Functions of the 3sg Possessive in Beserman Udmurt: Corpus Analysis

Timofey Arkhangelskiy (National Research University Higher School of Economics)  
timarkh@gmail.com

Maria Usacheva (Moscow State University)  
mashastroeva@gmail.com

Research was supported by the RFH grant 16-24-17003 “Интегральный анализ именной группы в финно-угорских языках: поддержание референции и кодирование информационной структуры высказывания”
Possessives in Udmurt

• Udmurt has full paradigm of possessives

• 3sg possessive is special, as apart from having purely possessive functions (see Едыгарова 2010), it often means something other than a prototypical possessive relation

• Only 3sg possessive is used in case compounding and for marking contrastive focus on adjectives:

(1) kyz-ze / *kyz-de    kor-de    uli-ja-z    pun
    thick-P.3SG.ACC / thick-P.2SG.ACC  log-P.2SG.ACC  bottom-ILL-P.3SG  put.IMP

‘Put the thick log to the bottom.’
3SG possessive

• The fact that P.3SG has developed non-possessive meanings in Udmurt and other Uralic languages has been often mentioned in the literature
  • [Кузнецова 2012]
  • [Simonenko 2014] for Komi, Khanty and Mari
  • [Тáнчzos 2016]
  • Sometimes P.3SG evolves into a definiteness marker (cf. Eastern Armenian)
3SG possessive

• Example of a 3SG possessive with a discourse function:

(2) pios.murt-\textit{ez} baš’t-iz baton
    guy-\textit{p.3SG} take-\textit{PST.3SG} loaf

    ‘The guy took a loaf.’

• It can attach to different parts of speech, but we only looked at nouns (including relational nouns)
Other possessives

• 2SG possessives can also be used in non-possessive functions:

(3) vaj so-ize=no gozâ-de
    bring.IMP that-P.3SG.ACC=ADD rope-P.2SG.ACC

‘Bring me that rope as well’ [the rope does not belong to the addressee]

• And, occasionally, 1SG possessives as well:

(4) ad’ami-je kâč’eke korka pâr-iz
    person-P.1SG some house.ILL enter-PST.3SG

‘That man [I was talking about] went into some house.’

• We will not cover this topic here
What we did

• We want to know when and how often P.3SG marker has non-possessive functions and what are the factors that trigger its appearance
• We annotated 12 texts of different genres from the Beserman corpus (collected in Shamardan, Udmurtia, in 2003–2016), which gave us about 2,000 nouns
• Each noun was annotated with a number of parameters (features)
• Then machine learning algorithms were used to see how well this set of parameters predicts the appearance of P.3SG and which parameters are most important
What we did

• [Serdobolskaya, Toldova 2012] for Pechora Komi:
Features

- referential status (def, weak, ref_indef, indef, generic,...)
- semantic class (rn, bp, kin, ..., other)
- animacy
- alienability
- uniqueness
- syntactic position (subj, DO, oblique, NPDep, addr, pred)
- topic / focus
- referential distance, distance to first occurrence, topic persistence
- the case of the dependent (if any)
- protagonism
- possessive relation
NP structure

• Two types of NPs with nominal dependents are possible in Beserman:

• N N

(5) korka koš’ag
    house window

• N-GEN N-P.3SG

(6) korka-len koš’ag-ez
    house-GEN window-P.3SG
NP structure

• With relational nouns as heads, the (discourse) possessive marker in N N can attach to the head or the dependent:

(7) korka-je  dor-e /  korka  dor-a-m  
    house-P.1SG  near-ILL /  house  near-ILL-P.1SG
    ‘to my house’

• In N-GEN N-P.3SG, it can attach only to the dependent, as the slot on the head is already occupied:

(8) korka-je-len  dor-a-z  
    house-P.1SG-GEN  near-ILL-P.3SG
    ‘to my house’
NP structure

