

# The comparative method

Community of Practice in Uralic Studies (COPIUS)  
Foundations of Uralic etymology

[www.copius.org](http://www.copius.org)

Version 1.4 (last updated 30 November 2021)



Co-funded by the  
Erasmus+ Programme  
of the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Canonical phonotactical form in Uralic

**(C)V(V)(C)CV-(C)CV(C)**

# Substance and structure

- language consists of both material substance (phonemes, letters, words, etc.) and its organization (language structure, syntax)
- Sequences of symbols of different kinds, sound waves, letters, signs etc. make up the substance. Structure on the other hand is what gives order to these sequences of symbols
- Some fields of study within linguistics like syntax concentrate on language structure, in historical phonology and lexicology, the main focus is on the language material
- Ferdinand de Saussure 1916: signifié ~ signifiant
- Antoine Meillet 1925: external versus internal factors

# What drives change in language?

- change in pronunciation, often for the sake of consuming less energy for pronunciation (classic Neogrammarian view)
- analogy, i.e. changes that restructure the language system so that it is more transparent and logical (classic Neogrammarian view)
- borrowing of elements from other languages (classic Neogrammarian view)
- code-switching, i.e. using many languages in communication (the latter half of 20th century)
- the wish to express oneself in an interesting, personal way (language sociology in the 1970s), “invisible hand” (Keller), changes tied to societal roles
- normativity, i.e. learning, orthographies, power relations
- education (explicit model behaviour acquired through education)
- media & fashions (implicit model behaviour)

# Etymology and historical phonology

- initially languages were thought to be related largely based on surface level similarities found in lexicon and at different levels of morphology. Gradually systematic recurring phonological correspondences were established, which also allowed the reconstruction of proto-languages
- etymology and historical phonology are tightly bound together and form a feedback loop (Häkkinen 1987). Known phonological correspondences are used in establishing cognates (Gleichsetzung) and in assessing their correctness. It is possible to find previously undiscovered cognates just by applying known phonological correspondences. In this case they don't provide new information but do reaffirm the validity of old assumptions. Usually new discoveries in historical phonology also lead to the discovery of previously unknown cognates and to the reassessment of old cognate sets

One cannot exist without the other

**historical phonology**



**etymology**

# The comparative method

- rooted in less than rigorous comparisons made since the Renaissance (language family trees)
- János Sajnovics 1770: *Demonstratio. Idioma ungarorum et lapporum idem esse.*
- Sir William Jones 1776: Latin, Greek and Sanskrit stem from the same ancestor
- the beginning of 19th century: comparison of lexicon and syntax, the notion of sound laws (Franz Bopp, Rasmus Rask, brothers Grimm)
- the 19th century Neogrammarians (*Junggrammatiker*): writing historical phonology of singular languages, research of loanword layers (Karl Bruggerman, August Leskien, Karl Verner)



# Notable milestones

- Wilhelm Thomsén *Über den Einfluss der Germanischen Sprachen auf den Finnisch-Lappischen. Eine Sprachgeschichtliche Untersuchung* 1869 / 1870  
→ scientific loanword research (cf. Mikkola 1890 *Die Ostseefinnischen Lehnwörter der Russischen*)
- Hermann Paul *Prinzipien der Sprachgeschichte* 1880  
→ theoretical basis of the Neogrammarian school and the regularity of sound change
- the later half of the 20th century: typology, internal borrowing = borrowing within a language family/between branches, substrates in the comparative method

# The comparative method

- diversity and variety are explained from a singularity known as a proto-language
- proto-language is an abstraction postulated based on the properties common to synchronic languages (Finnish, Karelian, Veps, Estonian, Votic, Livonian etc. → Proto-Finnic, Swedish, Danish, Icelandic, German, English, Gothic etc. → Proto-Germanic, Polish, Slovenian, Bulgarian, Russian etc. → Proto-Slavic)
- a few key terms **cognacy** (etymologically related words), **sound law** (Lautgesetz), perhaps better to use the more neutral **sound change**. Correspondences between modern languages derived from a proto-language through regular sound changes
- a sound correspondence between Finnish and Hungarian, word-initial *p*- in Finnish corresponds to *f*- in Hungarian: *puu* ~ *fa* 'tree', *pää* ~ *fej* 'head', *poika* ~ *fiú* 'boy, son', *pesä* ~ *fészek* 'nest', *pelätä* ~ *fél* 'to fear'

