

THE LAYERS OF SUBSTRATE VOCABULARY IN WESTERN URALIC

Luobbal Sámmol Sámmol Ánte (Ante Aikio)
Sámi Allaskuvla, Guovdageaidnu

Sub-Indo-European Europe: Problems, Methods and Evidence
(Leiden, Aug 30, 2021)

The topic of my talk

- Unknown, non-Uralic substrate languages in the western Uralic zone:
 - can they be identified?
 - can periods of prehistoric substrate influence be dated and located?
 - what do substrates reveal of Western Uralic linguistic prehistory?

URALIC LANGUAGES

F Finno-Ugric

FO Baltic-Finnic

- FO1 Finnish
- FO2 Karelian
- FO3 Veps
- FO4 Ingrian
- FO5 Estonian
- FO6 Votic
- FO7 Livonian

FS Sami languages

- FS1 Western Sami
- FS2 Central Sami
- FS3 Eastern Sami

FU Ugric

- FU1 Hungarian
- FU2 Mansi
- FU3 Khanty

FP Finno-Permic

- FP1 Komi-Zyrian
- FP2 Komi-Permyak
- FP3 Udmurt

FW Finno-Volgaic

- FW1 Mari
- FW2 Mordvinic

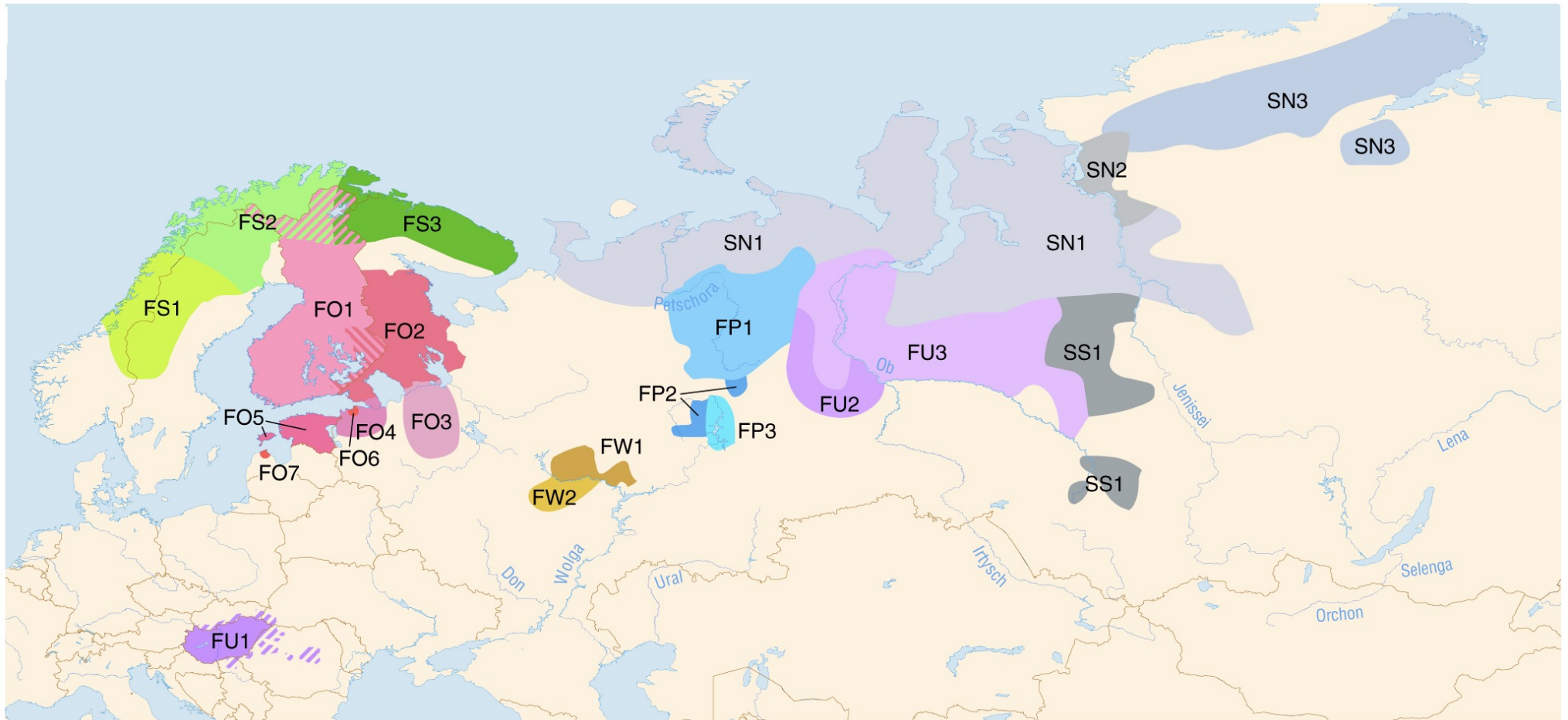
S Samoyedic

SN Northern Samoyedic

- SN1 Nenets
- SN2 Enets
- SN3 Nganasan

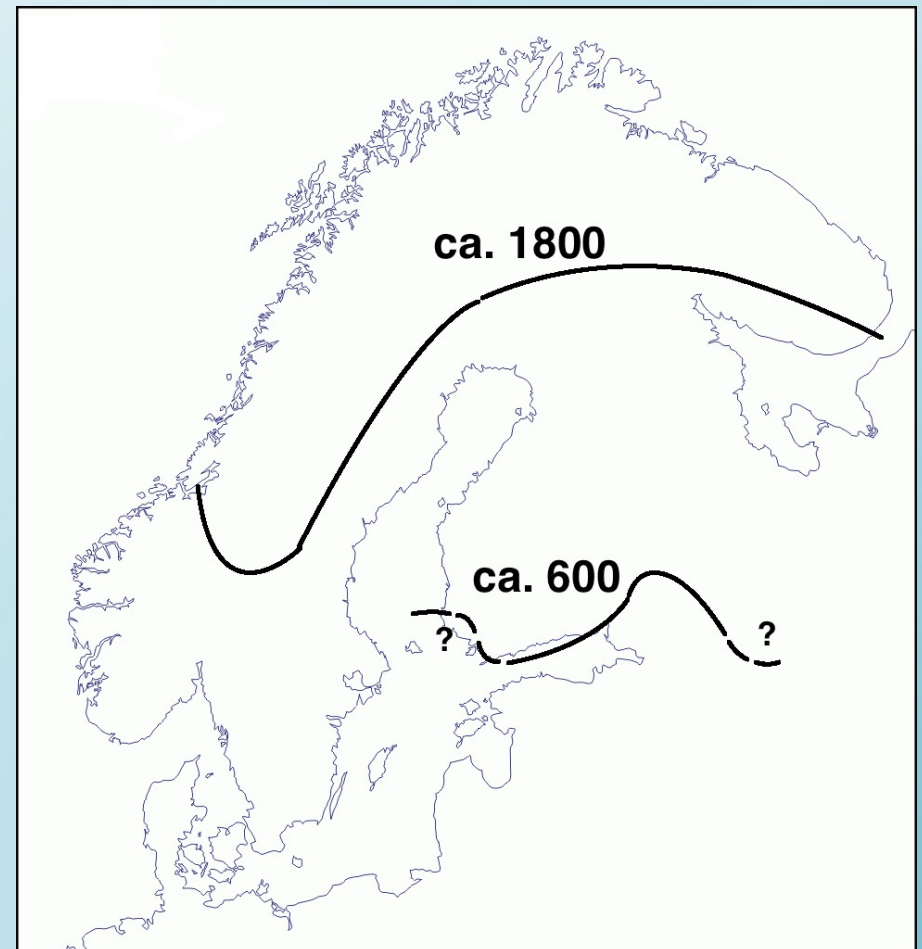
SS Southern Samoyedic

- SS1 Selkup



The prime example of substrate: Saami

- Saami languages are currently spoken in **Lapland**; earlier they were also spoken in Southern Finland and Karelia
 - **the extreme northern periphery** of the Uralic family
 - **one of the last regions** into which Uralic has spread
- Substrate influence is thus expected!



Identifying unattested substrates

- I have proposed a set of **five criteria** for the identification of substrate influence from an unattested language (Aikio 2004; 2012).
 1. **The quantitative criterion:** are potential substrate words present in large numbers?
 2. **The structural criterion:** do the words display non-native phonological or morphological structures?
 3. **The criterion of irregular correspondence:** do the words show irregular sound correspondences between languages or dialects?
 4. **The semantic criterion:** do the words tend to cluster in semantic fields typical of substrate vocabulary?
 5. **The onomastic criterion:** are there also etymologically opaque place-names displaying the same kinds of non-native structures?