• We counted [N RelN] as one occurrence:

  korka  0
  korkajez  3sg
  korka+vəlen  0
  korkajez+vəlen  3sg
  korka+vəla\text{z}  3sg
  korkajezlen+vəla\text{z}  3sg
NP structure

• Ordinary nouns: 1468 total, 75% non-possessive
• RelNPs: 130 total, 77% non-possessive
• It seems that possessive marking is independent from the NP type and from the choice of the host within [N RelN]
• Annotating heads and dependents in RelNPs separately would skew the results
Distribution of possessives in texts

- no possessive: 65.6%
- 1: 3.9%
- 2: 2.5%
- 3: 28%

- It is often hard to tell whether any given occurrence of P.3SG is "possessive" or "discourse":
  - "true possessive": 24%
  - "discourse possessive": 60%
  - ????: 16%
“True possessives”

• The genitive case of the dependent always triggers the appearance of the possessive marker

• Well, almost always:

(9) sakar  odig  kôl’-iz=na  môtam
      sugar  one  remain-PST.3SG=else  I.GEN

‘as for the sugar, I have only one left’

• In all 3 examples we have, the head is in the topic

• That said, we discard “true possessives” and only look at the rest
Overall prediction quality

• Machine learning: an algorithm looks at the annotated data (“training dataset”) and tries to learn rules so that it could predict the target variable (whether discourse P.3SG appears on a noun) based on other parameters

• After that, we check how well the rules have been learned by applying them against a separate test sample

• We applied 3 algorithms: SVM, decision tree, random forest

• SVM and random forest give 84% correct answers

• Therefore, our set of parameters explains when the P.3SG appears fairly well, but there are probably some other factors or free variation
Feature importance

1. semantic class (0.23)
2. referential status (0.19)
3. referential distance (0.17)
4. uniqueness (0.13)
5. syntactic position (0.11)
...

Feature importance

• alienability, animacy, protagonism, distance to the first mention proved to be irrelevant

• topic persistence gave inconsistent results with different window lengths

• topicality seems to be important (66% non-possessives in topic, 77% in focus), but we have low inter-annotator agreement
Animacy

- Animacy was found to be important for the choice of ACC/0 marking of the DO in Pechora Komi (Serdobolskaya, Toldova 2012)
- In our model its predictive strength was very low
- No hierarchy: people > inanimate >> animate
- Other factors seem to interfere (toy animals/people; animals as protagonists in fairy tales)
Syntactic position

- The probability of being marked decreases along the hierarchy:

<table>
<thead>
<tr>
<th></th>
<th>no possessive</th>
<th>p.3sg</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>65%</td>
<td>35%</td>
</tr>
<tr>
<td>direct object</td>
<td>71%</td>
<td>29%</td>
</tr>
<tr>
<td>oblique</td>
<td>78%</td>
<td>22%</td>
</tr>
<tr>
<td>nominal dependent</td>
<td>90%</td>
<td>10%</td>
</tr>
</tbody>
</table>
Definiteness

• Whether the referential status of a noun is “definite” is important.

• Nouns that are not definite usually (93.5%) do not have possessive.

• Still, 60% of definite nouns also do not have possessives (50% if you count the “true possessives”).

• Is it the case that p.3sg is blocked by other factors (as e.g. in English the definite article is incompatible with demonstrative pronouns and is not used with proper nouns)?
Definiteness: demonstratives

- Demonstrative pronouns in Uralic languages usually block discourse possessive marking (folklore)
- This is not the case in Beserman:

<table>
<thead>
<tr>
<th></th>
<th>no possessive</th>
<th>p.3sg</th>
</tr>
</thead>
<tbody>
<tr>
<td>with so ‘that’ / ta ‘this’</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td>definite without so/ta</td>
<td>60%</td>
<td>40%</td>
</tr>
</tbody>
</table>
Definiteness: uniqueness

• Items that refer to unique objects usually do not have possessive marking

• (more or less) globally unique: sun, police, army

• unique in the given context (the village or the river when describing the life in that village or during experiments)

• 92% nouns in each of the two classes are used without possessives
Definiteness: proper names

- In our sample, all 68 proper names are used without possessives.
- In fact, using п.3sg with proper names is not prohibited:

(10) kakoj, mon lada-ze ej=na až’-ôl!

