

# Mansi documentation: new approaches and new data

Daria Zhornik

Lomonosov Moscow State University  
Institute of Linguistics, Russian Academy of Sciences

*daria.zhornik@yandex.ru*

Linguistic Forum: Indigenous languages of Russia and  
beyond. Moscow, 04.04.2019

Introduction

○○○

Upper Lozva Mansi

○○○○○○○○○○

Existing materials

○○○○

UL Mansi documentation

○○○○○○○○○○

Conclusion

○○

# Overview

## Introduction

## Upper Lozva Mansi

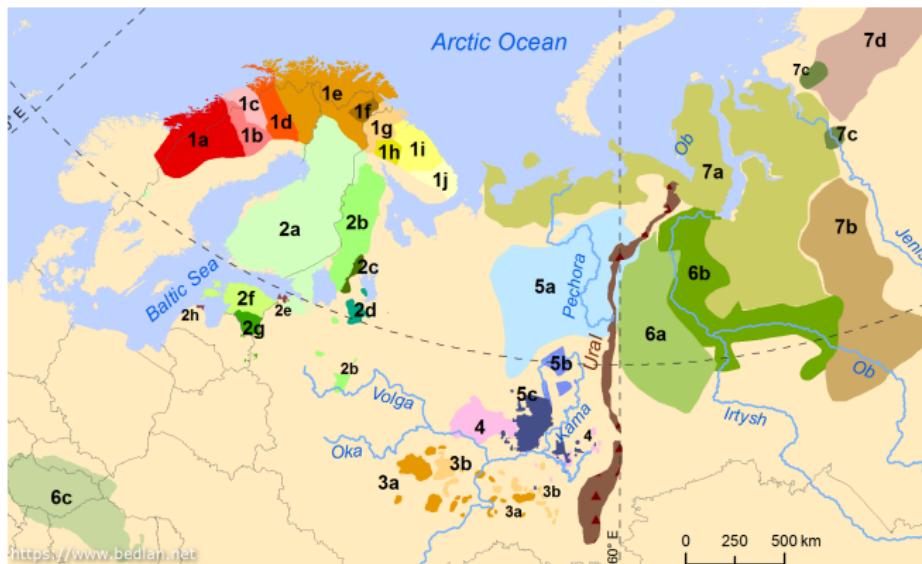
## Existing materials

## UL Mansi documentation

## Conclusion

# The Mansi language

- ▶ **Mansi** < Ob-Ugric < Finno-Ugric < Uralic
- ▶ An indigenous language of Russia (Western Siberia)



# Mansi dialects

- ▶ The size of the ethnic group is approx. 12,500, while the number of Mansi **speakers** is only 938 (2010 census)
- ▶ 4 Mansi dialect branches:
  - ▶ Western (†, beginning of the 20th century)
  - ▶ Southern (†, 1950s)
  - ▶ Eastern (†, end of the 20th century)
  - ▶ **Northern**

