolcsi, 1981, p. 263

    az   én kar-ja-i-m
    the I   arm-POSS-PL-1SG

    'my arms'

Locus of Marking Typology: Potential Problems

Hungarian is categorized as head marking in terms of Nichols and Bickel, 2013.

(4)   Hungarian adapted from Szabolcsi, 1981, p. 263

      az   én   kar-ja-i-m
      the   I   arm-POSS-PL-1SG

      'my arms'

én is the nominative form: it is used for marking the subject.

Locus of Marking Typology: Potential Problems

Hungarian is categorized as head marking in terms of Nichols and Bickel, 2013.

(4)   Hungarian adapted from Szabolcsi, 1981, p. 263

    az   én   kar-ja-i-m
    the I   arm-POSS-PL-1SG

    'my arms'

én is the nominative form: it is used for marking the subject.

Is this form truly unmarked? For me, the answer is no. It is **marked with nominative** since there are other cases in Hungarian.

Locus of Marking Typology: Potential Problems

In languages with case systems, cross-linguistic variation determines whether a dedicated genitive form (as in Russian), a morphologically unmarked form (nominative – as in Hungarian), or both (Turkish), can be used to indicate possession.

However, from a syntactic perspective, each form is assigned a case value.

Locus of Marking Typology: Potential Problems

In languages with case systems, cross-linguistic variation determines whether a dedicated genitive form (as in Russian), a morphologically unmarked form (nominative – as in Hungarian), or both (Turkish), can be used to indicate possession.

However, from a syntactic perspective, each form is assigned a case value.

Thus, the notion of unmarkedness must be rethought here.

My Approach: Problems with Juxtaposition

I acknowledge that my understanding of juxtaposition is inconsistent throughout my thesis. However, the concept itself might be problematic and requires a lot of further investigation.

POSSESSOR + POSSESSUM is considered juxtaposition if both the possessor and the possessum are morphologically unmarked, and one of the following conditions is met:

- ▶ The language does not have cases.
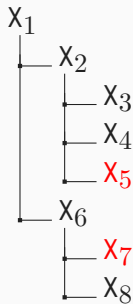- ▶ There is an option to morphologically mark the possessor.

Applying the Approach

- ► Genera are systematically assigned only in WALS (Dryer and Haspelmath, 2013).
- ► Merging WALS data with Glottolog data results in coverage of only 28.1% with genera.
- ► There is potential to increase the number of languages with known genera.

Applying the Approach

- ▶ Genera are systematically assigned only in WALS (Dryer and Haspelmath, 2013).
- ▶ Merging WALS data with Glottolog data results in coverage of only 28.1% with genera.
- ▶ There is potential to increase the number of languages with known genera.

X_1
X_2
X_3
X_4
X_5
X_6
X_7
X_8

Applying the Approach

▶ Genera are systematically assigned only in WALS (Dryer and Haspelmath, 2013).

▶ Merging WALS data with Glottolog data results in coverage of only 28.1% with genera.

▶ There is potential to increase the number of languages with known genera.

X_1
X_2
X_3
X_4
X_5
X_6
X_7
X_8
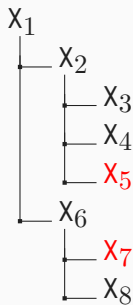
After implementing the algorithm and accounting for isolates not included in WALS, the coverage increased significantly to **75.8** percent. In practical terms, this means that I successfully annotated **6,506** languages from Glottolog with their respective genera.

Algorithm

► Modified the algorithm proposed by Cheveleva, 2023.

► Retained automatic suggestion of genera but introduced a more manual selection process.

► Algorithm now suggests genera for closer examination, empowering linguists in language selection.

► Created individual tables for each genus:
  ► Lengths of available grammars.
  ► Rationale for assigning languages to specific genera.

► Included details on:
  ► Common ancestor when assigning based on related languages.
  ► Related languages for languages sharing a common ancestor.

Ewe

Ewe possessive constructions can be expressed through several methods:

- ► **Possessive Linker**: This method involves a possessive linker *fe*, and is used in the 'alienable nominal construction' where the structure is $NP_{possessor}$ fe $NP_{possessum}$.
- ► **Juxtaposition**: This method, known as the 'inalienable nominal construction', also follows the structure $NP_{possessor}$ $NP_{possessum}$ without any connective.
- ► **Syntactic Compounding**: In this structure, the two nominals are compounded and marked with a high tone suffix at the end: $N_{possessor}$-$N_{possessum}$ + HIGH TONE SUFFIX.

Kwaza

The possessor – the modifier in the Kwaza possessive construction – is a personal pronoun or a noun. These constructions require a derivational possessive morpheme *-dy-*, which must be applied to the possessor and which must be followed by a classifier, usually the nominaliser *-hy*, which functions as a semantically neutral classifier, as exemplified in (5).

(5)   'si-dy-hy   ecui'ri
       I-POS-NOM buttock

     'my buttock'

Kwaza

Sometimes a choice is possible between a Neutral or a more specific classifier. When the specific classifier is etymologically related to the referent it classifies, the referent may be omitted, as in (6), as if the classifier were a cross-reference morpheme.

(6)   (a'xy)   'si-dy-xy
      (house) I-POS-CL:house

      'my house'

This is the only language in the sample that allows for omitting the possessum. From one hand, it could be analyzed in a way that the classifier morpheme is the possessum itself; however, it seems odd for me.

Other Parameters

There are three other parameters that worth mentioning:

1. Indexation properties: C-markers often vary based on the φ-features of the possessor, while D-markers typically do not.

2. Omission properties: C-markers can be used with omitted possessors, whereas D-markers cannot be used with omitted possessums.

3. Reflexive properties: Reflexive C-markers are rarely found.

Other Parameters

C-markers typically express the person value of the possessor, often reflecting number as well, i. e. they are indexing. Among languages in the sample, only three languages have invariable C-markers: Ewe, Ngardi, and Haida. However, in Ngardi and Haida, these markers are almost always used in contexts of third-person reference, except when the possessor is obvious from the previous context. This implies that these markers are *almost* indexing.

In contrast, D-markers are almost never indexing but rather registering. Among the languages in the sample, only Ruund and Kwaza have indexing D-marking strategies, both employing possessive classifiers.

Other Parameters

This difference leads to distinct patterns for C- and D-markers. Almost all C-markers can be used without an explicitly mentioned possessor, as they already index all its features. Conversely, only one D-marker can be used without an overt possessum: the one in Kwaza. Furthermore, possessums cannot be easily dropped without altering the distribution of the entire noun phrase since they are the heads. While it's possible to form a headless possessive noun phrase in contexts of ellipsis or predicative possession, I have not found mentions of omitting the possessum without changes in distribution or semantics for any language except Kwaza.