# Sound change

- there are two types of sound change, **absolute** and **conditioned**. In absolute change all instances of the sound in question are subject to change. In conditioned change, the change is dependent on the sound environment. Some typical factors for conditioned changes are:
  - 1) stressed and unstressed environments behave differently
  - 2) word-initial and word-internal positions often develop differently
  - 3) sounds often develop differently when preceded by a front vowel versus a back vowel
  - 4) assimilation to a neighbouring sound
- on a closer look the comparative method doesn't so much compare words but sound sequences
- note that phonological correspondence and sound change are two different things!

# Typical sound changes

- *\*ti, te > či, če > se, si, se*, cf. Fi *käsi ~ käden ~ kätenä* (< *\*käte : käten : kätenä*), Latin *centum* [k] > French *cent* ‘100’ [s]; Latin *centum* ~ Lith *šimtas* (< *\*kmtom*)
- *\*kV\_front > č*: Sw *kära, kyrka* ‘church’ (~ Fi *kirkko*, cf. Votic *čirikkö*), Latin *caballus* ~ French *cheval* ‘horse’
- *\*kV\_back > h*: Fi *kala* ~ Hung *hal*, Ru *bogatyj* ~ Fi *pohatta*
- *s, š > h*: Fi *hapan* ~ Md *šapamo* ~ Hung *savanyú* ‘bitter’, Sw *sju, sjuk*
- epenthetic vowels in clusters with a lateral: Sl *\*mleko* > Ru *moloko* ‘milk’, Fi *pölkkypää* > dial. *pölökkypää*

## Typical sound changes II

- chain shifts are typical of vowels, change in one part of the vowel system sets in motion a number of other changes, cf. the so-called Great Vowel Shift in English, Saami
- vowel lengthening in voiced environment: Fi *kärme* → *käärme* ‘snake’, *parma* → *paarma* ‘horse fly’
- rhotacism : *s* > *r*, cf. Fi *kuningas* ‘king’ < PGerm \**kuningaz* → Icelandic *konungur*, Sw *kung*, Latin *genus*: *generis*
- all types of assimilations, dental/alveolar nasals becoming velarized preceding a *k* (*pankki*, *punkki*), voiceless consonants becoming voiced between vowels (Fi *pata*, *sata* ~ Karelian *sada*, *pada*)

# Compare

- Sound laws are exceptionless: *Die Lautgesetze kennen keine Ausnahme* (Leskien)
- Every word has its own history: *Chaque mot a son histoire* (Meillet)
- Exceptions to sound change are usually explained by analogy, dialectal/internal borrowing or by some unknown conditioning factor
- Derivation, semantic change, borrowing, contamination, analogical leveling etc. make so that exceptionless sound laws very rarely explain every single detail of a word's history

# What is analogy?

- analogy is a counterforce to sound change
- “sound laws are regular, but cause irregularity; analogy is irregular, but increases regularity” (Hermann Paul)
- *lehti: lehden: lehteä* < *\*lešte : lešte: lešetä*
- *\*ti > si*      *\*š > h*      *\*kt > ht*
- *\*lakti : lakten: lakteta > laksi : lahden: lahtea* →
- *lahti: lahden: lahtea*

# Two parts of the comparative method

1. Internal reconstruction: based on the internal variation within one language:  
(*sormus* : *sormuksen* < \**sormukse*, *luukas* : *luukkaan* < \**luukkas* ~ *luukkasen*,  
*joutsen*: *joutsenen*, *sydän*: *sydämen*)
  - can be applied to proto-languages as well
2. Comparative reconstruction: based on several languages

Fi *peura*      Ka *petra*      Est *põder*      < PF \**petra*/\**pëtra* ‘deer;  
elk’

Fi *seura*      Ka *sepra*      Est *sõber*      < PF \**sepra*/\**sëpra*  
‘company’



# Phonemes and words

Finnish	p	ä		ä			
		y	d	i	n		
Saami	b	á	kŋ	i			
		a	ď	a			
Erzya	p	e					
		u	d'	e	m	e	
Mansi	p	ä	ŋk				
	v	ä	l	ø	m		