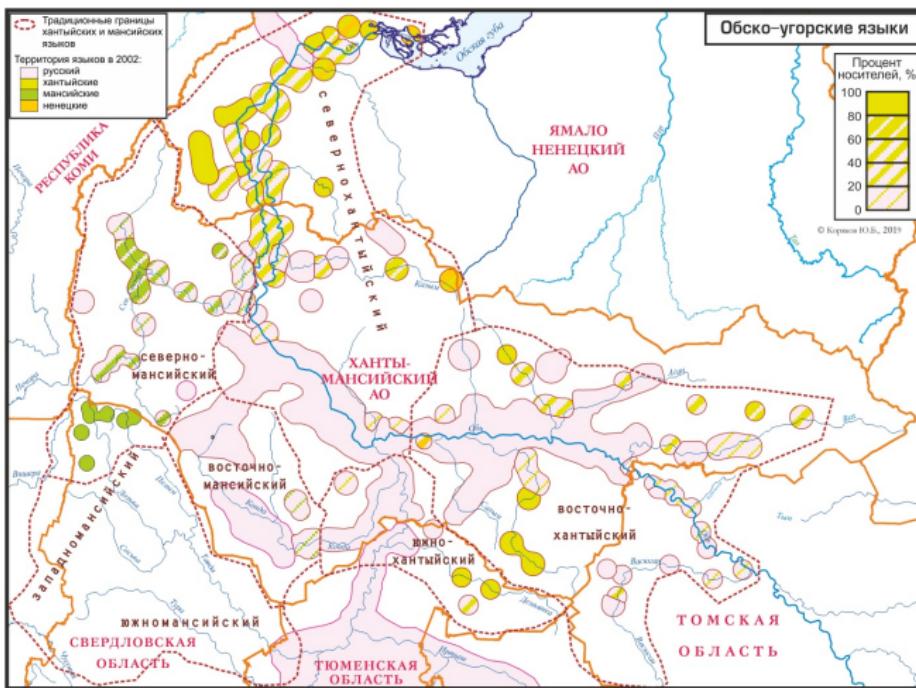
# Northern Mansi

- ▶ The Northern Mansi branch comprises Sosva (basis for standard Mansi), Sygva, Upper Lozva dialects.
- ▶ Most speakers: disperse distribution in the Khanty-Mansi Autonomous Okrug (Sosva and Sygva Mansi)
- ▶ The majority of Mansi people live in Russian villages or towns:
  - ▶ forced shift to Russian;
  - ▶ few opportunities for using Mansi;
  - ▶ all education is in Russian;
  - ▶ Mansi is even rarely taught as a separate subject.
- ▶ It seems that much less people really use the Mansi language than those mentioned in the census data.

# Upper Lozva Mansi

- ▶ **Upper Lozva (UL) dialect:** the only well-retained Mansi variety.
- ▶ Speakers live in the Sverdlovsk oblast: geographical and administrative separation from the main mass of Mansi people, a “language island”.
- ▶ Exact number of speakers unknown: 61 according to 2010 census, 108 according to some official data from 2017.
- ▶ The real number might be somewhere in between; about a half of these people constantly live in UL Mansi settlements.

## Mansi dialects (map by Yuri Koryakov)



# Living conditions

- ▶ UL Mansi: a closed community deep in the forest (approx. 150 km from the nearest small town Ivdel)
- ▶ UL Mansi settlements:
  - ▶ Ušma: approx. 20 residents;
  - ▶ Treskol'je (10km away): 7 permanent residents + frequently visiting relatives;
  - ▶ several more distant settlements (40-80km away) containing up to 5 people each;
  - ▶ solitary forest dwellers.
- ▶ UL Mansi have no access to electricity or most communications (internet, mobile connection etc.)

# Living conditions



# Languages of interaction

- ▶ All speakers are somehow related, most UL Mansi people originate from two families: Anjamovs and Baxtijarovs.
- ▶ They use exclusively Mansi in communication with each other, and switch to Russian only to interact with outsiders.
- ▶ Members of the community differ in terms of mobility and variety of interactions:
  - ▶ some speakers travel to “civilization” frequently and have business with Russians;
  - ▶ others always stay in their home villages and may leave only on exceptional occasions (to get to the hospital, withdraw money from the bank etc.)

# Education

- ▶ All UL children are fluent in Mansi and only start learning Russian at school;
- ▶ However, as they are forced to attend a boarding school (120km away from home), their use of Russian increases dramatically as soon as they begin their studies.
  - ▶ education is fully in Russian;
  - ▶ according to UL speakers, there are no competent Mansi teachers at the boarding school;
  - ▶ UL children have to communicate with everyone else in Russian
- ▶ The situation is tolerable, as UL children use Mansi to communicate with each other and during holidays they are taken back home.

# Standard Mansi competence

- ▶ UL Mansi is in some respect different from the Sosva dialect and even more so from Standard Mansi.
- ▶ Understanding written Mansi materials is difficult for UL Mansi speakers.
- ▶ A major issue is Mansi orthography, which does not accurately reflect some features of even Sosva dialect, not to mention UL.
- ▶ Another problem: newspaper articles and similar texts are heavily influenced by Russian syntax and contain numerous Russian words → some of these texts seem unnatural to UL speakers.

## UL vs. Sosva (= Standard) Mansi

- ▶ Loss of instrumental case in personal pronouns: forms like *anum-täl* ‘I.OBL-INSTR’ found e.g. in [Rombandeeva 1973] are deemed ungrammatical by UL speakers.
- ▶ Merger of ablative and instrumental cases in some instances:

(1) am xāp-əl/ xāp-nəl kwal-s-um  
I boat-INSTR boat-ABL get.up-PST-1SG  
'I got out of the boat'.

