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

Kola Peninsula, NW-Russia

East-Saamic < Saamic < Uralic

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)

Alexandra A. Antonova (1932–2014)

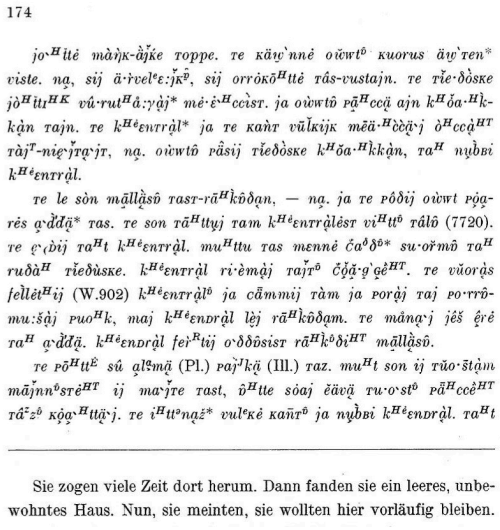
Fieldwork (Rießler et al., 2005+)

Alexandra A. Antonova (1932–2014)

Fieldwork

Literature

Old recordings



174

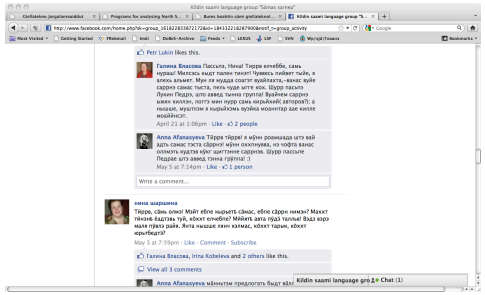
jo·^{H}ttė màǹk-ə̀ǰ̌ke toppe. te kāɲ̀·nnė oĭ̈wt^{b} kuorus ä̈ɣ̀ten viste. nạ, sij ä·rvel^{e}ɛ:ǰ̌k^{p}, sij orrȯkō^{H}ttė tâs-vustajn. te rĭ̈e·ɔ̀ɔ̀ske jȯ^{H}tt^{HK} vú·rut^{H}å:ɣạ̀j* mė·è·^{H}ccìsr. ja oȯ̀wtə̄ pą̄^{H}ccą̈ ajn k^{H}ɔ̌a·^{H}]k-kạ̀n tạjn. te k^{He}ɛntrạ̀l* ja te kaǹt vūĺ̌kijk mėä·^{H}ccą̈·j ȯ^{H}ccą̀^{HT} tàj^{T}-nię·ǰ̌rɑ·jr, nạ. oȯ̀wtə̄ pâsij rĭ̈eɔ̀ɔ̀ske k^{H}ɔ̌a·^{H}]kkạ̀n, ta^{H} nųɔ̀bbi k^{He}ɛntrạ̀l.*

te le sòn mą̄llą̀sə̄ rast-rā^{H}]kə̄ɔ̀ạn, — nạ. ja te pɔ̂ɔ̀ij oȯ̀wt pə̣ạ-rės ạ·ɗ̀ɗą̄ tas. te son tā^{H}ttuj ram k^{He}ɛntrạ̀lės^{T} vi^{H}tt^{b} tálə̄ (7720). te ɕ·(ɔ̀ij ta^{H}t k^{He}ɛntrạ̀l. mu^{H}ttu tas mɛnnė ča^{b}δ^{b}* su·oĭ̈mə̄ ta^{H} ruɔ̀à^{H} rĭ̈eɔ̀ùske. k^{He}ɛntrạ̀l ri·èmą̀j ta^{J}ǰ̌^{b} Ć^{b}ǫ̆ą̄·g·ǵ·ǫ̈ė^{HT}. te vūorą̀s feĺ̀ėt^{H}ij (W.902) k^{He}ɛntrạ̀l^{b} ja cämmij tàm ja pɔrą̀j taj po·rrə̂-mu:šą̀j puo^{H}k, maj k^{He}ɛnɒrạ̀l lę̀j rā^{H}]kə̂ɔ̀ạm. te mâɳą·j jéš ę̣rė ta^{H} ạ·ɗ̀ɗą̄. k^{He}ɛnɒrạ̀l fe^{R}ì^{R}tij o·ɔ̀ɔ̀ə̂sist rā^{H}]k^{b}ɔ̀i^{HT} mą̄llą̀sə̄.*

te pə̄^{H}tt^{E} sú ạl^{q}mą̄ (Pl.) paǰ^{J}kə̄ (Ill.) taz. mu^{H}t son ij tŭo·šṭą̀m mą̄ǰ̌nn^{b}stė^{HT} ij ma·ǰ̌re tast, ɔ̀^{H}tte sòaj ĕävą̄ ru·o·st^{b} pā^{H}ccė^{HT} tâ^{z}^{b} kə̣ǫ̆·^{H}ttą̈·j. te i^{H}tt^{e}nąž vul^{e}kė kañt^{b} ja nųɔ̀bbi k^{He}ɛnɒrạ̀l. ta^{H}t*

