



Typological Database of the Ugric Languages: establishing a long-term project

DIGITAL HUMANITIES CONFERENCE

BUDAPEST, 29.05.2018.

TIMÁR BOGÁTA – ASZTALOS ERIKA – F. GULYÁS NIKOLETT – HORVÁTH LAURA – SZABÓ DITTA – HAVAS FERENC (ELTE)



Preview

- About the project
- What we established so far: utdb.elte.hu
- Usage
- Problems to be solved
- Present activities, tasks, plans



Phase 1: Typological Database of the Ugric Languages

- OTKA K 104249
- 2012–2015
- Online database about the typological features of the Ugric languages
- Analysed languages:
 - Synja Khanty
 - Surgut Khanty
 - Northern Mansi
 - Hungarian

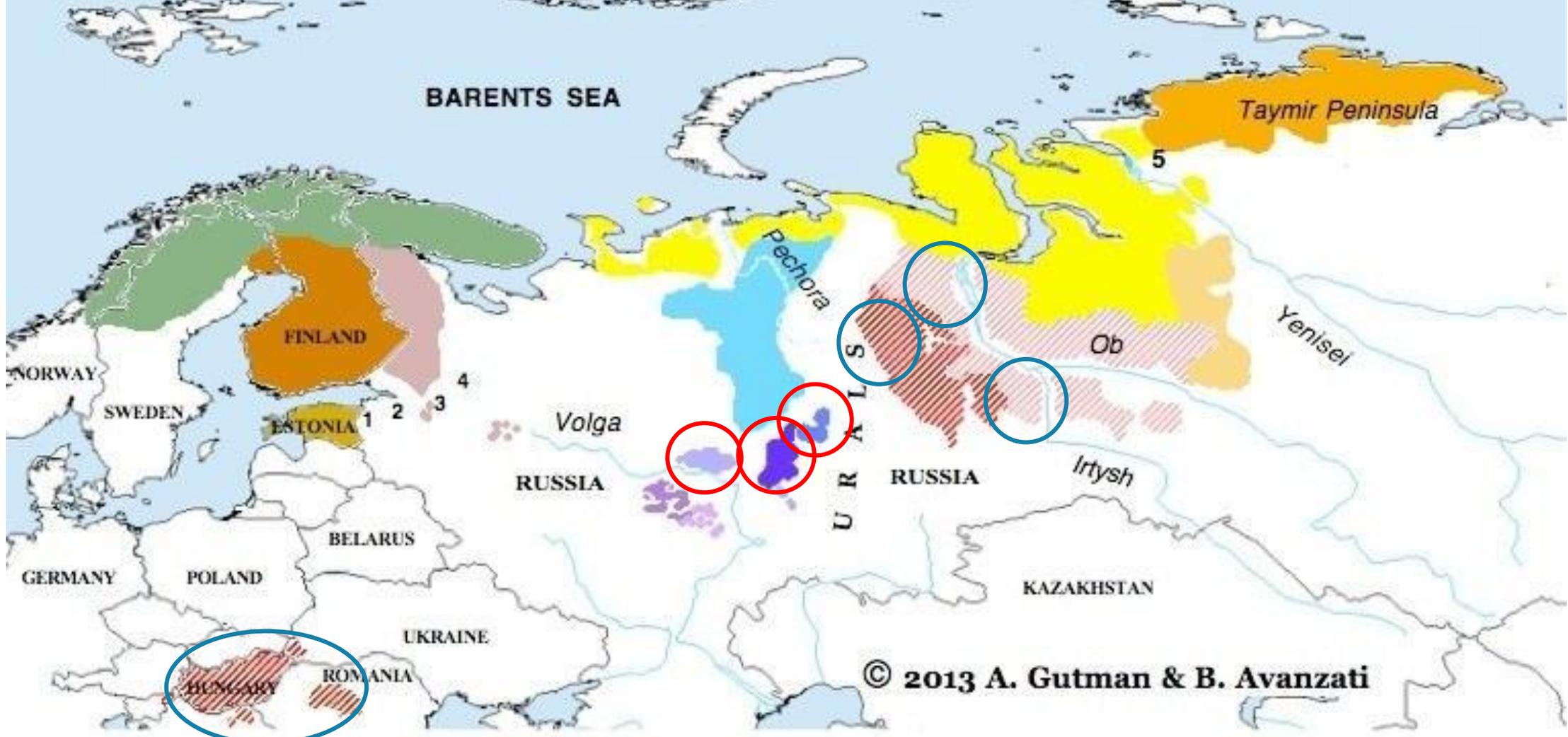
Phase 2: Typological Database of the Volga Area Finno-Ugric Languages

