The syntax and prosody of Tundra Nenets interrogatives

Katalin Mády Réka Metzger Nikolett Mus Péter Rebrus
Uwe Reichel

Research Institute for Linguistics, Hungarian Academy of Sciences

Workshop on Uralic Prosody, Helsinki, 21 March 2019

Background

- The Uralic languages spoken in the Russian Federation are under a strong Russian influence.
- This contact often leads to the restructuring of the originally SOV-type Uralic languages to the Russian SVO-type, i.e. from head-final to head-initial.
- There are social factors that may slow down the contact-induced changes.
- Some (word order) correlations may show whether a (potential) typological restructuring has already begun in a language (Greenberg 1966).

Background (cont.)

- Tundra Nenets is an endangered and understudied language (c. 20,000 speakers, EGIDS 6b (threatened)).
- There are three Tundra Nenets dialects: Western, Central and Eastern.
- Tundra Nenets text collections either contain a very limited amount of tokens or do not provide representative samples of Tundra Nenets and its dialects.
- The vast majority of the available audio recordings contain elicited data.
- The culture of Tundra Nenets is predominantly an oral one.

Background (cont.)

- The word order of Tundra Nenets (Northern Samoyedic, Uralic) declaratives is SOV.

(1) Sergei Masha-m? meńe(-da).
Sergei Masha-acc love.3sg/(-3sg.sg)
'Sergei loves Masha.'
S O V

- The information structure of clauses seems to have an effect on their word order.
  - The topic tends to occupy the sentence-initial position (but it can remain *in situ*, i.e. in the position for its constituent type).
  - The focus may appear in the immediately preverbal position (but it is an optional position for the focused element).

Background (cont.)

- In content questions, wh-phrases may appear in 3 structural positions:

1. The wh-phrase remains *in situ*, i.e. it is in the position for its constituent type.
   - This position correlates with the basic word order of the language, i.e. it is the 'expected' syntactic position of a wh-phrase (Greenberg 1966).
   - Based on a corpus study, 67.5% of the wh-phrases appear *in situ* (478/708; Mus 2015).

(2) xíba Masha-m? meńe(-da)?
who Masha-acc love.3sg/(-3sg.sg)
'Who loves Masha?'
S_{WH} O V

(3) Sergei xíba-m? meńe?
Sergei who-acc love.3sg
'Whom loves Sergei?'
S O_{WH} V

Background (cont.)

2. The wh-phrase occupies the immediately preverbal position, i.e. it appears in the (optional) position of the focused constituent.

- Possible reasons:
  - The object can be a topical object.
  - The wh-subject can be focused.

(4) Masha-m? xíba meńe(-da)?
Masha-acc who love.3sg/(-3sg.sg)
'Who loves Masha?' [topicalized object]
'WHO loves Masha?' [focused wh-phrase]
O S_{WH} V

Background (cont.)

3. The wh-phrase is in the sentence-initial position, i.e. in the (optional) position of the topic.

- Possible reasons:
  - The wh-object is more specific than the subject (É. Kiss 1993 'specificity constraint').
  - The subject is in the focus position.
  - The influence of the Russian language can have an effect on the position of the wh-phrase.

(5)

xurka	po-ʔ	peĺa-mʔ	pidaraʔ	xarwo-bta-daʔʔ
which	year-gen	half-acc	2pl	want-tr-2pl

'Which part of the year do you like?' [specific wh-phrase]

'Which part of the year do YOU like?' [focused subject]

O_{WH} S V

Background (cont.)

Q1 Is there a syntactically distinct clause type whose function can be associated with the function of questions?

Q2 What kind of intonation patterns do the interrogatives have in Tundra Nenets?

Q3 Can we find systematic similarities/differences in the marking strategies of the interrogatives in the dialectal variations of Tundra Nenets?

Q4 Are there any contact-induced changes in the language or in its dialectal variations?