# UL vs. Sosva (= Standard) Mansi

- ▶ Merger of phonemes <γ> and <j>: *sajrap* [sajrap]  
‘axe’, *pəγ* [pəj] ‘boy’
- ▶ Archaic allomorphs of lative and locative cases: *-na/-ne* vs.  
Sosva *-n*, *-ta* vs. Sosva *-t*.
- ▶ Lexical differences: *xajtnut* ‘wolf’ vs. Sosva *sāliuj*.
- ▶ etc.

# Existing Mansi documentation

- ▶ The Mansi language has been documented by various scholars since the middle of the 19th century:
  - ▶ Antal Reguly, 1843—1844
  - ▶ August Ahlqvist, 1858—1859, 1877
  - ▶ Bernát Munkácsi, 1888—1889
  - ▶ Artturi Kannisto, 1901—1906
  - ▶ Nikolai Chernetsov, 1930s
- ▶ Kannisto's texts were published in 1950s by M. Liimola as “Wogulische Voksdichtung”, 5 volumes of Mansi texts.
- ▶ Munkácsi's work was later turned into [Munkácsi, Kálmán 1986], an extensive dialectal dictionary.
- ▶ Chernetsov's materials contain numerous Mansi texts, which have not yet been processed in any way.

## Existing Mansi documentation

4-

55 wörn da minassat. akw mat  
sorop ö! ha iomöhi. da lakwsa-  
lidnsö. akw ert fau nuplano  
da ängw atps. lawi  
-annum ul älelen. am nanan  
jursqate sun  
-ia lawe öt, nay jursqate  
fanansla de jursqate ös. sun  
da minassat. fa sleet  
uit konazsun da iale öt.  
fa lönje öt, da xöpaltööt da  
iöyxerteöt. iajäjte akw ert  
lawi  
-e lawi, mansr potzla a'  
di manri'ö fi atum utane  
fotsum utane. potzla' wä-  
re öt!  
akw ert iajäjte äymsl  
li jemts. lawi:  
-öt. lawi, sat pis ker  
ärwip ampar +ot öli. sun  
jallen, sat pis ker ärwip  
ampar pattan sun fallen

# Existing Mansi descriptions

- ▶ The only existing Mansi grammars are:
  - ▶ A grammar of Eastern Mansi, based on A. Kannisto's texts: [Kulonen 2007];
  - ▶ A grammar of Northern Mansi (Sygva and Sosva dialects), based on author's field data and her own speaker's competence: [Rombandeeva 1973];
  - ▶ A grammar of Sosva Mansi, based on A. Kannisto's texts, thus reflecting the language of 1904-1905: [Murphy 1968];
  - ▶ A brief sketch, which summarizes previously available descriptions of Northern Mansi: [Riese 2001].

# Materials on Upper Lozva Mansi

- ▶ UL data is absent from these descriptions, only Rombandeeva rarely mentions UL wordforms in her work.
- ▶ Some UL texts are to be found among A. Kannisto's and N. V. Chernetsov's materials, but to our knowledge they have not yet been analyzed.
- ▶ Gábor Székely has collected a few UL wordlists in 2005.
- ▶ T. P. Bakhtiyarova (a native UL speaker) has published a 2000 word Mansi-Russian dictionary of her dialect in Khanty-Mansiysk in 2016.
- ▶ To sum it up, some data from UL Mansi exist, however, there is neither an annotated text corpus nor a comprehensive grammar.

# Our project

- ▶ Supported by the RFBR grant №18-012-00833A “*Dynamics of phonetical and grammatical systems of Ob-Ugric languages*”
- ▶ **Team:** Daria Zhornik, Sophie Pokrovskaya (main investigators), Fyodor Sizov (software development), Vladimir Plungian (general supervision), Lev Kozlov (data processing), Tatiana Bakhtiyarova (language consultant).
- ▶ Fieldwork on the Upper Lozva dialect:
  - ▶ 3 field trips to Ušma and Treskol'je in 2017-2018;
  - ▶ elicitation tasks;
  - ▶ text recordings;
  - ▶ lexicon collection;
  - ▶ experiments;
  - ▶ language games.