Sie zogen viele Zeit dort herum. Dann fanden sie ein leeres, unbewohntes Haus. Nun, sie meinten, sie wollten hier vorläufig bleiben.

E-mail, social media, ...



Language documentation meets language technology

Our goal

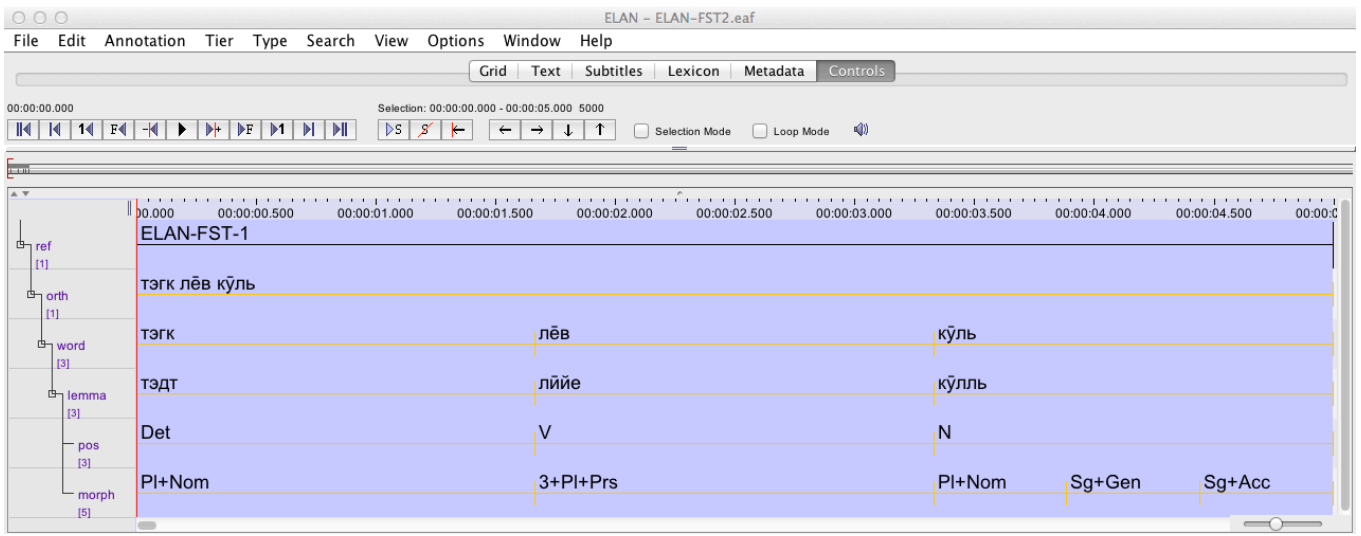
- develop better interaction between language documentation and language technology
- building better corpora, which make empirically verifiable analyses possible
  - more data (>500.000 word tokens for Kildin)
  - diverse data (language modes, registers, genres, etc.)
  - contextualized data ("thick" metadata: speaker biographies, meta-documentation, etc.)

Our approach

- rather shallow annotation
  - orthographic transcription, no interlinearization
  - but better automated morphosyntactic glossing