ThEA

Theoretical and experimental approaches to dialectal variation and contact-induced change: a case study of Tundra Nenets

- Research network and research infrastructure
  - 2018–2022
  - Research Institute of Linguistics, HAS
  - National Research, Development and Innovation Office
  - participants
    - Katalin Mády
    - Réka Metzger
    - Nikolett Mus (PI)
    - Uwe Reichel
    - Péter Rebrus

Outline

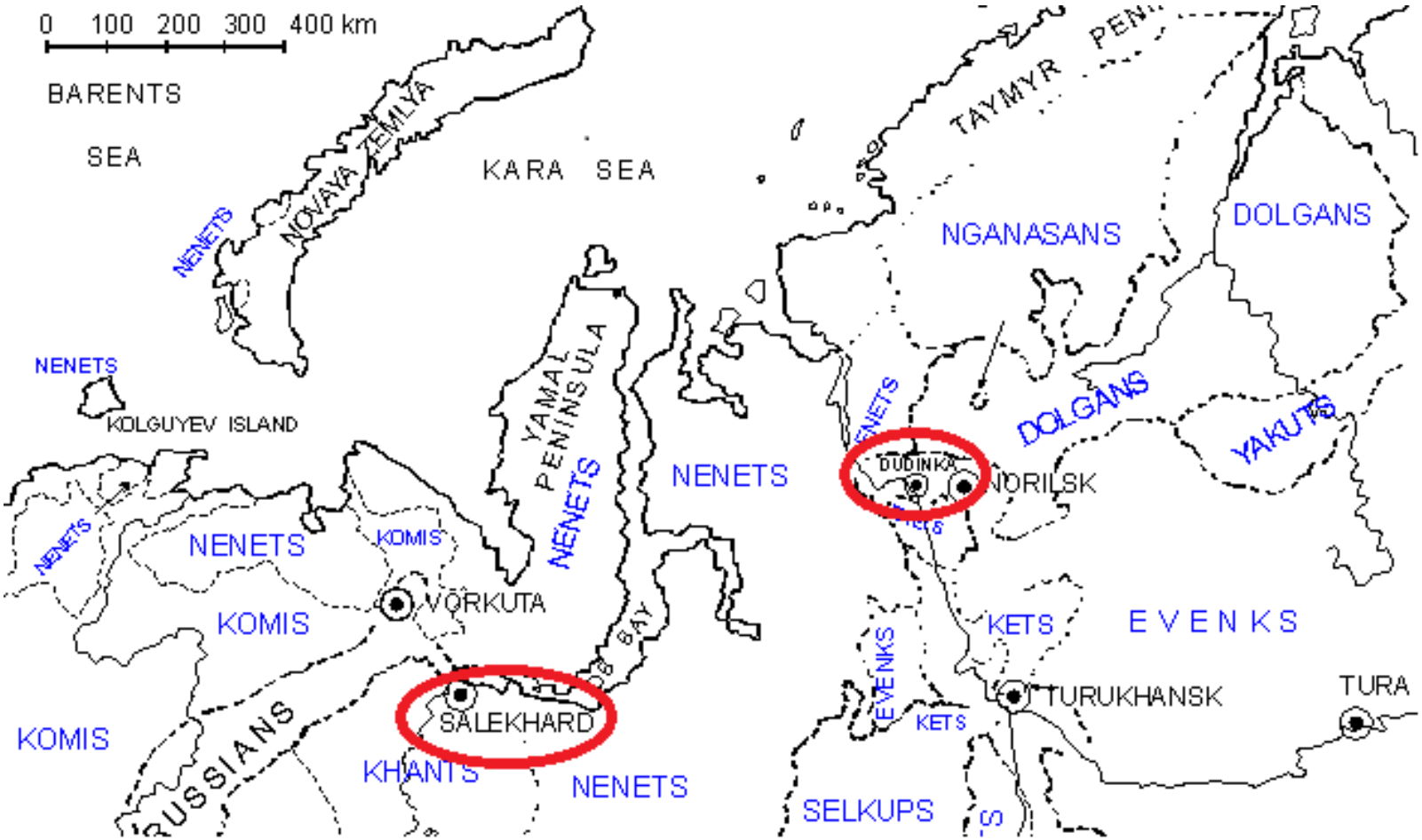
1. Aims and expected results
2. Methodological considerations and methods
3. Data processing

