

FEDOR ROZHANSKIY (Tartu—St. Petersburg), *MIKHAIL ZHIVLOV* (Moscow)

VOTIC AND INGRIAN CORE LEXICON IN THE FINNIC CONTEXT: SWADESH LISTS OF FIVE RELATED VARIETIES

Abstract. The main goal of this paper is compiling the Swadesh lists for five Finnic varieties: Votic, Estonian, Finnish, and the Soikkola and Lower Luga dialects of the Ingrian language. The lists are compiled using the methodology developed by the Moscow School of Comparative Linguistics. The meaning of the target words is specified not just with the translation equivalent but also with the context. Words for the lists are selected in cooperation with native speakers who help to choose the most suitable word from several synonymic variants. The resulting lists contain 111 words. For each word, etymological comments are provided. The paper also offers some preliminary observations concerning the core lexicon of the discussed varieties. In particular, we investigate the lexicostatistical distances between the languages and analyse the directions of borrowings. One of the conclusions of the research is that the lexicostatistical difference between closely related languages does not have a strong correlation with their genetic distance. The three minor varieties (Votic and two Ingrian dialects) are more similar to each other than to either of the major languages (Estonian and Finnish). The latter demonstrate more language-specific items in the core lexicon.

Keywords: Finnic languages, Votic, Ingrian, Swadesh list, lexicostatistics, comparative studies.

1. Introduction

The core idea of lexicostatistics is the analysis and comparison of wordlists compiled from the most stable part of the lexicon. The compilation of such lists is not a trivial task that can be solved through simple searching of translation equivalents in a dictionary. Synonymy, dialectal variation, and other factors significantly influence the composition of the list and correspondingly the final result of the research.

In this article, we present the 111-word Swadesh lists for five Finnic idioms. The core of the research are the wordlists for three minor varieties: a dialect of the Votic language, and two dialects of the Ingrian language. These are analysed and compared with the wordlists for two major Finnic languages: standard Estonian and standard Finnish.

The research has the following goals:

- (1) to compile wordlists of five Finnic varieties applying the same methodology;
- (2) to analyse and compare the materials from the minor varieties with the data from the major languages;
- (3) provide comments on particular words in order to make the content of the lists and the differences between the analysed varieties more transparent;
- (4) to draw a lexicostatistic picture of the minor varieties in the context of the major Finnic languages;
- (5) to make some other preliminary observations based on the compiled wordlists.

The existing lexicostatistical research on the Uralic languages rarely uses explicit Swadesh lists. In most cases the compiled list is not accessible to a reader (see, for example, Taagepera 1994; Syrjänen, Honkola, Korhonen, Lehtinen, Vesakoski, Wahlberg 2013¹). The paper Hofírková, Blažek 2012 is an exception as it gives the wordlists for many languages including Finnish, Estonian and Votic. However, the method of compilation of a wordlist as well as sources of the data are not always transparent there. For example, the list of sources does not contain any dictionary of the Votic language that leads us to conclude that secondary sources (such as etymological dictionaries) were used to obtain data. Many flaws both in transcription² and the choice of words³ increase this impression. Another piece of recent research that uses Swadesh lists is Tillinger 2014. Tillinger analyses Saami languages and, among other things, gives the Swadesh lists of several European languages including Finnish and Estonian. In the Appendix, we comment on the differences between Tillinger's and our Swadesh lists.

In the current article we present both the explicit wordlists and the transparent methodology of their compilation.

The article consists of four sections. Section 1 provides the basic information: (a) the main facts about the Votic and the Ingrian languages, (b) a description of data and methods of the research, (c) transcription conventions. Section 2 presents the annotated wordlists. In Section 3 (Discussion), we formulate our preliminary observations of the wordlists. Section 4 (Conclusions) contains a short summary of the results.

¹ When our article was already submitted to the journal it became known that the dataset used by Syrjänen, Honkola, Korhonen, Lehtinen, Vesakoski, Wahlberg (2013) is now open for online access, see <https://www.bedlan.net/data>. The lexical lists from this dataset did not try to solve the synonymy problem: several words are often given for the same definition.

