# The Votic Use-Case of Morfologilabbet 2.0

Kristian Kankainen

February 2018

## 1 Main task

- morphology expansion
  - only nouns vs not only nouns ...
- for creating a prescriptive morphological dictionary
- for creating a written standard for Votic
- lexicon expansion will be needed

## 2 Resources

- 1000 entry dictionary (vot-est-rus-fin-eng)
  - with sort-of morphology)
  - I'm uncertain whether to swell it
- 30'000 entry phonetic dialectal dictionary
  - with much information
  - and example sentences with translations into Estonian and Russian
  - I re(tro)-digitalize it, I will bring it to Karp
  - but that's a different project (next project for mfl)
- corpus of my supervisor's text book (spelling variation)
- grammar description (in Russian)

## 3 Vision

- only use pextract – maximal structural units
  - to map other classifications on top of
  - be part of language documentation / documentary linguistics
- to couple my Votic morphological dictionary to GF (pextract-lmf2gf)

# 4   Setting

- no users of the language
- language submersion program "Ämmässe vunukassaa"
  - "from grandmother to grandchild"
  - more than 10, rising trend
  - one week once a year
  - Votic in the computer

# 5   Done work

- pextract-xml(DFDL parser/serializer) as an unstable API for the pfile
- pextract-lmfas a stable API for the pfile
- pextract-lmf2gf
  - has nice lemma agnostic features
  - work in progress, no flexible type system yet

# 6   Work flow

1. add full example paradigm for the first word
2. take next word from the candidate list
   (a) if matches existing paradigm
       i. then goto 2
       ii. else goto 1
3. change variable instantiation for naming existing paradigms
4. inspect variable instantiations (quality vs quantity)
5. add higher level class information for features that can collapse/group existing paradigms
6. keep my Votic GF synchronized