[RUS] I.NOM PN-p.3sg.acc neg.pst.1sg=yet see-ITER

[Have you seen Lada’s daughter?] — I haven’t even seen Lada herself yet!

- But п.3sg is used here for reasons other than marking definiteness.
Definiteness

• Even if we remove unique objects and proper names from the sample, still 42% of definite nouns will not be marked

• Therefore, $P_{3SG}$ cannot be described as the definiteness marker

• Out of these, subjects and direct objects are more likely to be marked
Topicality and referential distance

• The P.3SG marker attaches to topic nouns more frequently
• This trend is even more pronounced for subjects
• Referential distance: number of clauses between the current and the previous mention of the object in discourse
• With the increase if the distance, the object loses its activation
Topicality and referential distance

• It is evident that difference between, say, 15 and 16 clauses is less significant than between 2 and 3

• We divided this parameter into segments of exponentially growing lengths:

<table>
<thead>
<tr>
<th>segment</th>
<th>0</th>
<th>1</th>
<th>2-3</th>
<th>4-7</th>
<th>8-13</th>
<th>14-23</th>
<th>24-40</th>
<th>41-68</th>
<th>69-...</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref. dist.</td>
<td>0</td>
<td>1</td>
<td>2-3</td>
<td>4-7</td>
<td>8-13</td>
<td>14-23</td>
<td>24-40</td>
<td>41-68</td>
<td>69-...</td>
</tr>
</tbody>
</table>
Topicality and referential distance

True referential distance in clauses (smoothing by averaging over neighbors: at least 30 nouns for each point)
Topicality and referential distance

Referential distance in distance classes
Topicality and referential distance

• The probability of possessive marking increases as referential distance increases from 0 through 2 (24% → 40%), then stays at the same level, and then drops little by little

• This indicates that 3SG is used to reactivate the object that was mentioned at some point earlier, but lost its activation

• This can happen to both definite and indefinite nouns

• Nouns that need such reactivation are often topical

• The drop on greater distances can be explained by the fact that if the object was last mentioned too long ago, it loses its activation completely and has to be reintroduced
Contrast

• P.3SG is very frequent in contrastive focus:

(11) a gord mašina-ez-len odig pal es-ez paš’=uk?
and red car-P.3SG-GEN one side door-P.3SG open=EMPH

‘[They also opened both doors of the white car.] — And is one of the red car’s doors open?’

• Among nouns that denote one object out of a small set, only 31% lacked possessive marking

• In such cases, P.3SG can be described as having possessive semantics, this set being interpreted as the possessor
Conclusions - 1

• “discourse” P.3SG is more frequent than the “possessive” one

• There is no single parameter that explains its use

• Referential status, semantic class of the noun, syntactic position, referential distance, uniqueness, and topicality predict the appearance of P.3SG with 84% accuracy

• Animacy, alienability, topic persistence, and distance to first mention do not influence the probability of P.3SG marking

• Assignment of P.3SG to a RelNP seems to be independent from the choice of the host within that phrase
Conclusions - 2

• Referential status of the noun is important, but P.3SG has not yet evolved into a definiteness marker.

• P.3SG is used for reactivating the topic that was mentioned previously but lost activation (even for indefinite nouns), especially if it is in the subject position.

• P.3SG is used in contrastive contexts, where it can be interpreted as marking a possessive relation with respect to a small set.
References


- Tánczos, O. Towards a unified account of the suffix -ez/jez in Udmurt. Talk at SLE 2016


- Corpus of spoken Beserman: http://beserman.ru/corpus/
Thank you!