- NKFI-125282
- 2017–2021 (48 months)
- ELTE BTK Department of Finno-Ugristics
- Analysed languages:
 - Meadow Mari
 - Udmurt
 - Komi-Permyak

(→ long-term plan: Typological Database of the Uralic Languages)

Features and principles

- Database available in 3 languages:
 - Hungarian
 - English
 - Russian
- focus on morphology and syntax
- **Current** information and data about the analysed languages
- weekly meetings and discussions at the Department of Finno-Ugristics
- Team: Ferenc Havas, Erika Asztalos, Nikolett F. Gulyás, Laura Horváth, Ditta Szabó, Bogáta Timár



© 2013 A. Gutman & B. Avanzati

Baltic-Finnic		Volgaic		Permic		Ugric		Samoyedic		Unclassified	
	Finnish		Mari		Udmurt		Hungarian		Nenets		Saami
	Karelian		Moksha		Komi-Permyak		Mansi		Selkup		
	Estonian		Erzya		Komi-Zyrian		Khanty		Nganasan		
	1. Votic								5. Enets		
	2. Ingrian										
	3. Veps										
	4. Ludian										



WALS vs UTDB

WALS

- languages of the world
- typological presentation of the certain languages in detail is not an aim
- about 192 parameters
- phonology, morphology, syntax, lexicon
- spectacular but not always efficient
- not updated, features wrong data

UTDB

- only one language family (particular branches) – microtypology
- detailed presentation of the main morphosyntactic features of the Ugric (volgaic, uralic ...) languages
- ~213 parameters
- morphology, syntax
- Suitable for making a comparative analysis



Usage of the database: structure

- About
 - The structure of the database, sources of parameters, writing system, glossing, licence, how to cite
- Search
- Browsing
- Languages
- Sources
- Authors
- Contact us

Problems to be solved

- UTDB in Drupal system → is it really suitable for our project and plans?
- UTDB in 3 languages → 3 separated sites
- Migration from a private server to an official server of ELTE
- Glossing problems

Glossing issues

- if present tense is marked in Udmurt, should the absence of it be marked in Mari and Permyak?

- Homonymy

mon u-g malpa-śkə - SG1

so u-g malpa-śkə - SG3

- should etymology be a factor?

pal-en o-m-əl

know-PST2 NEG-1SG-is?

- Gerund? Converb? If converb, is it CVB or CNV?

Other issues

- What can be considered as data?
- Dialectal variety?
- Language contact phenomena, especially with Russian
Predicative adjective without copula in Meadow Mari:
 - literature says it's agrammatical
 - but they use it all the time
- different data in each source, e.g. the number of cases in Meadow Mari:
 - 12 (Bereczki 1990)
 - 10 (Alhoniemi 1985)
 - 8+3 (Kangasmaa-Minn 1998)
- Schrödinger's phonology: e.g. Udmurt $\ddot{o} \neq$ Meadow Mari \ddot{o}

Present activities, tasks, plans

- Working out the Udmurt, Komi-Permyak and Meadow Mari features
- field trips, involvement of native language consultants
- Specification of the writing system used at the certain languages (variances between the languages)
- Unify glossing
- Solving technical issues (migration from the private server)
- Formal corrections



References

Dryer, Matthew S. 2001. *Matthew Dryer's Typological Database* Buffalo: Department of Linguistics, University at Buffalo. <http://linguistics.buffalo.edu/people/faculty/dryer/dryer/database>

Dryer, Matthew S. – Haspelmath, Martin (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>

Michaelis, Susanne Maria – Maurer, Philippe – Haspelmath, Martin & Huber, Magnus (eds.) 2013. *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/>

OUL. Ob-Ugric Database: analysed text corpora and dictionaries for less described Ob-Ugric dialects. http://www.oudb.gwi.uni-muenchen.de/index.php?abfrage=welcome_oudb&navi=about_oudb



Thank you for your attention!

utdb.elte.hu