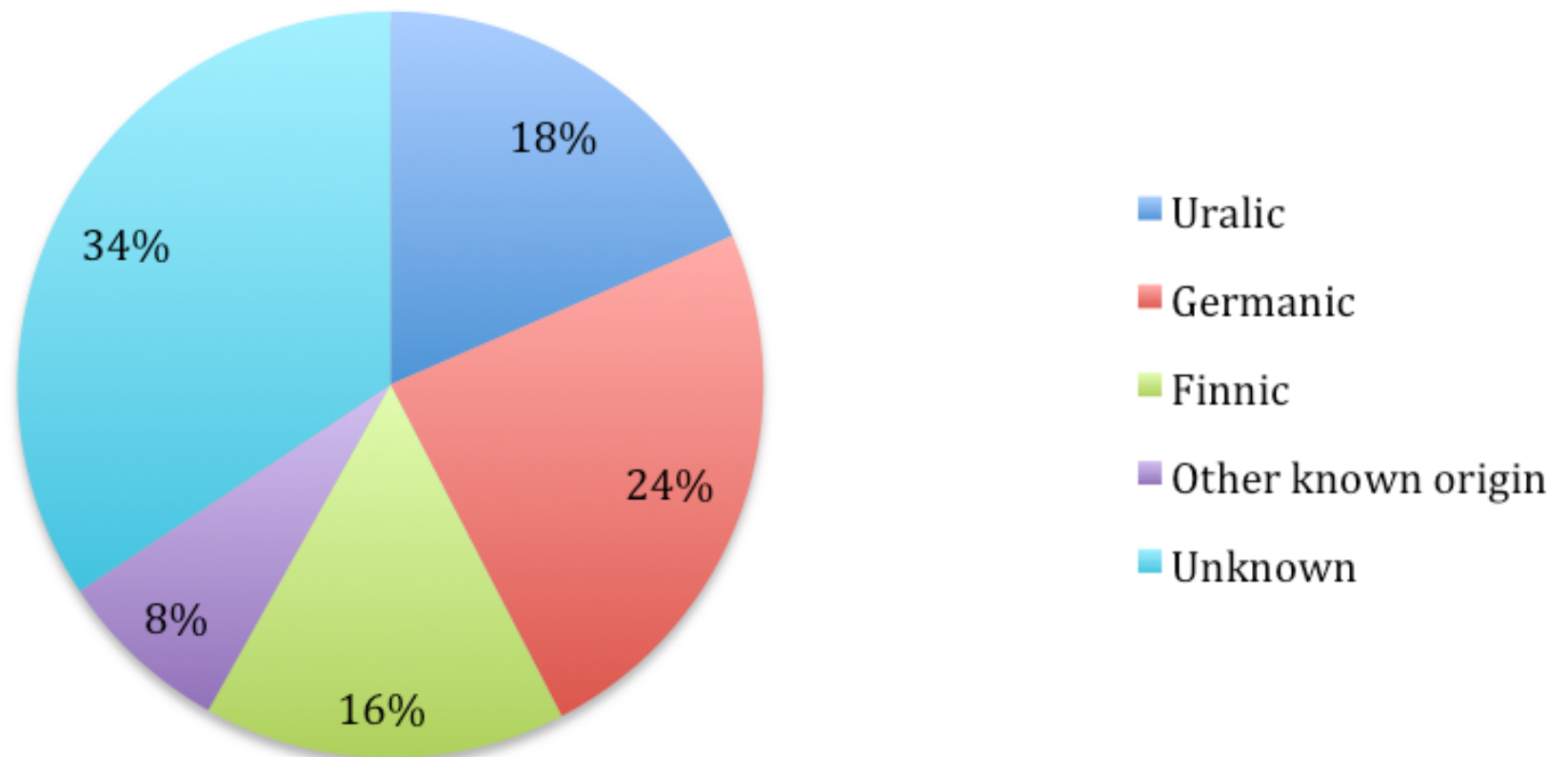
Saami etymology: the big picture

- My corpus for the etymological study of Saami languages:
 - **3421 reconstructed Saami word-roots** and their etymologies
 - Included are all monomorphemic word-roots that:
 - are attested in at least three Saami languages
 - show regular sound correspondence

Saami etymology: the big picture

- My corpus for the etymological study of Saami languages:
 - **3421 reconstructed Saami word-roots** and their etymologies
 - Included are all monomorphemic word-roots that:
 - are attested in at least three Saami languages
 - show regular sound correspondence
- **Key results:**
 - Well over a half of Saami vocabulary stems from just three known sources: Uralic, Finnic, Germanic
 - A third of the vocabulary is of unknown origin

