Possession in North Saami: Rich Morphology in Competition with an Analytic Construction

Laura A. Janda
Lene Antonsen
WALS Feature 57a: Possessive affixes

North Saami
An ongoing language change: NPx is being replaced by Refl

These two examples were found only a few pages apart in Kirsti Paltto’s novel Ája

NPx (possessive suffix with HIGH morphological complexity):

Son manai latnjasis.
3S.NOM go.IND.PRET.3S room.ILL.SG.PX.3S
‘He went into his room and lay down on the bed.’

Refl (analytic construction with reflexive genitive pronoun):

son... gavccui loki̱i iežas latnii...
3S.NOM climb.IND.PRET.3S upstairs.ILL.SG REFL.GEN.3S room.ILL.SG
‘she.. climbed upstairs into her room...’
Research questions

• Is language change always motivated by social value alone, or can there be inherent fitness values? Cf. Blythe & Croft (2012) “S-curves and the mechanisms of propagation in language change”
• Is morphological complexity disadvantaged in a context of language contact?
• What happens to the residue of a language change?
• What kinds of things are really inalienable?
Overview

1. The S-curve in our data
2. Inherent syntactic fitness
3. Inherent semantic fitness
4. Morphological complexity in the face of intense contact
5. A new vocative?
6. Conclusions
1. The S-curve in our Data

- **Literary texts**: 0.53M words, three age groups, two geographic regions
  - 2272 examples, full analysis by hand
  - 1530 examples, full analysis by hand
- **Newspapers**: 10M words from three newspapers (1997-2011)
  - 29964 examples of words with frequency ≥5, partial automatic analysis, a lot of cleaning by hand

- **Total**: 33633 examples
Tagging

• Construction (PossCon): NPx, Refl
• Reference: Anaphoric, Endophoric, Exophoric, Generic
• Possessum (the one that is possessed; PM):
  – Case: Nom, Acc, Gen, Ill, Com, Loc, Ess
  – Number: Sg, Pl
  – Semantic class: Kin, Human, Body...
• Possessor (the one that possesses; PR):
  – Case: Nom/Verb, Acc, Gen, Ill, Loc
  – Person and number: 1Sg, 2Sg, 3Sg, 1Du, 2Du, 3Du, 1Pl, 2Pl, 3Pl
  – Semantic class: Human, Animal, Nature...
• Source, Generation (Old vs. Mid vs. Young), Geography (East vs. West)

See examples illustrating Reference, Possessum Case, and Possessor Case on the handout

Only Anaphoric and Endophoric are used in statistical analysis
Distribution of possessive constructions

Old
Mid
Young
NT
News

Refl
NPx
The S-curve: longitudinal data from literary texts, showing only anaphoric and endophoric use.
No evidence that high frequency helps to retain NPx
News data: Pearson's correlation = -0.14, p = 0.0001, 95% confidence interval: -0.2 - 0.07
<table>
<thead>
<tr>
<th>Lemma</th>
<th>Meaning</th>
<th>total # exx</th>
<th>proportion NPx</th>
</tr>
</thead>
<tbody>
<tr>
<td>preassadieðáhus</td>
<td>‘press release’</td>
<td>293</td>
<td>97%</td>
</tr>
<tr>
<td>reive</td>
<td>‘letter’</td>
<td>133</td>
<td>92%</td>
</tr>
<tr>
<td>virgi</td>
<td>‘job’</td>
<td>140</td>
<td>89%</td>
</tr>
<tr>
<td>áhčči</td>
<td>‘father’</td>
<td>306</td>
<td>83%</td>
</tr>
<tr>
<td>sárdni</td>
<td>‘sermon, speech’</td>
<td>178</td>
<td>81%</td>
</tr>
<tr>
<td>eadni</td>
<td>‘mother’</td>
<td>276</td>
<td>78%</td>
</tr>
<tr>
<td>eallu</td>
<td>‘herd’</td>
<td>104</td>
<td>75%</td>
</tr>
<tr>
<td>viellja</td>
<td>‘brother’</td>
<td>115</td>
<td>71%</td>
</tr>
<tr>
<td>nieida</td>
<td>‘daughter’</td>
<td>139</td>
<td>65%</td>
</tr>
<tr>
<td>bánári</td>
<td>‘son’</td>
<td>187</td>
<td>64%</td>
</tr>
<tr>
<td>mánna</td>
<td>‘child’</td>
<td>645</td>
<td>62%</td>
</tr>
<tr>
<td>boazu</td>
<td>‘reindeer’</td>
<td>183</td>
<td>58%</td>
</tr>
<tr>
<td>eamit</td>
<td>‘wife’</td>
<td>121</td>
<td>55%</td>
</tr>
<tr>
<td>girji</td>
<td>‘book’</td>
<td>158</td>
<td>53%</td>
</tr>
<tr>
<td>namma</td>
<td>‘name’</td>
<td>220</td>
<td>50%</td>
</tr>
</tbody>
</table>
CART-analysis