² For example, *munö* 'egg' instead of *muna*, *polottaa* 'burn (tr)' instead of *põlõttaa*, *anna* 'give' instead of *antaa* (*anna* is the 2Sg imperative form; but the infinitive is given for other verbs in the list), *põlvi* 'knee' instead of *põlvi*, *tundõ* 'know' instead of *tunta*, *жуҕала* 'say' instead of *жуҕала*, *seemee* 'seed' instead of *seemene*, *kelten* 'yellow' instead of *keltein*. NB! Here we give Votic forms in the spelling for Kattila and the neighbouring varieties of Votic. Most publications on Votic including Hofírková, Blažek 2012 follow this system of spelling.

³ For example, 'green' is rather *rohoin* than *viher* as *viher* means 'unripe'; 'to lie' is *ležiä* but not *magata* as *magata* means 'to sleep'; *sato* is a rare and dialectally restricted word for 'precipitation' and a common word for 'rain' is *vihma*; the main word for 'this' is *kase* and *se* means 'this' or 'that' depending on a context (see comments to this item in the wordlist below); 'to hear' is *kuulla* but *kuullua* is 'to be heard; to listen to smb'.

1.1. Languages

Votic and Ingrian are minor Finnic languages on the verge of extinction. Votic belongs to the southern branch of the Finnic languages and is the closest relative of Estonian; Ingrian belongs to the northern branch and is the closest relative of Finnish and Karelian.

The last generation of Votic and Ingrian fluent speakers was born in the early 1930s. Their deportation to Finland during the Second World War, a ban on living in their native settlements after the war and the negative attitude of Russian people towards speakers of minority languages led to the rapid extinction of both Votic and Ingrian (see more details in Рожанский, Маркус 2013). At the moment, the most optimistic calculations give no more than five Votic and twenty Ingrian speakers (representing two dialects).

Most Votic dialects are extinct. Krevin — the language of the Votic population relocated to Latvia in the 15th century — died out by the middle of the 19th century. The last speaker of Eastern Votic died in 1976 (Ernits 2005 : 87), and the last recordings of Central Votic were made in the 1970s. Most probably there are no fluent speakers of the mixed Votic-Ingrian Kukkuzi variety (Suhonen 1985; Markus, Rozhanskiy 2012), though there were a few in the mid-2000s. The last speakers of Votic represent the Western dialect (Vaipoli Votic), which shows some contact-induced Ingrian influence (Rozhanskiy, Markus 2015).

The last speakers of Ingrian represent the Soikkola and Lower Luga dialects. Two other traditionally distinguished Ingrian dialects are already extinct. Oredeži Ingrian died out in the second half of the 20th century (Laanest (Лаанест 1993 : 62) considered it already moribund in the 1960s); Hevaha Ingrian became extinct around the turn of the millennium.

In the current research, we use Votic data from the Western dialect and Ingrian data from both the Soikkola and Lower Luga dialects. The analysis of two Ingrian dialects is not redundant but is one of the key goals of the research. According to the hypothesis formulated in Rozhanskiy, Markus 2014, the Lower Luga dialect (traditionally described as the most specific Ingrian dialect⁴) is in fact a very specific convergent variety based on Ingrian and Votic but also influenced by Ingrian Finnish and Estonian. Many Votes shifted completely to this variety and changed their identity.

The contact situation for the analysed minor languages can be briefly described as follows. Votic has had intensive contact with Russian during the last millennium. The Western Votic variety discussed in this article was influenced by Ingrian (in fact, all Votic villages in the Lower Luga area had a mixed Votic-Ingrian population in the 20th century, and there were plenty of mixed Votic-Ingrian families). Central Votic was in close contact with Ingrian Finnish; however, due to the difference in religion, mixed marriages were not typical.

Ingrians also had contact with the Russian population but it seems that Soikkola Ingrian was less influenced by Russian than Votic. Also, there are no evident traces of Votic or some other Finnic influence on Soikkola Ingrian.

⁴ See, for example, Лаанест 1966 : 146, 161: "As we can see the largest number of differences is between the Lower Luga dialect and three other dialects", "At present, the problem of origin of the Lower Luga dialect cannot be finally solved".

On the contrary, Lower Luga Ingrian had intensive contact not only with Russian but also with Ingrian Finnish, Votic, and (in the southern part of the area) with Estonian.

As most of the Finnic population from Western Ingria was deported to Finland during the Second World War, most of the speakers had some experience in the Finnish language.