The origins of Saami word-roots



The prime example of substrate: Saami

All five criteria are clearly fulfilled (see Aikio 2012; 2004):

1. Lots of potential substrate words
2. The words often show non-native phonotactic features
 - Uralic origin unlikely or impossible
3. Some words show irregular correspondences between languages or “duplets” within the same language
 - Parallel or consecutive borrowings of the same word
4. Clear semantic clustering
 - animal names, reindeer vocabulary, topographic features, weather terms, etc.
5. Confirmed by place-names
 - names of major topographic features are often opaque and structurally non-native looking

What about Finnic?

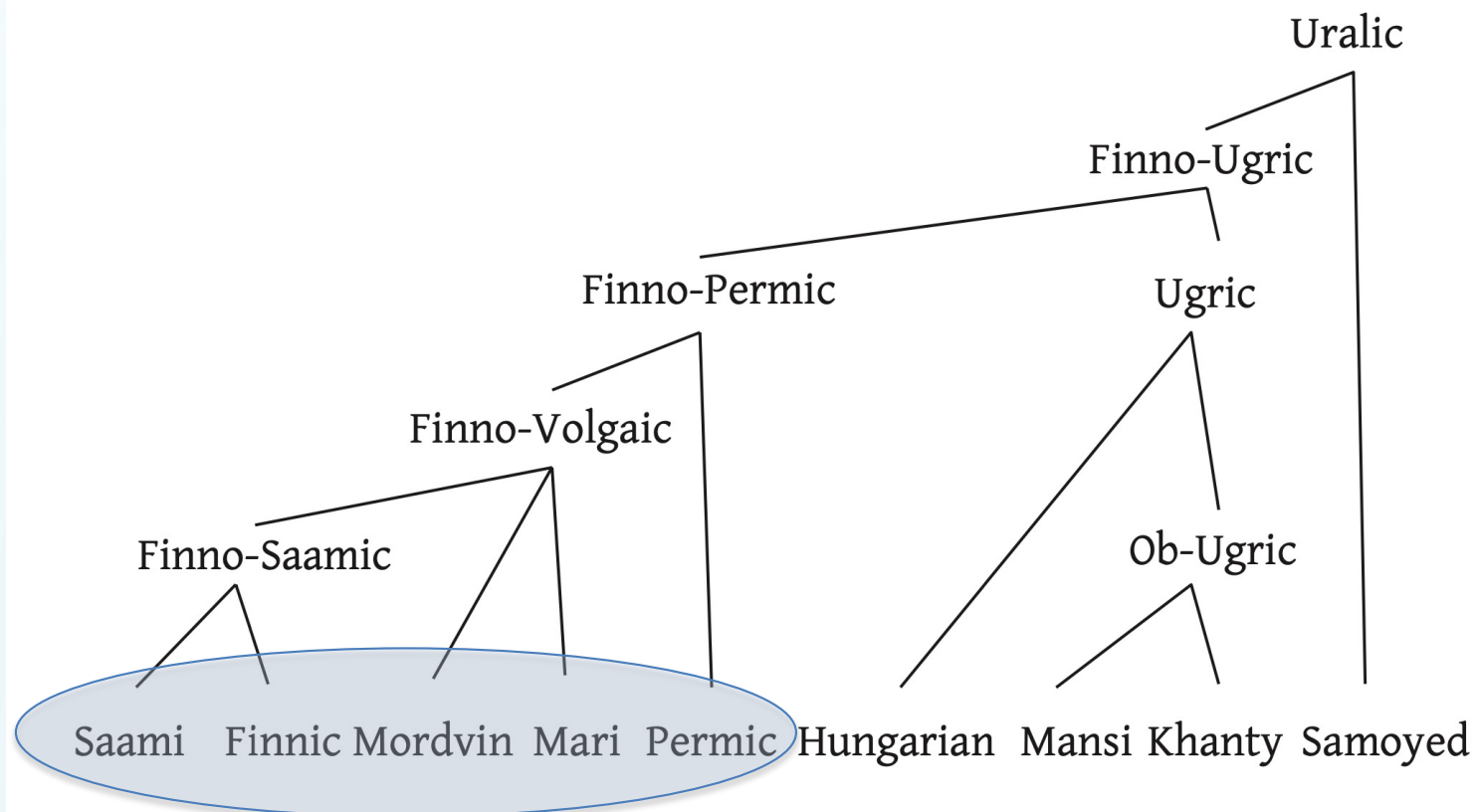
- Do we find something similar to the case of Saami in **Finnic**, the neighboring Uralic branch?
- The issue has been analyzed by Petri Kallio (unpublished):
 - Looking at distinctly Finnic vocabulary, the answer seems to be **no** (or at least not a clear “yes”).
 - far less words of unknown origin
 - non-native structural features and irregular correspondences not as frequent
 - little semantic clustering, many basic vocabulary items

What about Finnic?