Theoretical starting point

- We will focus on two speech communities:
  - the one spoken on the Yamal Peninsula by traditional reindeer herders,
  - the other spoken in Dudinka, by people who settled down in the city and live a (more or less) urban life.

Theoretical starting point (cont.)

The Tundra Nenets communities



Theoretical starting point (cont.)

H1 the two dialectal variations of Tundra Nenets are influenced by the Russian language to a different extent:

- the dialect spoken on the Yamal Peninsula is supposed to represent a more conservative head-final structure;
- signs of breaking up the SOV structure in the variation spoken in Dudinka is assumed.

Subject of the research

- Our general hypothesis will be tested on a specific type of clauses, i.e. on interrogatives.

⇒ Interrogatives are presumed to be universal (Sadock & Zwicky 1985; Huddleston 1994; König & Siemund 2007; Velupillai 2012).

⇒ Interrogative marking strategies show correlations with the basic word order of languages.

Expected results

- a comprehensive description of interrogative clauses (content, polar, alternative, etc. questions) in Tundra Nenets
- a comparative analysis of interrogatives in Tundra Nenets dialects
- an online database of Tundra Nenets interrogative structures
- a parallel (and comparable) corpus of Tundra Nenets spoken data (collected during our fieldworks)

- a typological questionnaire
- a language-independent data collection toolkit

1 Aims and expected results

2 Methodological considerations and methods

3 Data processing

Experimental methods in the field

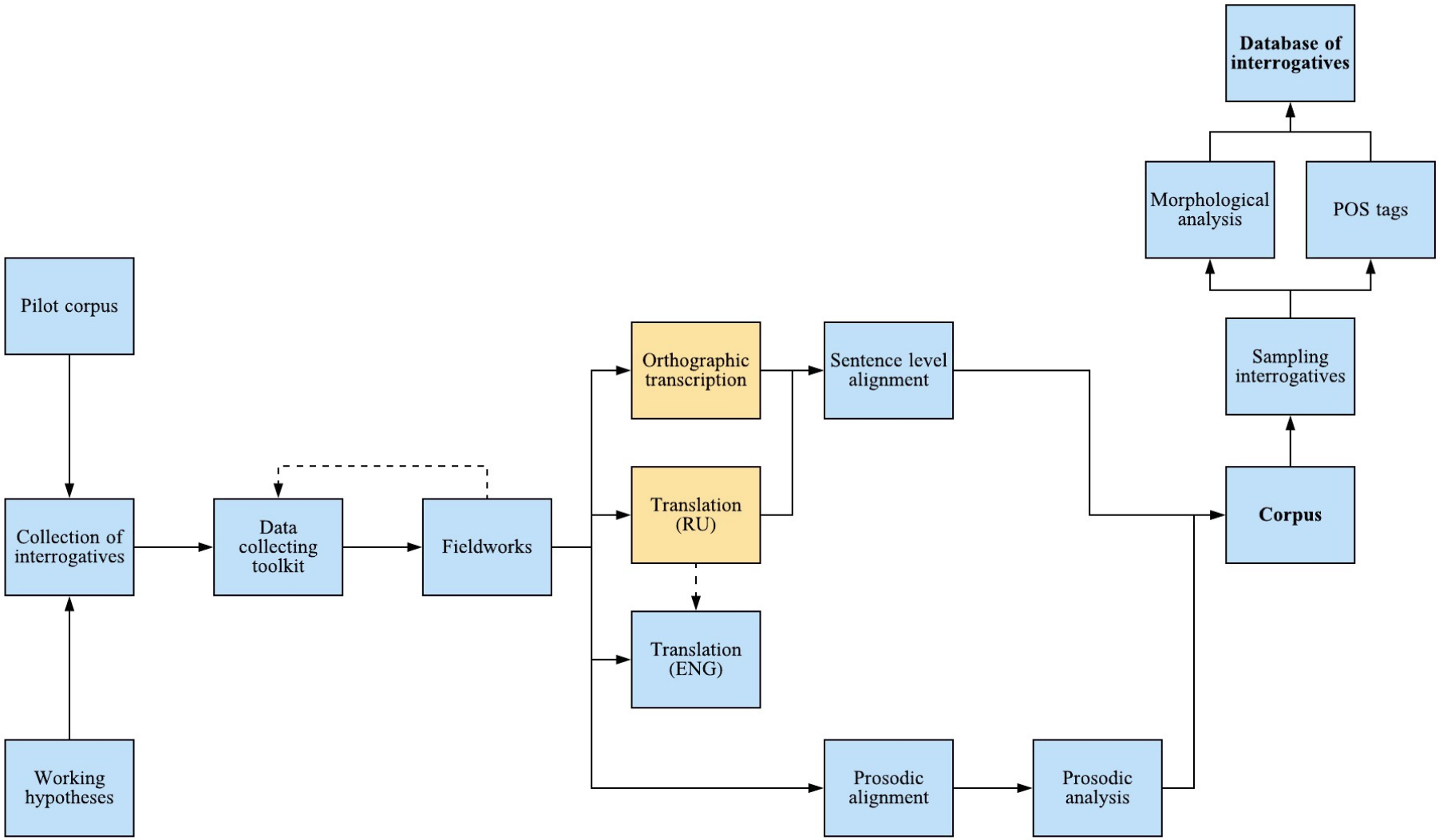