1.2. Data and methods

The last decades have witnessed a renewal of interest in lexicostatistics and glottochronology. Different scholars use different mathematic algorithms: some work with the classical Swadesh method or its modified versions (Hofírková, Blažek 2012), some use methods borrowed from evolutionary biology, such as maximum parsimony or Bayesian phylogenetic inference (Chang, Cathcart, Hall, Garrett 2015; Honkola 2016). All these studies have one thing in common: they use lists of basic lexemes with fixed meanings. Different authors use different wordlists, for example, the original Swadesh 200-word list (Swadesh 1952), the Swadesh 100-word list (Swadesh 1971) and various modifications thereof (Kassian, Starostin, Dybo, Chernov 2010), the ASJP 40-word list (Holman, Wichmann, Brown, Velupillai, Müller, Bakker 2008), the Leipzig-Jakarta list (Tadmor 2009), and others. A useful catalogue of such lists can be found on the Concepticon site (<http://concepticon.clld.org>).

We are convinced that the key problem with lexicostatistics lies not so much in the mathematics, as in the lexicography. Whatever algorithm we choose to apply, if our initial data are not sufficiently accurate, the well-known maxim "garbage in — garbage out" will aptly describe the result. The core problem while compiling the wordlists is synonymy. For every meaning on the list in every language included in the comparison we must use the most neutral basic word representing this meaning. However, the standard meanings in the various lists of basic vocabulary are usually represented by English words (or words of some other natural language). It is clear that English words need not have one-to-one semantic equivalents in other human languages. For example, English 'hand' may be translated into Russian as 'рука' or 'кисть', depending on the context. So, the compilation of a reliable basic vocabulary list requires a semantic specification of items on this list. Such a specification can be done in several ways:

- (1) using more than one language for the list of basic meanings counting on more specific meanings in at least one of the languages;
- (2) giving additional comments and explanations to narrow the meaning of the basic word;
- (3) choosing a specific context to narrow down the meaning of an item on the list.

We chose the Swadesh wordlist and its particular modification because it is one of a few, if not the only, basic lexicon list for which a detailed semantic specification is available (Kassian, Starostin, Dybo, Chernov 2010). This standard was already used in hundreds of wordlists compiled for the Global Lexicostatistical Database project (GLD 2018), as well as in publications not affiliated with this project (Грунтов, Мазо 2015).

The selected standard allows us to use all the mentioned methods of resolving the choice between synonymic variants. First, the list of basic

words is given in both English and Russian. Second, the comments specifying the meaning of the basic words are given. Third, every basic word has several contexts that narrow down the meaning. Thus, it becomes possible with minimal exceptions to choose the most neutral word that is not too general or too specific, is not stylistically marked, and is not too bookish or too colloquial.

In this article, we present the 111-word modified Swadesh lists for five Finnic idioms, compiled on the basis of the following methodology. The compilation of lists had two stages. During the first (preliminary) stage, the lists were compiled with the help of dictionaries⁵ and/or the authors' competence. During this stage, some items had several variants in case there were no evident reasons to select the most suitable word. During the second stage, the lists were checked with the help of native speakers (see Acknowledgements section) and (for the minor languages) corpora of elicitation and narratives were also used.⁶ The native speakers annotated the meaning and usage of words and translated the sentences with the contexts. The final decision of which word should be added to the list was made exclusively by F. Rozhanskiy.

Etymological comments are based on standard etymological dictionaries (SSA; EES; UEW; LÄGLOS) and other sources on Finnic and Uralic etymology. The final decisions on etymology were made exclusively by M. Zhivlov.

1.3. Transcription conventions

Two of the analysed languages, Estonian and Finnish, have a literary tradition; Ingrian had a literary tradition only for a short period in the 1930s; and Votic has always been an unwritten language.⁷ In this paper, we use the following transcription conventions.

For *E s t o n i a n* and *F i n n i s h*, the standard orthography is used.

Our *V o t i c* transcription is similar to the one used by Tsvetkov (1995) but has some minor differences. First, we use *j* instead of the traditional Finnic *i* as the second part of diphthongs, e.g. *kejg* 'all' not *keig* (see the discussion in Маркус, Рожанский 2017 : 351–352). Second, the final reduced vowels are spelled as *ɔ̂* (back vowel) and *ə* (front vowel) instead of *ɛ* and *e* respectively, e.g. *rintɔ̂* 'breast', *tšülmə* 'cold'. The long vowels are transcribed with double letters for comparability with other languages.