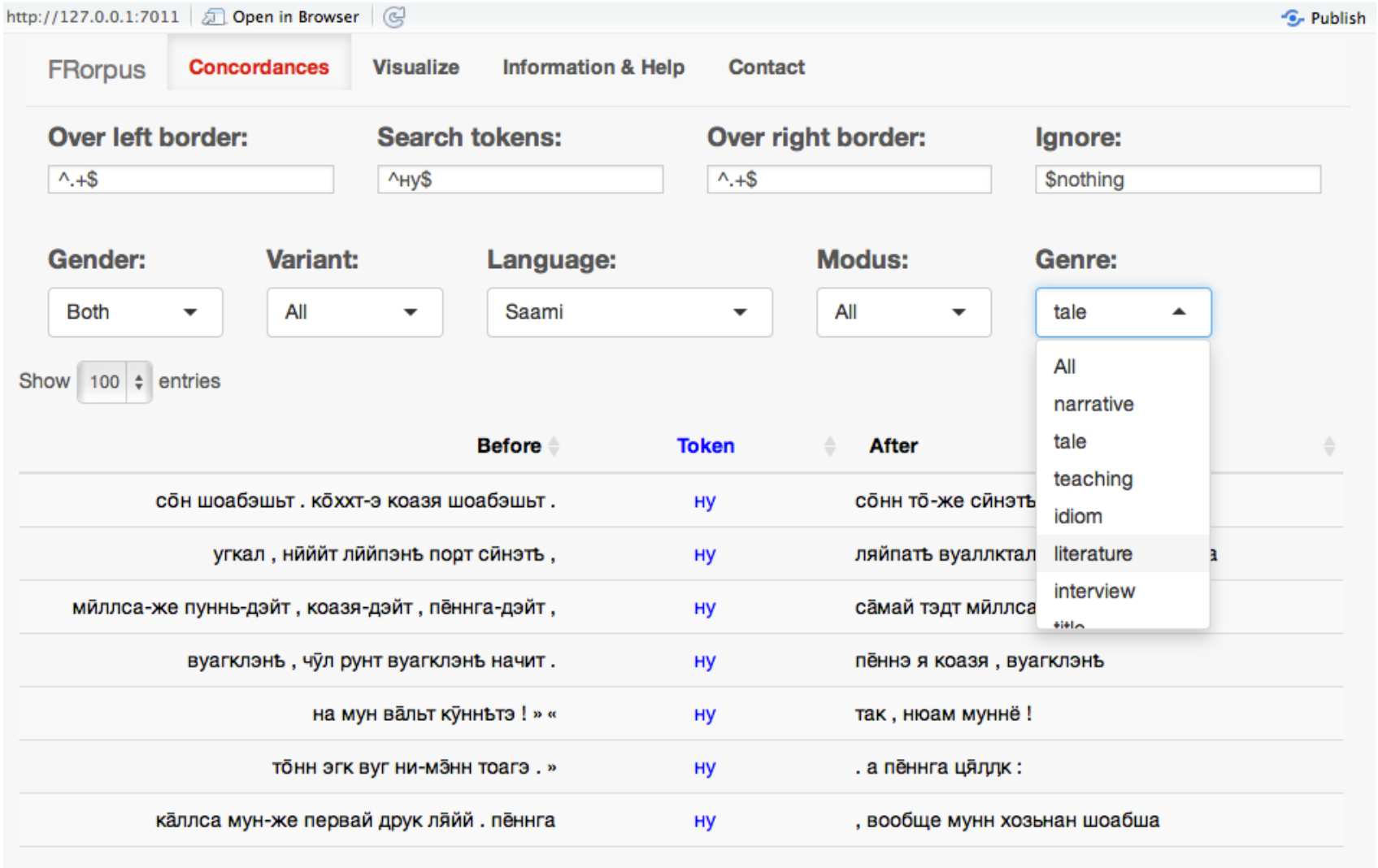
ELAN → FST → ELAN (Blokland et al., 2015)

token	lemma+PoS+Gloss	‘These are fish.’
тэгк	тэдт+Det+Pl+Nom	
лӣв	лӣйе+V+3+Prs	
кӯль	кӯлль+N+Pl+Nom	
	кӯлль+N+Sg+Gen	
	кӯлль+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)