“Classification and regression trees and Random forests”:

– Optimal sorting of data
– Results similar to regression, but appropriate for non-parametric data
– Bootstrapping and measurement of variable importance
CART-analysis for Literary texts + NT (Anaphoric + Endophoric data only)

PossCon ~ PMCase + PMClass + PRCase + Author
Variable importance for Literary texts + NT

PossCon ~ Generation + PMClass + PRCase + PMCase + Geography
CART analysis for News

PossCon ~ PMCase + PMClass + PRPERSNum
Measurement of variable importance for News

PMClass

PMCase

PRPersNum
What this analysis tells us

• We have an ongoing language change following an S-curve
• The most important factors are:
  – Case of possessum and possessor
    >> syntactic fitness
  – Semantic class of possessum
    >> semantic fitness
2. Inherent syntactic fitness

- Case of possessum
- Case of possessor
  - For case-marking of possessum and possessor, we see that Refl predominates in non-prototypical uses
- Replaceability
  - NPx is always replaceable, Refl is not

Refl has greater inherent syntactic fitness
Possessum Case (Literary texts + NT data)

<table>
<thead>
<tr>
<th>Case</th>
<th>Acc</th>
<th>Gen</th>
<th>Loc</th>
<th>Ill</th>
<th>Com</th>
<th>Nom</th>
<th>Ess</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1099</td>
<td>434</td>
<td>397</td>
<td>145</td>
<td>303</td>
<td>78</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>334</td>
<td>152</td>
<td>109</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>90%</td>
<td>80%</td>
<td>70%</td>
<td>60%</td>
<td>50%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Legend:
- Orange: Refl
- Green: NPx
Possessum Case

- Prototypical case for possessum is Accusative
  - ~40% of all uses of both NPx and Refl have Accusative
- Both NPx and Refl are robustly attested for Genitive, Locative, Illative, and Comitative
- Nominative and Essive are least prototypical for possessum
  - Here Refl has a strong advantage

- Refl is robust in all uses, and especially strong in non-prototypical uses
## Possessor Case (Literary texts + NT data)

<table>
<thead>
<tr>
<th>PR Case</th>
<th>NPx</th>
<th>%</th>
<th>Refl</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td>1810</td>
<td>79 %</td>
<td>484</td>
<td>21 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Verb</td>
<td>656</td>
<td>73 %</td>
<td>243</td>
<td>27 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Accusative</td>
<td>44</td>
<td>62 %</td>
<td>27</td>
<td>38 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Illative</td>
<td>10</td>
<td>43 %</td>
<td>13</td>
<td>57 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Locative</td>
<td>21</td>
<td>34 %</td>
<td>40</td>
<td>66 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Genitive</td>
<td>18</td>
<td>19 %</td>
<td>76</td>
<td>81 %</td>
<td>100 %</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>2559</strong></td>
<td><strong>74 %</strong></td>
<td><strong>883</strong></td>
<td><strong>26 %</strong></td>
<td><strong>100 %</strong></td>
</tr>
</tbody>
</table>
Possessor Case

