

# The Uralic Typology Database Project

Ferenc HAVAS, Budapest  
Congressus Undecimus Internationalis  
Fenno-Ugristarum  
12. 8. 2010 Piliscsaba

## A short (pre)history of the project

2005: Initiative launched at the Tenth International Congress of Finno-Ugrists held in Yoshkar-Ola, Mari El

2008: A closer delineation of the project in 2008 at a dedicated international conference at Vienna University

2008: Establishment of a steering committee

2008–2009: Presentation of the project at different conferences (Bratislava, Khanty-Mansijsk, Moscow, Tallinn, Szeged)

2010: CIFU 11

# **What is a typological database of the Uralic languages ?**

General outline of its structure and purpose

A database is a grid (table) consisting of *columns* and *rows* that yield *cells* in their intersection.

[illegible]

The columns represent Uralic languages...

[illegible]

... whereas the rows stand for *parameters*  
ie. typologically salient features of languages.

...						
<b>basic word order</b>						
<b>possessive construction</b>						
<b>conditional clause</b>						
...						

The content of a cell depicts the way  
the given parameter materializes in a given language.

	...	6.	7. Erzya- Mordvin	8.	...	
...						
46. ...						
47. ...						
<b>48. possessive constructions</b>			typological specification of <u>possessive constructions</u> in <u>Erzya- Mordvin</u>			
49. ...						
50. ...						
...						

Displaying a given parameter's specification  
for several languages:

	...	6. Mari (Cheremis)	7. Erzya- Mordvin	8. Moksha- Mordvin	...	
...						
46. ...						
47. ...						
<b>48. possessive constructions</b>	...	typological specification of <u>possessive constructions</u> in <u>Mari</u>	typological specification of <u>possessive constructions</u> in <u>Erzya- Mordvin</u>	typological specification of <u>possessive constructions</u> in <u>Moksha- Mordvin</u>	...	
49. ...						
50. ...						
...						



## Displaying a given parameter's specification for a whole subgroup of the language family:

	Finnish	Estonian	Karelian	Livonian	Ingrian	Veps	Votic
...	...	...	...	...	...	...	...
<b>48. possessive con- structions</b>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Finnish</u>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Estonian</u>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Karelian</u>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Livonian</u>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Ingrian</u>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Veps</u>	typological specifi- cation of <u>possessive</u> <u>con-</u> <u>structions</u> in <u>Votic</u>
...	...	...	...	...	...	...	...

# Displaying several parameter specifications for a given language:

	...	6.	7. Erzya- Mordvin	8.	...	
...						
46. ...			...			
47. possessive phrases			typological specification of <u>possessive phrases in Erzya- Mordvin</u>			
48. numeral phrases			typological specification of <u>numeral phrases in Erzya- Mordvin</u>			
49. adpositional phrases			typological specification of <u>adpositional phrases in Erzya- Mordvin</u>			
50. ...			...			
...						

# Dialects

## How many columns do we need?

- The number of living Uralic languages: at least 19 items
- Some items commonly considered as collective designations of several dialects
- Some items considered sometimes as languages, sometimes as dialects
- Often significant differences between dialects of the same language
- Conclusion: as many items (separate columns) should be taken into account as there are *typologically* different dialects
- All these items can equally be called and regarded as (representing) *dialects* in our typological database

[illegible]

**Karelian** is a language closely related to Finnish, with which it is not necessarily mutually intelligible. Karelian is spoken mainly in Republic of Karelia, Russia. Dialects spoken in Finnish Karelia (North Karelia and South Karelia) are not considered Karelian but Savonian dialects or Southeastern dialects of Finnish. Karelian is spoken by about 100,000 people mainly in the Republic of Karelia, Russia but notable Karelian-speaking communities can also be found in Tver region. Karelian is also spoken in Finland where Karelian-speakers are estimated to be around 5,000. Karelian belongs to the Finno-Ugric languages, and is distinguished from Finnish by some important extensions to the phonology and the lack of influence from modern 19th and 20th century Finnish. It cannot merely be classified as a Finnish dialect with Russian influences, because it has original innovations and it may differ considerably from Finnish. In the Republic of Karelia Karelian has official status as a minority language. Since the late 1990s there have been moves to pass special language legislation, which would give Karelian an official status on par with Russian. In Finland Karelian has official status as a non-regional national minority language. There is no standard Karelian language, although the Republic of Karelia's authorities have recently begun to attempt standardization. Each writer writes in Karelian according to his own dialectal form. The script is the Latin alphabet as used for Finnish with letters added.

