Kildin Saami documentary sociolinguistics
Investigations on the variation and change in function words

Niko Partanen, Michael Rießler, Evgenia Zhivotova

Diversity Linguistics: Retrospect and Prospect
MPI EVA Leipzig, May 1–3, 2015

May 1, 2015
Co-authors and ongoing projects

Grammatical description (Kildin Saami varieties)
Michael Rießler | Uni Freiburg

Methodology in documentary sociolinguistics (Komi and contacting Barents languages)
Niko Partanen | Uni Freiburg

Function words in language contact (different indigenous languages of Russia)
Evgenia Zhivotova | MPI EVA Leipzig
Kildin Saami (Glottolog: kild1236)

Language sociology

- 100 “active” / 700 “passive” speakers (Scheller, 2011)
- revitalization attempts, incl (written) language planning

Changes induced by language shift to Russian

- all levels of grammar (for variation and change in function words specifically, see Kert, 1975, Karvovskaya, 2011, Zhivotova, 2014a,b)
- “revitalization-induced change” (due to purism) (Rießler et al., 2013)
Kildin Saami documentary sociolinguistics

Introduction

Language documentation

Alexandra A. Antonova (1932–2014)

Fieldwork (Rießler et al., 2005+).
Kildin Saami documentary sociolinguistics

Introduction
Language documentation

Alexandra A. Antonova (1932–2014)

Fieldwork

Old recordings

Literature

E-mail, social media, ...

Sie zogen viele Zeit dort herum. Dann fanden sie ein leeres, unbewohnetes Haus. Nun, sie meinten, sie wollten hier vorläufig bleiben.
Language documentation meets language technology

Our goal

- develop better interaction between language documentation and language technology
- building better corpora, which make empirically verifiable analyses possible
  - more data (>500,000 word tokens for Kildin)
  - diverse data (language modes, registers, genres, etc.)
  - contextualized data (“thick” metadata: speaker biographies, meta-documentation, etc.)

Our approach

- rather shallow annotation
  - orthographic transcription, no interlinearization
  - but better automated morphosyntactic glossing
Kildin Saami documentary sociolinguistics
Language documentation meets language technology
Language technology

ELAN → FST → ELAN (Blokland et al., 2015)

token lemma+PoS+Gloss ‘These are fish.’

тэгк тэдт+Det+Pl+Nom

лēв лиййе+V+3+Prs

куль кууллъ+N+Pl+Nom
кууллъ+N+Sg+Gen
кууллъ+N+Sg+Acc
Corpus data interface

ELAN → R (Partanen, 2015)
The connector “and”

Kildin Saami /ja/ inherited (← Saamic < Germanic)

- я /ja/
  - NP coordinator, VP coordinator, S coordinator

Kildin Saami /i/ borrowed (< Russian)

- э /i/
  - NP coordinator, VP coordinator, S coordinator

Ja menn sōnn tes’ oat, munn vāll’te i rūv’teguejm pierrez pīmm’e-denn. ‘And it was lying here; I took (it) and cut the paper with the scissors.’

([sjd20070421kudckullj-008] NEA, spoken, informal)
The connector “and”

Token distribution across four speakers

- и /i/ Russian (code-switching)
- э /i/ < Russian (borrowing)
- я /я/ ← Saamic (inherited)
Kildin Saami documentary sociolinguistics
Case study on connector “and”

Variation

Variation in two speakers/writers across text modes 1

AAA (f, 1932) and NEA (f, 1939)

- female L1 speakers
- teachers and language activists
- productive writers and consultants for linguists
- different dialects

Control group: PPJ (m, 1936), PPZ (f, 1913)

- two non-writers recorded by earlier projects
Variation in two speakers/writers across text modes 2

Our (sub-) corpus data: Tokens across spoken and written mode

![Graph showing variation in tokens across spoken and written modes for different speakers and dialects.]
Kildin Saami documentary sociolinguistics
Case study on connector “and”
Variation

Variation in two speakers/writers across text modes 3

<table>
<thead>
<tr>
<th>Token</th>
<th>AAA</th>
<th>NEA</th>
<th>PPJ</th>
<th>PPZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>э я</td>
<td>200</td>
<td>100</td>
<td>200</td>
<td>100</td>
</tr>
</tbody>
</table>

Dialect:
- Arsjogk
- Kiillt
- Lujaavv'r
Summary

Exemplary case study

- inherited /ja/ and borrowed /i/ compete over the same functions
- inherited /ja/ is preferred in written variants (in line with Zhivotova, 2014a,b)

More detailed investigations, including statistics, will be carried out on a larger and more diverse corpus.

- Applying language technology in language documentation helps us to create more systematically annotated corpora, rather than eclectic data collections.
- Ultimately, the multimodal corpora we create will be useful for quantitative investigations into synchronic variation and change in Kildin Saami.
References


References II


References III


References IV