- Do we find something similar to the case of Saami in **Finnic**, the neighboring Uralic branch?
- The issue has been analyzed by Petri Kallio (unpublished):
 - Looking at distinctly Finnic vocabulary, the answer seems to be **no** (or at least not a clear “yes”).
 - far less words of unknown origin
 - non-native structural features and irregular correspondences not as frequent
 - little semantic clustering, many basic vocabulary items
 - Finnic may instead have a substrate from a **known source**.
 - The **Baltic loanwords of Finnic** show semantic features consistent with **substrate** rather than superstrate!

What about “Western Uralic”?

- **Western Uralic** refers here to “Finno-Volgaic” and “Finno-Permic” groupings in the traditional (but now obsolete) taxonomy of Uralic



Western Uralic – previous research

- **Häkkinen (2009: 37–40)**
 - **Finnic and Mordvin** moved to the broad-leaf forest zone south of the Ladoga-Volga line, acquiring substrate
 - Substrate words include tree names and some agricultural terms
 - The theory is based on a very small number of words
- **Zhivlov (2015):**
 - analyzes 20 potential substrate words in "Finno-Volgaic" vocabulary
 - non-Uralic phonological features and irregular sound correspondences
 - semantic clustering (agriculture and food processing; livestock terms; tree names)

Western Uralic – previous research

- **Aikio (2015: 43–47)**
 - Uralic etyma containing **the phoneme *š** tend to display anomalous features:
 - the phoneme is very rare in Proto-Uralic etyma
 - distribution skewed towards the western part of the family (“Finno-Permic” / “Finno-Volgaic”)
 - non-native phonotactic features
 - semantic clustering: animal husbandry, agriculture, natural environment, “primitive tools”
 - Also many other traditional “Finno-Permic” / “Finno-Volgaic” etymologies may belong to the same stratum.

Compiling more material

- **Research material:**
 - the vocabulary traditionally reconstructed for “Proto-Finno-Volgaic” or “Proto-Finno-Permic”
- **Method:**
 - identify potential substrate words by three criteria:
 - foreign phonological / phonotactic features
 - irregular sound correspondences
 - semantic clustering
- **Result:**
 - 83 candidates for substrate words identified (still likely to increase)

Analyzing the material - phonology

- 37/83 (45%) of the words contain **the phoneme *š**
- Often there are also **non-native consonant clusters**



***wakštVrV**
'maple'



***wešnä**
'wheat / spelt'



***päkšnä**
'lime-tree'

Analyzing the material - phonology

- Some words have **word-initial *r-**.



?*riņiši 'drying kiln'

?*räppä(-nä) 'smoke hole'

Analyzing the material - phonology

- Irregular correspondences are frequent:

Finnish *lypsää* 'to milk'

< ***lüpsä-** or ***lүpćä-**

Mordvin *lovso*, *lofca* 'milk'

< ***lupsV** or ***lipsa**

Mari *lüştem*, dial. *lüştem*,
laštem 'to milk'

< ***lüstä-** ? < ***lüps-tä-**

Komi *lįstį-* 'to milk'

< PNo ***lüćtV-** or ***lućtV-** (? < ***lü/upć-tA-**)



Analyzing the material - phonology

- Irregular correspondences are frequent:

Finnish *pähkinä*, *pähkenä* '(hazel)nut'

< ***päškinä** (?)