# Parameters

## Parameter sources:

- WALSL
  - = World Atlas of Language Structures, <http://wals.info>  
(or search for WALSL)
- Matthew Dryer's Typological Database
  - <http://linguistics.buffalo.edu/people/faculty/dryer/dryer/database> (or search for Matthew Dryer)
- My own improvements and supplements



## Parameter specification:

- Set of parameters: a revised and enhanced inventory of typologically salient grammatical features
- Definition of parameters: precise but not oversophisticated definitions (indicating alternative terminology if necessary)


**Details: the parameter values**

## Parameter value specification:

- Set of parameter values: a possibly full set of different attested patterns of realizing the given parameter in the world's languages
- Definition of parameter values: precise and unequivocal definitions
- Format of parameter values: a possibly transparent code (string of abbreviations)
- Coding the *mixed* situations in a given dialect: combining two value codes using a restricted set of linking symbols (such as & for “equally present” and / for “both present with the first one being dominant”)

## **Details: displaying parameters**

# The placement of possessive pronouns

- **Parameter:** Placement of possessive pronouns
- **Values:**   NoPoss(NB)  
                  Poss(NB)N  
                  Nposs(NB)
- **Comments**                      ← 
- **Hints:** D 31b

# Comments

- explanation and definition of the **meaning** of a given parameter
- assistance in **identifying** the relevant phenomenon
- explanation of each possible **value** of the given parameter

# The placement of possessive pronouns

- **Possessive pronouns** are representatives of a specific part of speech – they are non-bound (NB) grammatical words that mark person and/or number (occasionally also class or gender) of the possessor next to a noun or noun phrase depicting something possessed (the possessee). They can be **placed** before or after the possessed noun (noun phrase).
- The existence of *possessive* pronoun as an independent part of speech can only be stated in a language if the grammatical words marking person and/or number of the possessor are not identical to (some form of) *personal* pronouns. For example, an ordinary genitive form of a personal pronoun representing the possessor cannot be considered a possessive pronoun. Similarly, an ordinary personal pronoun obligatorily extended with some other grammatical element in the possessor function (like the definite article in Hungarian az én...) is not a possessive pronoun either.
- Values:
  - **NoPoss(NB)**: there are *no possessive pronouns* as an independent part of speech in the given dialect.
  - **Poss(NB)N**: there are possessive pronouns in the dialect and they are placed *before* the possessed nouns (noun phrases).
  - **NPoss(NB)**: there are possessive pronouns in the dialect and they are placed *after* the possessed nouns (noun phrases).

# The agreement of adnominal adjectives with their head nouns

- **Parameter:** Agreement of adnominal adjectives
- **Values:**
  - NoAgr
  - NoAdjAgr
  - AdjAgr
  - AdjNumAgr
  - AdjCaseAgr
  - AdjClassAgr
  - AdjNumAdjCaseAgr
- **Comments**
- **Hints:**



# Agreement of adnominal adjectives comments (1)

- **Agreement of adnominal adjectives** is an obligatory *marking* of class/gender, number and case (or at least one of these features) of the nouns syntactically governing, and semantically modified by, adnominative (ie. non-predicative) adjectives *within the morphological shape of these adjectives themselves*. We can consider the agreement of adjectives depicting primary properties like size, shape, colour etc. as prototypical and we should examine their agreement with the head nouns in non-nominative (oblique) case forms, if available.<sup>1</sup>

# Agreement of adnominal adjectives

## comments (2)

- Values:

**NoAgr:** There is no marking of either class (gender) or number or case in the given dialect so no agreement between an adnominal adjective and noun can take place.

**NoAdjAgr:** No agreement in adjectival phrases, though adnominal agreement in other types of phrases occurs.<sup>2</sup>

**AdjAgr:** Agreement of adjectives takes place relating to all grammatical features of the head noun in attributive phrases.

**AdjNumAgr:** Though head nouns have several grammatical features, agreement of adnominal adjectives takes place only relating to number.<sup>3</sup>

**AdjCaseAgr:** Though head nouns have several grammatical features, agreement of adnominal adjectives takes place only relating to case form.

**AdjClassAgr:** Though head nouns have several grammatical features, agreement of adnominal adjectives takes place only relating to class/gender.

**AdjNumAdjCaseAgr:** Though head nouns have more than two grammatical features, agreement of adnominal adjectives takes place only relating to number and case.<sup>4</sup>

# Agreement of adnominal adjectives

## comments (3)


<sup>1</sup>If there are several values of the parameter that are characteristic of the given dialect, we can link together different values with the symbol „&” if they occur (as types) evenly or with the symbol „/” if the first value is dominant but the second one occurs as well. For example, **AdjAgr&AdjNumAgr** would mean a dialect in which agreement in all respects and relating only to number occurs with the same frequency; whereas, **AdjAgr/AdjNumAgr** is a dialect in which agreement in all respects is a general rule but there is a limited but considerable number of cases (in the paradigm) where agreement in number only (and not in case, for example) takes place.

<sup>2</sup>This means that agreement relating to class/gender, number and case (or at least one of them) takes place in other, non-adjectival attributive structures, e.g. with adnominal determiners.