The *S o i k k o l a* *I n g r i a n* transcription is close to Nirvi 1971 but the short geminates are transcribed with double letters and a breve in all

⁵ The following dictionaries were used: Tsvetkov 1995 and VKS 2013 for Votic, Nirvi 1971 for Ingrian, EVS for Estonian, БФРС and НБРФС for Finnish.

⁶ Our corpora were collected during fieldtrips organized by Fedor Rozhanskiy and Elena Markus in 2003–2018. The Soikkola Ingrian corpus contains about 650 hours of recordings; the Votic and Lower Luga Ingrian corpora contain about 250 hours of recordings each.

⁷ We mean that Votic has never been taught in school or had a written standard that was regularly used by native speakers to communicate and to read printed materials. However, besides various texts transcribed by linguists as speech samples there were a number of texts in Votic published for native speakers or other people studying Votic (e.g. Муслимов, Кузнецова, Николаева, Гореликов, Ефимов, Ефимова 2003; Heinsoo 2015; 2018).

phonetic contexts, e.g. *valkkia* 'white', not *valkia*. The long mid-high vowels of the first syllable (that in some idiolects merged with the long high vowels *uu*, *üü*, *ii*) are marked with a circumflex accent below: *kõõri* 'bark', *õõ* 'night', *ššemen* 'seed'. The sibilant fricatives are *š* and *ž* (instead of *s* and *z*), e.g. *šuur* 'big', *męžž* 'man'.

There are no authoritative sources for the transcription of Lower Luga Ingrian, which also exhibits very significant phonetic variation between different varieties. We represent long mid vowels in the first syllable as diphthongs, e.g. *kuori* 'bark', *üö* 'night', *siemen* 'seed', although their diphthongization is usually much weaker than in Finnish. The final reduced vowels that can be realized as short, voiceless, or dropped are marked with small capital letters: *savvu* 'smoke' [*savvu* ~ *savvu* ~ *savv*], *hantÄ* 'tail' [*hantä* ~ *hantä* ~ *hant*].

Proto-Finnic and Proto-Uralic reconstructions are written in a system based on the UPA. In this system affricates are written as single symbols.

2. The wordlists

1. all [все]	Est. <i>kõik</i>	Vot. <i>keĵg</i>	L-L. <i>kai</i>
	Soi. <i>kaig</i>	Fin. <i>kaikki</i>	

This word exists in most Finnic languages. It goes back to Proto-Finnic **kaikki* 'all', possibly of Baltic origin (cf. SSA 1 : 275; EES 199).

2. ashes [зола]	Est. <i>tuhk</i>	Vot. <i>tuhkš</i>	L-L. <i>tuhka</i>
	Soi. <i>tuhka</i>	Fin. <i>tuhka</i>	

This word exists in most Finnic languages. It goes back to Proto-Finnic **tuhka* 'ashes', borrowed from Germanic (cf. SSA 3 : 319; LÄGLOS III : 307).

3. bark [кора]	Est. <i>koor</i>	Vot. <i>koori</i>	L-L. <i>kuori</i>
	Soi. <i>kõõri</i>	Fin. <i>kaarna</i>	

Proto-Finnic **koori* 'bark' goes back to Proto-Uralic **kari* 'surface, crust, skin, bark' (Aikio 2015 : 52), which is certainly not the main word for 'bark' in Proto-Uralic (the meaning 'bark' is represented only in Finnic). Fin. *kaarna* exists in some Finnic languages and possibly has a Baltic origin (cf. SKES 1987 : 135; SSA 1 : 265–266). The word *kuori* also exists in Finnish but has a more general meaning, and the word *kaarna* looks more natural in the test contexts. The word *kaarna* is known in Ingrian with the meaning 'cork; fir or pine bark' (Nirvi 1971 : 148).

4. belly [живот]	Est. <i>kõht</i>	Vot. <i>vattsš</i>	L-L. <i>vatsA</i>
	Soi. <i>vatsa</i>	Fin. <i>vatsa</i>	

Votic, Ingrian and Finnish preserve Proto-Finnic **vacca* 'belly'. Pace Rédei (UEW 547), this word has no acceptable etymology: the proposed Mansi cognate has irregular vocalism and is restricted to North Mansi. Estonian *kõht* goes back to Proto-Finnic **koktu* 'belly' (cf. EES 199); *vats* exists as a dialectal variant. In colloquial Finnish, the word *maha* looks more natural in the test sentences. The difference between **vacca* and **koktu* in Proto-Finnic may have been that of '(external) belly' vs '(internal) belly/stomach'.