Mordvin *päštä*, *päščä* (etc.) '(hazel)nut'

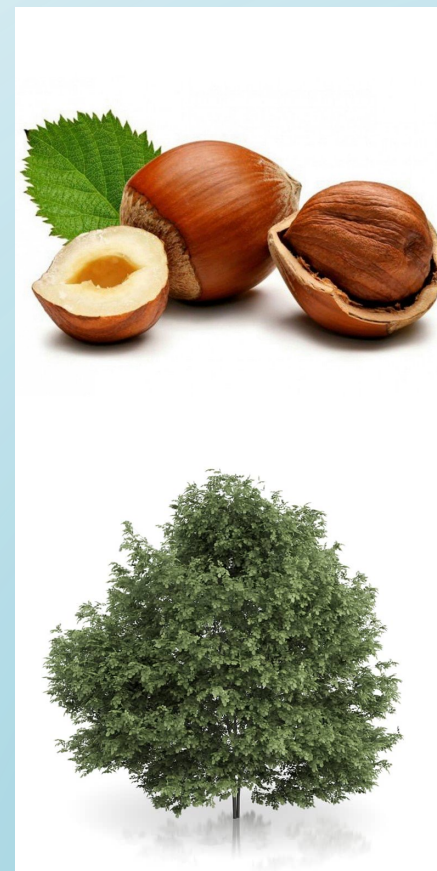
< ***päšti**

Mari *pükš* 'hazelnut'

< ?***pekši**

Udmurt *paš-*, *puš-* 'hazel(nut)'

< ***pVškV** or ***pVkšV**



Analyzing the material – semantic clustering



***tammi** 'oak'



***särnä** 'ash'



***ša/u(w)p(k)a** 'aspen'



***le/i(j)p(p)ä** 'alder'



***pVškV(nä)** 'hazel'



***wakštVrV** 'maple'

Analyzing the material – semantic clustering

- 18(+) words connected with **agriculture** or **animal domestication**
 - 'spelt / wheat', 'malt / barley', 'slash-and-burn field', etc.
 - 'pig', 'cow / horse', 'to milk'
- 30(+) words describing the **natural environment**
 - tree names, bird names, etc.
- 19(+) words for tools and other objects of material culture
 - 'splint holder', 'pestle', 'grindstone', 'back of knife or ax blade', 'mortar', etc.
- The three categories cover about **80%** of the material (67/83)!

The five criteria applied to the material

1. The quantitative criterion

- 83(+) words is far less than the over 1000 in Saami, but still quite a large portion of the shared Western Uralic vocabulary

2. The structural criterion

- Non-Uralic phonological features are very frequent

3. The criterion of irregular correspondence

- Most words show irregularities ($60/83 = 73\%$)

4. The semantic criterion

- Strong semantic clustering (80 % of words fall into just three semantic categories)

5. The onomastic criterion

- Remains unclear in light of present research

Interpretation

- There is a very strong case for early substrate influence of unknown “š-languages” on Western Uralic.
- **Dating:**
 - Irregular sound correspondences imply that the Western Uralic branches had already diverged
 - Contemporaneous with Indo-Iranian contacts?
- **Location:**
 - During the early phases of westward spread of Uralic?
 - No doubt before the spread of Finnic and Saami to the Baltic Sea region

Further questions

- A large part of the vocabulary (65%) is **missing from Saami**.
 - Did Pre-Proto-Saami originally develop farther from the influence of “š-languages”?
 - Did the Saami branch originate in Western Uralic agricultural communities, and later lose this vocabulary?
- Contact between **Indo-European (Indo-Iranian)** and “š-languages” looks like a plausible hypothesis.
 - Indo-Iranian loans are most prominent in Western Uralic (excluding Saami), and show broadly the same types of irregular correspondences
 - Did Western Uralic assimilate speakers of “š-languages”, simultaneously coming into contact with Indo-Iranian?

Thank you!

Key references:

- Aikio, Ante 2004: An essay on substrate studies and the origin of Saami. — *Mémoires de la Société Néophilologique de Helsinki* 63: 5–34.
- Häkkinen, Jaakko 2009: Kantauralin ajoitus ja paikannus: perustelut puntarissa. *Journal de la Société Finno-Ougrienne* 92: 9–56.
- Kallio, Petri (unpublished paper): Substrates in Finnic.
- Luobbal Sámmol Sámmol Ánte (Ante Aikio) 2012: An Essay on Saami Ethnolinguistic Prehistory. *Mémoires de la Société Fenno-Ougrienne* 266. 63–117.
- Luobbal Sámmol Sámmol Ánte (Ante Aikio) 2015: The Finnic ‘secondary *e*-stems’ and Proto-Uralic vocalism. — *Journal de la Société Finno-Ougrienne* 95: 25–66.
- Zhivlov, Mikhail 2015: Неиндоевропейский субстрат в финно-волжских языках. (Presentation handout, available at <https://rggu.academia.edu/MikhailZhivlov>)

Contact:

anaikio1977@gmail.com