- Prototypical case for possessor is Nominative/subject-agreement on the finite verb (“Nom/Verb”)
  - 96% of all NPx and 82% of all Refl uses are **Nom/Verb**
- Refl has strong advantage in non-prototypical uses, particularly **Locative** and **Genitive**
- **Refl is robust in all uses, and especially strong in non-prototypical uses**
Replaceability

- **NPx can always be replaced** by either Refl or null marking of possession or a Genitive pronoun
  - NPx can be replaced by Refl in exx. (1-12)
  - NPx can also be replaced by Genitive pronoun or Ø, as in:
    - *Máhttájeaddji ii leat stuorit go oahpaheaddjis* [Original with NPx]
      “The disciple is not greater than his teacher ”
    - *Máhttájeaddji ii leat stuorit go su oahpaheaddji* [Genitive pronoun]
    - *Máhttájeaddji ii leat stuorit go oahpaheaddji* [Ø]
- But the converse is not true for **Refl**, which **cannot always be replaced**:
  - *Muhto dien ádjagis lea iežas suollemasvuohta.* (KP2: 107)
    “But that spring has its secret.”
  - *Muhto dien ádjagis lea suollemasvuohtas.*
Summary for Syntactic Fitness

• Refl is well-represented across full syntactic spectrum
• Refl is predominant in peripheral uses
• NPx is always replaceable
• Refl is sometimes irreplaceable
• Recall also that Refl predominates in Generic Reference (a type of impersonal construction)
3. Inherent semantic fitness

Distribution across semantic classes for possessum
- **NPx** focuses on “core possession” of Kin, Body, and Property
  - This focus grows tighter over time
- **Refl** shows more even distribution across semantic classes
Distribution of NPx vs. Refl across semantic classes
4. Morphological complexity in the face of intense contact

- NPx is much more morphologically complex than Refl
- North Saami is under intense pressure from both Germanic (Norwegian and Swedish) and Finnish
- Many North Saami speakers have reclaimed the language as adults
- Language contact and 2nd language learners can lead to morphological simplification (Trudgill 2002, McWhorter 2007, Bentz & Winter 2013)
- In this situation, there may be an advantage for the morphologically simpler form: Refl
Reflexive pronoun is built entirely of otherwise existing morphology.

**Reflexive Pronoun + Substantive**

| Sg1 | iežan  
| Sg2 | iežat  
| Sg3 | iežas  
| Du1 | iežame 
| Du2 | iežade 
| Du3 | iežaska
| Pl1 | iežamet 
| Pl2 | iežadet 
| Pl3 | iežaset 

A substantive has 10 forms:

<table>
<thead>
<tr>
<th>guoibmi “partner”</th>
</tr>
</thead>
<tbody>
<tr>
<td>NomSg</td>
</tr>
<tr>
<td>GenSg=AccSg</td>
</tr>
<tr>
<td>IIIsg</td>
</tr>
<tr>
<td>LocSg</td>
</tr>
<tr>
<td>ComSg=LocPl</td>
</tr>
<tr>
<td>NomPl</td>
</tr>
<tr>
<td>GenPl=AccPl</td>
</tr>
<tr>
<td>IIIPl</td>
</tr>
<tr>
<td>ComPl</td>
</tr>
<tr>
<td>Ess</td>
</tr>
</tbody>
</table>
NPx requires a large quantity of unique morphology -- See Handout

- [N. Saami has 3 types of substantive stems: vowel stems, consonant stems, and contracted stems]
- NPx expands the paradigm of a substantive from 10 forms to 91
  - see 81 additional forms for guoibmi “partner” on handout with forms involving NPx-unique morphology boldfaced
- There are two full sets of 9 possessive suffix forms, one set attaches after vowels and one set attaches after consonants (compare Gen/AccSg with IllSg in handout)
- NPx conditions unique changes in case endings: IllSg -i/-ii > -s/-asa-; LocSg -s/-is > -st/-isttá/-istti-; IIIPl -ide/-iidda > -idas/-iiddás-
- With NPx certain case endings have additional morphophonemic variants depending on the type of stem: LocSg, Com Sg=LocPl, AccPl=GenPl, IIIPl
- NPx conditions additional morphophonemic alternation in the stem, e.g. i ~ á
- NPx requires insertion of the possessive suffix inside the Comitative Plural ending
5. A new vocative?

