

# Designing a Lexicographic Database for an Online Dictionary

Timofey Arkhangelskiy (National Research University  
Higher School of Economics)

R. Idrisov (Institute of Linguistics, RAS)

N. Serdobolskaya (Russian State University for the  
Humanities)

M. Usacheva (Moscow State University)



# Beserman Lexicographic Project

- Long term lexicographic project aiming at description of the Beserman dialect of Udmurt
- A lexicographic database (~4000 entries) and a corpus (~75,000 tokens) were collected during fieldwork

## **Target audience:**

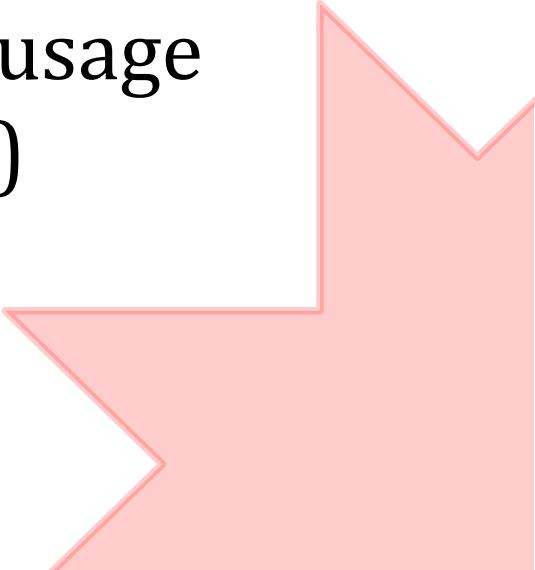
- researchers (→ need to have a publicly accessible online version)
- native speakers (→ possibility of exporting the data to a printable document)

# The Beserman dialect

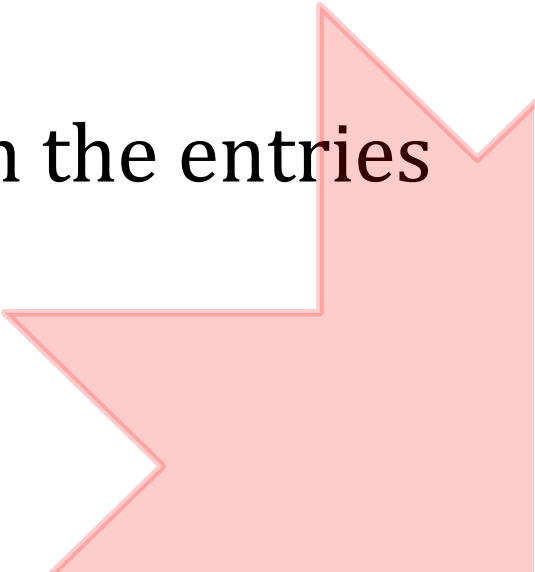
- a dialect of Udmurt (Uralic > Permian)
- spoken by ~2000 in NW Udmurtia and bordering regions
- no established written form
- quite poorly documented



# Tools used

- TshwaneLex for the dictionary database
  - SIL Fieldworks for the corpus
  - both have an option of exporting the data to XML
  - an additional python/flask module for connecting the two and providing an online interface (one can search a word in the dictionary and its usage examples in the corpus simultaneously)
- 

# Entry templates

- entries conform to a flexible hierarchical template
  - the fields may be marked as publicly accessible or restricted (for the lexicographer's use only)
  - the template can be altered (and has been altered many times) in the course of the project without adverse consequences for the database
  - there are several types of links between the entries
- 



# Thank you!