<sup>3</sup>Phenomena such as the adjective obligatorily taking some special non-nominative (oblique) shape next to the modified noun, displaying a form which nonetheless only marks the number (and not, for example, the case form) of the head, should also be considered here.

<sup>4</sup>Following this pattern, we could set further values as well, e.g. **AdjNumAdjClassAgr** representing a dialect in which the adnominal adjective would only agree with its head in number and class/gender but not in case form.

**Details: specifying  
the parameter value for a dialect**

	...	Hungarian	...
...			
<ul style="list-style-type: none"> <li>Parameter: Agreement of adnominal adjectives</li> <li>Values:           <ul style="list-style-type: none"> <li>NoAgr</li> <li>NoAdjAgr</li> <li>AdjAgr</li> <li>AdjNumAgr</li> <li>AdjCaseAgr</li> <li>AdjClassAgr</li> <li>AdjNumAdjCaseAgr</li> </ul> </li> <li><u>Comments</u></li> <li>Hints:</li> </ul>		<b>NoAdjAgr</b> 	
...			

# Background information

Cf.

*a fekete kutyá-val*

ART<sub>DEF</sub> black-~~Ø~~ dog-**INSTR/COMIT**

‘with the black dog’

but

*ez-zel a kutyá-val*

this-**INSTR/COMIT** ART<sub>DEF</sub> dog-**INSTR/COMIT**

‘with this dog’

# Editing the database

- Creating the IT-apparatus of the database and its technical implementation as an online device
- Presenting the core stock of criteria (inventory of dialects, parameters and parameter values) for the database in English and Russian
- Researchers of the different dialects specify the parameter values for the particular dialects, prepare the background information for them
- Inserting of new dialects or parameters if necessary



## Adding a new dialect:

[illegible]

Adding a new parameter:

...						...
Basic word order						...
Possessive construction						...
New parameter						...
Conditional clause						...
...						

## Supervisory Board:

- invitation of authors
- call for offers
- peer review
- approval
- translation
- insertion

# Using the database

- The database is an online device with access provided to any researcher.
- Using the database means collecting parameter values for a particular set of typological parameters in a particular set of Uralic dialects.
  - With dialects, the particular set may consist of one or several items or all of them, whereas the number of the selected parameters should range from one to a reasonable limit.

- Displaying the search results in a form of a printable and storable table (grid)
  - Actually, the resulting table is a subset of the hidden (virtual) comprehensive database.
  - It is also equipped with the relevant hiperlinks:
    - in the column headings: textual identification and possibly closer characterization of the related dialects;
    - in the row headings: parameter specifications;
    - in the cells: background information relating to the specified parameter values.

**Positive “side effects”  
of editing the database**

- Contribution to the improvement of the grammatical description of the Uralic languages
  - Unforeseen gaps in the research on certain dialects  
→ motivation for in-depth inquiries into the dialects themselves; supply of topics for conference talks, articles, monographs
- Void cells to be filled in → supply of topics for graduate students' assignments, theses, essays, PhD dissertations



**Proceeding step by step**

*Pilot project:*

## Typological Database of the Ugric Languages

- Yugra University , Khanti-Mansiysk
- ELTE University, Budapest
  - a dedicated point in the workplan within the valid contract between the two universities
  - A pending application for a National Scientific Research Foundation grant in Hungary

Setting up further pilot projects?

Typological Database  
of the Ugric  
Languages...

Setting up further pilot projects?

Typological Database  
of the Permic  
Languages...

Setting up further pilot projects?

Typological Database  
of the Cheremis and Mordvin  
Languages...

Setting up further pilot projects?

Typological Database  
of the Finnic  
Languages...

Setting up further pilot projects?

Typological Database  
of the Samoyedic  
Languages...

Setting up further pilot projects?

Seeking investigators and  
research institutes/departments...



## Introduction to the project:

- **Uralic Typology Database Project website:**
  - <http://www.univie.ac.at/urtypol/index.html>
  - (or search for “urtypol” in your browser)

### In Russian:

- **Uralic Typology Pages:**
  - <http://uralictypology.pbworks.com/>
  - (or search for “uralictypology” in your browser)
- Article «Проект типологической базы данных уральских языков».  
*Финно-угорский мир*, 2009/4, 42–46., 2010.

- The PowerPoint presentation you have just seen will soon be accessible in both English and Russian on the

**Uralic Typology Database Project website:**

- <http://www.univie.ac.at/urtypol/index.html>
  - (or search for “urtypol” in your browser)

Thank you for your attention.  
Questions? Comments?

Ferenc HAVAS, Budapest  
Congressus Undecimus Internationalis  
Fenno-Ugristarum  
12. 8. 2010 Piliscsaba  
*havas.ferenc@btk.elte.hu*