# Phonemes and words

Hungarian

f

e

j

v

e

l

ő

Komi

p

o

m

v

e

m

PU

\*p

ä

ŋ

i

\*w

ü

ð

i

m

i

# Phonemes and words: comparison

Finnish	p	i		i
	k	ä	s	i
Saami	b	a	tn	i
	g	ie	ht	a
Erzya	p	e	j	
	k	e	d'	
Mari	p	ü	j	
	k	i	d	

# Phonemes and words: comparison

Komi	p	i	ń
	k	i	
Khanty	p	e	ŋk
	k	a	t
Hung	f	o	g
	k	é	z

# Internal reconstruction

*käsi: käden: kättä < \*käte : käten : kätetä* ‘hand’

*lasi: lasin: lasia* ‘glass’

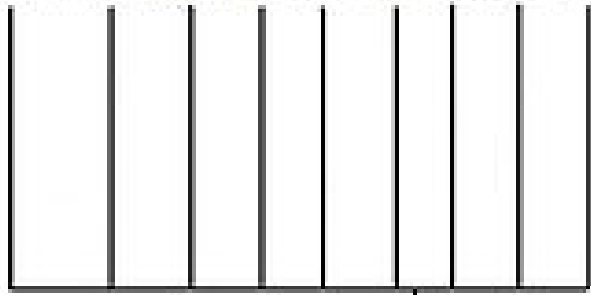
*kuusi: kuusen: kuusta < \*kuuse : kuusen : kuusetä* ‘spruce’

*lehti: lehden: lehteä < \*lešte : lešten : lešetä* ‘leaf’

*lahti ~ laksi: lahden: lahtea < \*lakte : lakten : lakteta* ‘bay’

# Taxonomy of Uralic languages I

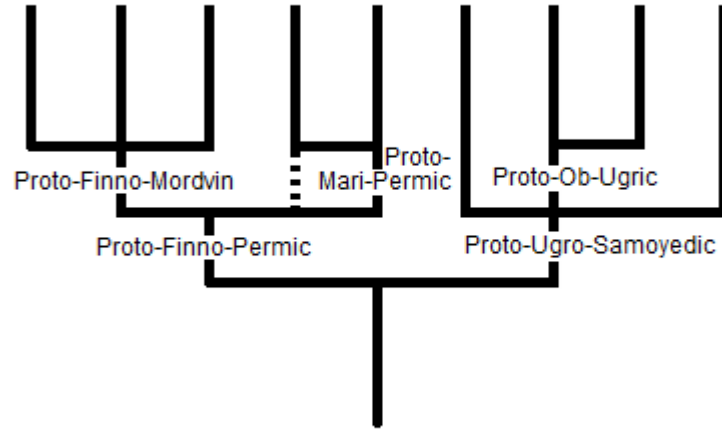
Saami Bälmeonnuozzi Mordva Mari Permi Utkusi Mansi Hanti Samoyedi



Uralilainen karttakieki

(Uralilainen sukupuu Tapani Salmisen mukaan (1999))

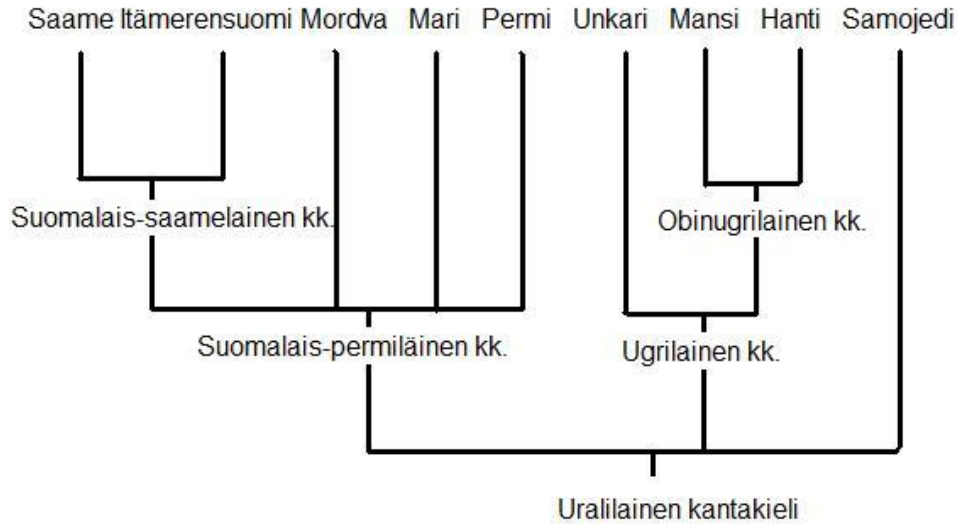
Saami Finnic Mordvin Mari Permic Hungarian Mansi Khanty Samoyed