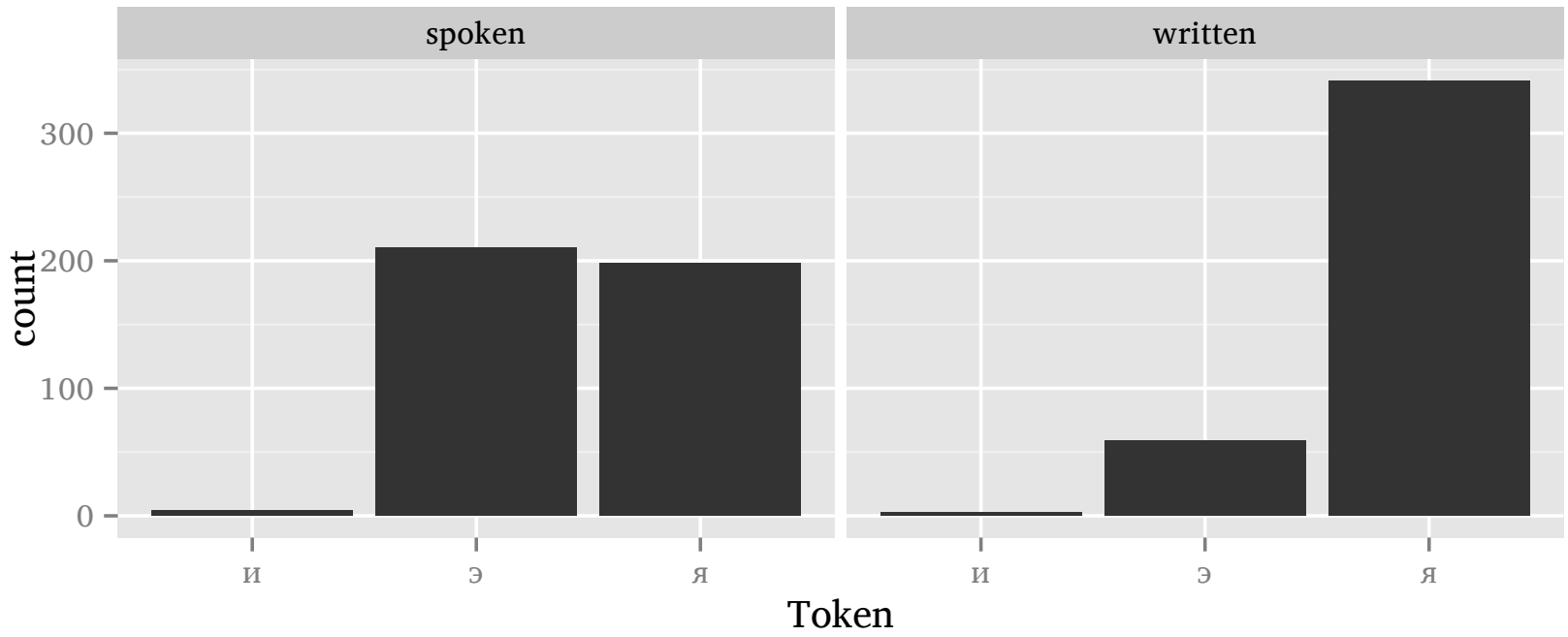
- э /i/
- NP coordinator, VP coordinator, S coordinator

Ja menn sōnn tes't oarr, munn vāll'te i rūvv'teguejm pierrse pɨmm'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)