Methods of documentational linguistics

- Focus on the speaker's performance,
- collecting reliable, representative and natural data,
- documenting metadata,
- techniques
  - observed events,
  - staged events.

Methods of experimental linguistics

- Focus (also) on the speaker's competence,
- measuring grammaticality, preference and reaction times,
- systematic manipulation of variables,
- techniques
  - sentence-picture matching,
  - questionnaires.

Workflow


Database of interrogatives
Morphological analysis
POS tags
Pilot corpus
Sampling interrogatives
Orthographic transcription
Sentence level alignment
Collection of interrogatives
Data collecting toolkit
Fieldworks
Translation (RU)
Corpus
Translation (ENG)
Working hypotheses
Prosodic alignment
Prosodic analysis

1. Aims and expected results

2. Methodological considerations and methods

3. Data processing

Preprocessing

Sentence level alignment

- Following the process of UraLUID (RIL HAS, OTKA 118079)
  - conversion of the `.txt` files (Cyrillic transcritpion and the ENG/Ru translations) into `.tsv`,
  - uploading the texts to (No)SketchEngine $\rightarrow$ EN/RU/YRK$_{\text{Y}}$ and EN/RU/YRK$_{\text{T}}$ parallel corpus,
  - creating ELAN and Praat annotation files automatically by Pympi module of python3.

Preprocessing (cont.)

Signal-text alignment

- Orthographic transcription of the data,
- pronunciation dictionary generation by rule-based grapheme-phoneme mapping,
- Montreal forced aligner (McAuliffe et al., 2017) for acoustic model training and signal-text alignment on phone and word level (does not require pre-aligned training material)