# Our goals in UL Mansi description and documentation

- ▶ Step 1: documentation of the modern state of UL Mansi
- ▶ Step 2: synchronic description of UL Mansi based on the data gathered
- ▶ Step 3: investigation of the diachronic development of UL Mansi
- ▶ Step 4: determining the place of UL Mansi in the Mansi dialectal continuum
- ▶ Step 5: providing a grammar of UL Mansi
- ▶ Step 6: UL Mansi revitalization and language support.

## Step 1: documentation

- ▶ Gather data (texts, elicitation materials) from all available speakers, which range from school children to elderly speakers, to get a representative account of existing variation.
- ▶ Collect accurate data on the sociolinguistic among UL Mansi.
- ▶ **Today:** 23 speakers, 13-78 years old, 7 hours of texts, 40 hours of elicitation sessions
- ▶ Employ experimental methods and gamification (interaction in pairs and groups, also on less common topics and in new circumstances) alongside traditional methods.
- ▶ Create a searchable corpus of UL texts; pilot version available at <http://digital-mansi.com/corpus>.

## Step 2: synchronic description

- ▶ Process and analyze all materials gathered.
- ▶ Identify gaps in the data and fill them.
- ▶ Provide a comprehensive description of UL Mansi in its own terms accounting for extensive variation found in the dialect.

## Step 3: diachronic analysis

- ▶ Process, annotate and analyze old UL Mansi texts:
  - ▶ A. Kannisto's texts from 1904: digitalized and partly annotated by members of E. Skribnik's ObBabel project in Munich, <http://wwwbabel.gwi.uni-muenchen.de/>
  - ▶ V. N. Chernetsov's texts from 1938: our joint attempt with TSU to digitize and annotate texts from his archive, RSF grant 19-18-00329 "*Interpreting linguistic and cultural history of Mansi: ethnographic, folklore and linguistic data from Valery Chernetsov's archive*", 2019-2021.
- ▶ Analysis of these texts in comparison to those gathered in the field allow us to trace the evolution of UL Mansi as far as 115 years back.

## Step 4: dialectal divergence

- ▶ Although most Mansi dialects are now extinct, Kannisto's and Chernetsov's text collections (as well as some materials by Reguly) contain large amounts of data on these language varieties.
- ▶ Also, data from Sosva and Sygva dialects is still available for gathering.
- ▶ Comparison of numerous linguistic phenomena across texts in various dialects would allow us to suggest ways of Mansi dialectal divergence and also determine the individual path of UL Mansi as compared to some Proto-Mansi state.

## Step 4: dialectal divergence

- ▶ This analysis is even more feasible if we perform it on the basis of a multidialectal corpus, where all texts are annotated in a comparable data.
- ▶ Thus, the `digital-mansi.com` corpus is multidialectal.
- ▶ The pilot version now features only UL and Standard Mansi subcorpora, which are constantly updated, however, they are to be supplied by multiple dialectal subcorpora after enough data is processed.

## Step 5: grammar writing

- ▶ After all the previous steps have been performed, it is possible to provide an accurate corpus-based grammar of the dialect.
- ▶ It should feature:
  - ▶ history of the UL Mansi population and modern sociolinguistic situation;
  - ▶ synchronic description of UL Mansi, including variation;
  - ▶ diachronic development of the described phenomena;
  - ▶ information about contact influence;
  - ▶ divergence from other Mansi dialects.

## Step 6: language support

- ▶ At all steps of our work, we try to identify the linguistic needs of UL Mansi.
- ▶ Their main concern is orthography.
- ▶ Our task: to develop an orthographic system which would accurately reflect UL Mansi traits.
- ▶ Then, together with native speakers, we need to produce and publish text materials in this orthography.
- ▶ Also, we try to extend the domain of UL Mansi language use by engaging speakers in various kinds of games involving Mansi language use etc.

# Conclusion

- ▶ To sum it up, the main goal of our project is to document UL Mansi by processing **new data** and applying (somewhat) **new approaches**.
- ▶ **New data:** field recordings + archival texts which have been lying dormant for ages;
- ▶ **New approaches:**
  - ▶ combining traditional and “experimental” methods in data collection;
  - ▶ in-depth analysis of idiolectal variation, possible due to well-pronounced sociolinguistic differences between speakers;
  - ▶ integration of UL Mansi data into a diachronic context;
  - ▶ multidialectal corpus-based approach;
  - ▶ interaction with native speakers at all steps of research.

# Thank you for your attention!