- Nominative + possessive suffix may be evolving into a new Vocative
- In preceding statistics we excluded Exophoric Reference
- Nearly all examples of Exophoric use are NPx, and half of those are what we call “Exophoric Vocative”:
  
  *Gula, mánážan.* (KP2: 6)

  mánážan

  child.DIM.NOM.SG.PX.1S

  “Listen, oh my (little) child.”

- Exophoric vocative:
  - Always: NPx, Nominative, 1Sg reference
  - Often: Diminutive, Proper Names

- Note that NomPl+possessive suffix is extremely rare, restricted only to 1Sg reference and address forms

- NPx with Nom 1Sg may be leaving the paradigm to become a derived form of address (cf. Daniël’ 2009, Andersen 2012 on recent evolution of a new vocative in Russian)
Conclusions

- Replacement NPx > Refl is ongoing language change shaped as S-curve
- This change is not easily explained by frequency or alienable vs. inalienable
- Possible factors:
  - Refl is more flexible syntactically
  - Refl is more flexible semantically
  - Refl is morphologically simpler
  - Paradigm of NPx is undermined by evolution of new Vocative
Take your linguistic data to the bank!

TROLLing
- is an international archive of linguistic data and statistical code
- is built on the Dataverse platform from Harvard University and complies with DataCite, the international standard for storing and citing research data
- is compliant with CLARIN (Common Language Resources and Technology Infrastructure in the EU), the EU research infrastructure for language-based resources
- assigns a permanent URL to each post
- uses metadata that ensures visibility and retrieval through international services
- is professionally managed by the University Library of Tromsø and an international steering committee.

Authors of scholarly works around the world are welcome to deposit their data in TROLLing, along with citations of their publications. Conversely, authors can reference their data by citing their TROLLing posts in their articles.

Visit us at
http://opendata.uit.no/
http://site.uit.no/trolling/

“In the age of Big Data, the creation of a general repository of datasets and statistical models for linguistic research is a welcome development. It will stimulate more research and new analyses.” -- Maria Polinsky, Director of the Polinsky Language Sciences Lab at Harvard University

“TROLLing will revolutionize research in linguistics and drive the discipline forward: making data publicly available significantly reduces the risk of bogus results, avoids duplication of efforts and facilitates large-scale analysis of meticulously annotated datasets.” -- Dagmar Divjak, Reader, Russian and Slavonic Studies, University of Sheffield

“TROLLing is crucial for the field of linguistics as it takes the next steps towards becoming more empirical. For the first time, it will be possible for researchers to deposit their primary linguistic data (the foundation for all research) in a central freely accessible on-line repository so that colleagues around the world have access to the same data. This invaluable resource will promote on-going academic exchange on an empirical basis.” -- Hans Boas, Professor, Department Germanic Studies and the Department of Linguistics, University of Texas at Austin

“TROLLing is exactly what our field needs - with the potential to become the most useful data resource in linguistics.” -- Marit Westergaard, Professor, Center for Advanced Study of Theoretical Linguistics, UiT The Arctic University of Norway

“I would like to recommend that scholars deposit their data at TROLLing. I strongly believe that sharing of data and methods for analysis can play a key role in the growth of cognitive linguistics. It will be beneficial for the community of linguists to have a single searchable repository rather than having data scattered about in many places.” -- Laura Janda, Professor, Center for Advanced Study of Theoretical Linguistics, UiT The Arctic University of Norway