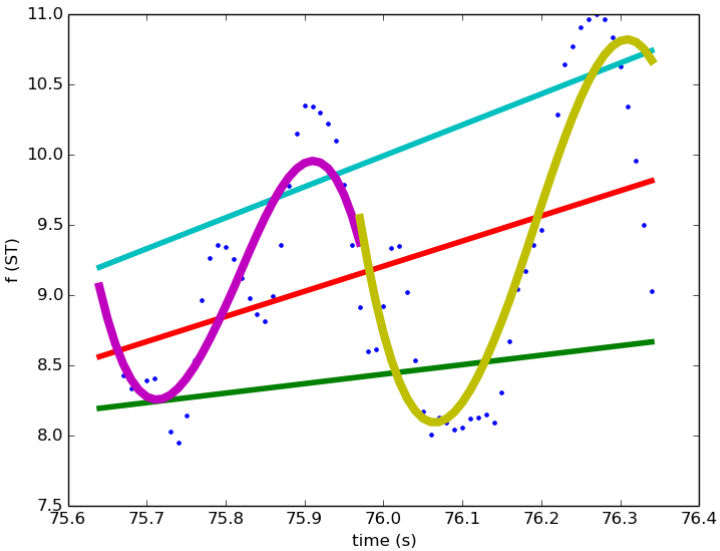
$\longrightarrow$ localization of relevant parts in interrogatives.

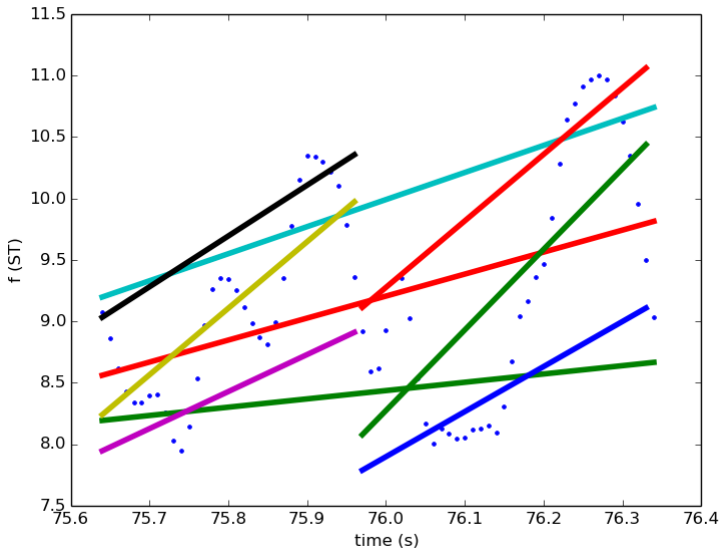
Prosodic structure

- syllable nucleus extraction,
- prosodic phrase boundary and pitch accent detection (Reichel, 2017).

Prosodic analyses

Superpositional stylization





- Register level and range patterns of interrogatives for intonation phrases and pitch accents,
- pitch accent and boundary tone inventory.

Typology of interrogatives

Intonation of interrogatives in languages

- interrogative phrases obligatorily initial (German, Russian),
- interrogative phrases not obligatorily initial (Tundra Nenets, Hungarian).

Are Tundra Nenets wh-interrogatives marked by

- a specific question intonation,
- a specific pitch accent on the wh-word,
- a certain boundary tone,
- else, e.g. subsequent deaccentuation?

Now it's time for your expert suggestions!

Elicitation techniques:

- Are there advantages of read questions during initial data collection?
- Will dialogues such as maptasks or Mr. X (Who am I) work with educated speakers of Tundra Nenets?

Equipment:

- Are there disadvantages of laptop usage in urban areas? Alternative: digital recorder with external microphone(s).
- Rural speakers might have difficulties with head-mounted microphones. Any high-quality alternatives among clip microphones? Omnidirectional vs. cardioid?

Text-aligned prosodic annotations:

- chunk and IP segmentation,
- syllable and pitch accent time stamps,
- anything else that can be done automatically?

Acknowledgements

Thank you for your attention!

The support of the research project "Theoretical and experimental approaches to dialectal variation and contact-induced change: a case study of Tundra Nenets" (NKFI 129235) is gratefully acknowledged.