Jaakko Häkkinen 2007

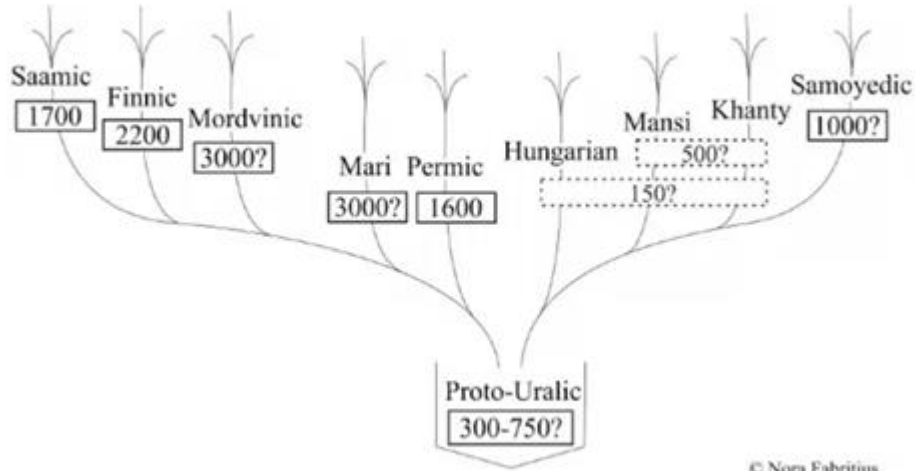
Proto-Uralic

# Taxonomy of Uralic languages II



*Uralilainen sukupuu Ulla-Maija Kulosen mukaan (2002).*

# Shared vocabulary between branches





# Proto-Uralic lexicon

- How many words can be reconstructed for PU?
- UEW = Uralisches etymologisches Wörterbuch, Károly Rédei et. al. 750 words (divided into Proto-Uralic + Proto-Finno-Ugric + Finno-Permic proto-languages)
- Janhunen 1981: Uralilaisen kantakielen sanastosta: 150 certain PU words i.e. such words that have a cognate in Samoyedic (later research has found around 20-30 new cognates that follow regular sound correspondences)
- Sammallahti 1988: 350 words (Proto-Uralic + Proto-Finno-Ugric + Finno-Permic, stricter criteria for cognates than in UEW)
- Ante Aikio, Uralic etymological dictionary (UED) in progress : even 900 words, lists all cognates between any two Uralic languages (excluding neighbouring languages with long lasting contacts like Finno-Saamic)

# Three layer structure

Modern languages (Votic, Skolt Sami, Udmurt, Hungarian, Vach Khanty, Erzya, Hill Mari, Tundra Nenets etc.)



Immediate proto-languages for individual branch (Proto-Finnic, Proto-Saami, Proto-Mordvin, Proto-Mari)



Proto-Uralic/Proto-Finno-Ugric

## Three layer structure II

- proto-languages for individual branches (välikantakieli in Finnish) are based on the comparison between the dialects of a language, Proto-Mari is based on the comparison between Meadow Mari, Hill Mari, Northwestern dialects, Volga dialect etc., Proto-Mansi on the comparison between Tavda Mansi, Konda Mansi, Pelymka Mansi, Sosva Mansi etc.
- the goal is to explain observed variation in modern languages to a common, often invariable, ancestor
- Proto-Uralic is a so-called second level reconstruction in that it's based on the comparison between these lower level proto-languages
- internal reconstruction can be applied to Proto-Uralic and other proto-languages as well