Variation in two speakers/writers across text modes 1

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

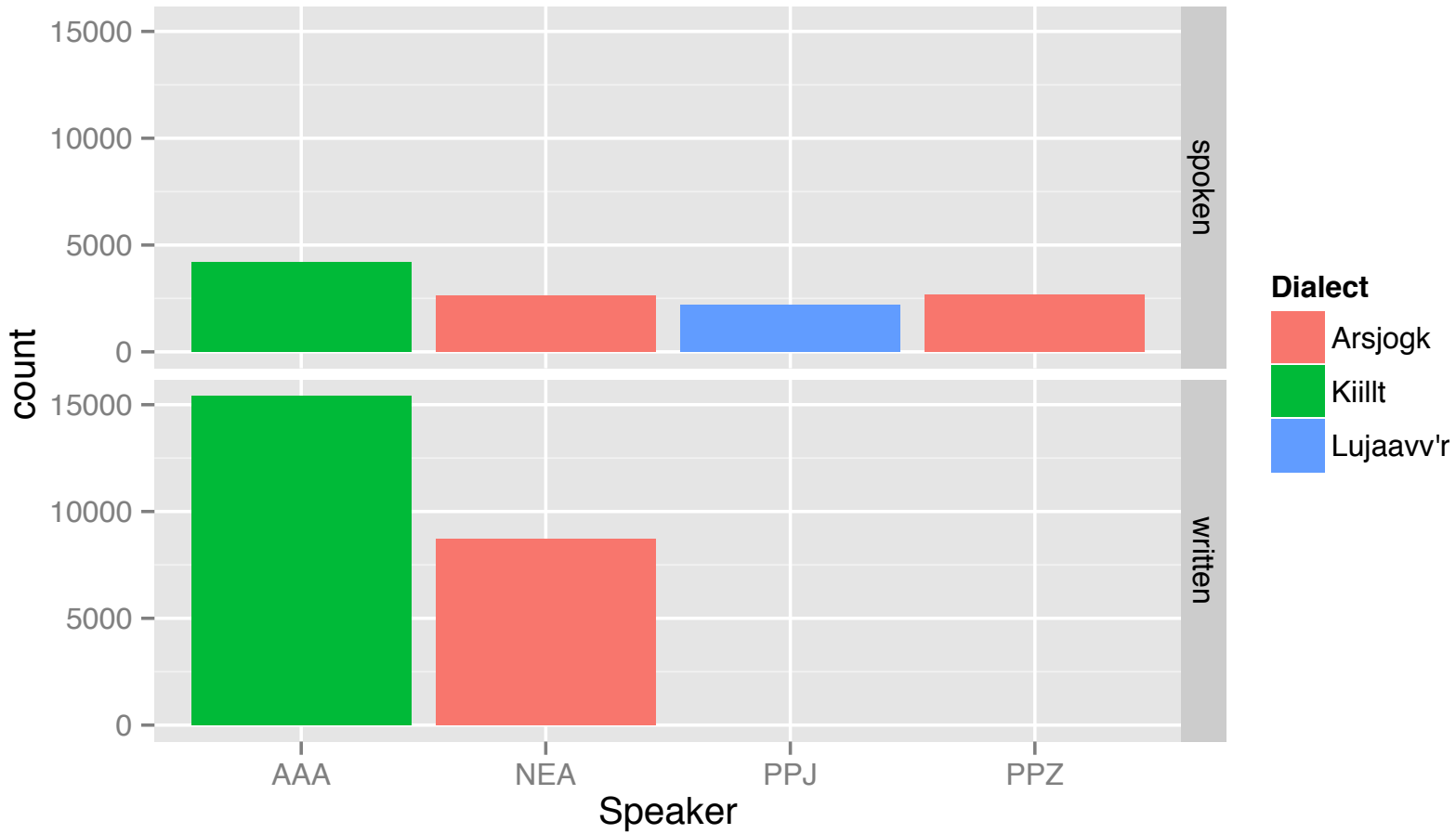
- female L1 speakers
- teachers and language activists
- productive writers and consultants for linguists
- different dialects

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

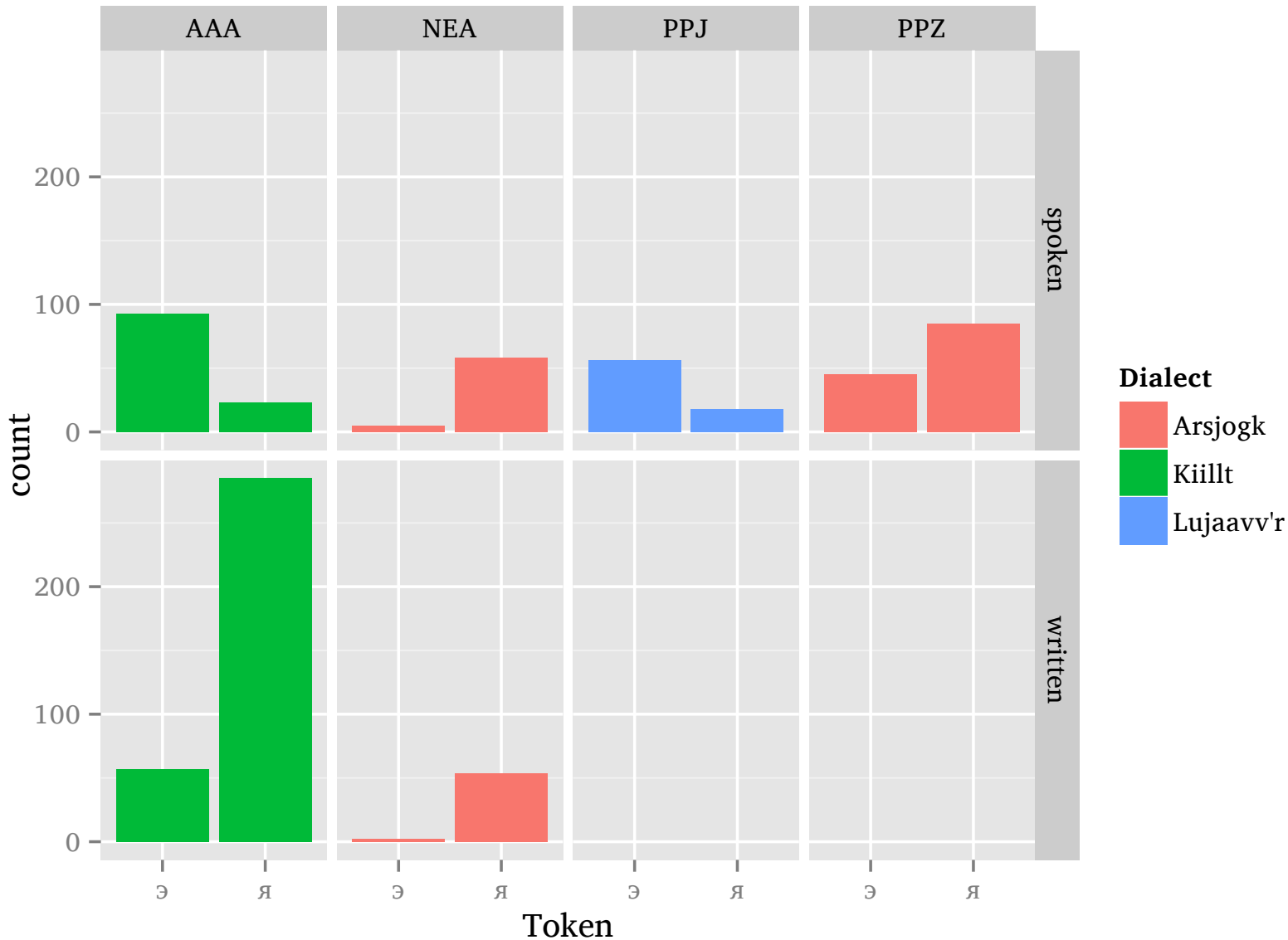
- two non-writers recorded by earlier projects

Variation in two speakers/writers across text modes 2

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



Variation in two speakers/writers across text modes 3



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 detailled investigations, including statistics, will be carried out on a larger and more divers 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 I

Blokland, Rogier et al. (2015). "Language documentation meets language technology". In: *Proceedings of the First International Workshop on Computational Linguistics for Uralic Languages*. Ed. by Tommi A. Pirinen et al. Tromsø. In press.

Haspelmath, Martin (2007). "Coordination". In: *Language Typology and Syntactic Description*. Ed. by Timothy Shopen. 2nd ed. Cambridge, pp. 1–51.

Karvovskaya, Elena (2011). "Focus particles in endangered languages". In: *Proceedings of Conference on Language Documentation and Linguistic Theory 3*. Ed. by Peter K. Austin et al. London, pp. 161–170.

Kert, G[eorgij] M[artynovič] (1975). "Saamskij jazyk". In: *Osnovy finno-ugorskogo jazykoznanija*. Vol. 2. Ed. by V. I. Lytkin et al. Moskva, pp. 203–247.

References II

Partanen, Niko (2015). *FRelan. Parse and analyze ELAN files*. R package version 0.1. URL: http://github/izvakomi/FRelan.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Rießler, Michael et al. (2005+). "Kola Sámi Documentation Project (KSDP)". In: *Endangered languages. DoBeS archive*. Nijmegen. URL: http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI363060%23.

Rießler, Michael et al. (2013). "Purism in language documentation and description. " In: *Keeping languages alive. Documentation, Pedagogy and Revitalization*. Ed. by Mari Jones et al. Cambridge, pp. 83–97.

References III

Scheller, Elisabeth (2011). "The Sámi language situation in Russia. Finno-Ugric minorities". In: *Ethnic and linguistic context of identity. Finno-Ugric minorities*. Ed. by Riho Grünthal et al. Uralica Helsingiensia 5. Helsinki, pp. 79–96.

Wickham, Hadley (2009). *ggplot2: elegant graphics for data analysis*. URL: http://had.co.nz/ggplot2/book.

Xie, Yihui (2013). *knitr: A general-purpose package for dynamic report generation in R*. URL: http://yihui.name/knitr/.

Zhivotova, Evgenia (2014a). "Function words in a situation of language contact. Kildin Sámi and Russian". poster presented at The Young Linguists' International Conference, Stockholm, 28 June 2014.

References IV

Zhivotova, Evgenia (2014b). “Function words in a situation of language contact. Kildin Sámi and Russian”. Conference paper presented at The 2nd Silesian Meeting of Young Scholars, Szczyrk, Poland, 6–7 November 2014.