5. big, large [большой]	Est. <i>suur</i>	Vot. <i>suur(i)</i>	L-L. <i>suur</i>
	Soi. <i>šuur</i>	Fin. <i>iso</i>	

This word exists in most Finnic languages and goes back to Proto-Finnic **suuri* 'big', borrowed from Germanic (cf. SSA 3 : 224–225; EES 491;

LÄGLOS III 253–254). In Finnish, the word *suuri* which is semantically close also exists. However, our Finnish consultant considered *iso* to be the main word. The Finnish word may be an archaism, replaced in other Finnic languages by a Germanic loanword. Proto-Finnic **iso* 'big', derived from **isä* 'father', has a striking parallel in Moksha *oću* 'big', derived from *oća* 'paternal uncle' (cf. SSA 1 : 228; UEW 78), cf. also Finnish *eno* '(maternal) uncle' and *enemmän* 'more'.

6. bird [птица] Est. *lind* Vot. *lintu* L-L. *lintu*
Soi. *lindu* Fin. *lintu*

Proto-Finnic **lintu* 'bird' is either an isolated word or an irregular reflex of Proto-Uralic **lunta* 'bird, goose' (cf. SSA 2 : 80; EES 242; UEW 254). Livonian and dialectal Estonian data show that Proto-Finnic **lintu* was originally polysemous 'bird, flying insect, wild animal'. The polysemous word 'bird / wild animal' is found also in Samoyed and Ob-Ugric, although Finnic, Ob-Ugric, and Samoyed words with these meanings are not related.

7. to bite [кусать] Est. *hammustada* [*hammustama*] Vot. *purra* L-L. *purra*
Soi. *purra* Fin. *purra*

Proto-Finnic **purę-* 'to bite' goes back to Proto-Uralic **puri-* 'to gnaw, bite' (cf. SSA 2 : 438; EES 393; UEW 405–406). The verb *pureda* exists in Estonian but *hammustada* (derived from *hammas* 'tooth') is considered a more neutral word.

8. black [черный] Est. *must* Vot. *mussâ* L-L. *musta*
Soi. *mušta* Fin. *musta*

Proto-Finnic **musta* 'black' has no plausible etymology (cf. SSA 2 : 183; EES 289; LÄGLOS II 276).

9. blood [кровь] Est. *veri* Vot. *veri* L-L. *veri*
Soi. *veri* Fin. *veri*

Proto-Finnic **veri* 'blood' goes back to Proto-Uralic **weri* 'blood' (cf. SSA 3 : 427; EES 598–599; UEW 576).

10. bone [кость] Est. *luu* Vot. *luu* L-L. *luu*
Soi. *luu* Fin. *luu*

Proto-Finnic **luu* 'bone' goes back to Proto-Uralic **l̥wi* 'bone' (cf. SSA 2 : 114; EES 256; UEW 254–255). In Estonian, there is also a word *kont* (of Finnic origin, SSA 1 : 398; EES 175) that possibly broadened its meaning from 'shin' to 'bone'. This word was considered as more colloquial and less neutral.

11. breast [грудь] Est. *rind* Vot. *rintâ* L-L. *rinta*
Soi. *rinda* Fin. *rinta*

There are several hypotheses for the origin of Proto-Finnic **rinta* 'breast' (cf. SSA 3 : 80; EES 429). It is improbable that it is a borrowing from Germanic (LÄGLOS III 158–159). Koivulehto (2008 : 315–317) has suggested a Slavic origin. Proto-Saami **rentē* 'breast' is a Finnic loanword. In Finnish, there is also a word *povi* of Uralic origin (cf. SSA 2 : 408; UEW 395) that can be used at least for the second context (*His breast (chest) was decorated with ornaments*). However it is rarely used and should not be considered the main word. There is no special word for 'woman's breast' but there is a Finnic word for 'teat' that in some idioms has an extended meaning 'woman's breast': Est. *nänn*, Vot. *nännə*, L-L. *nännä*, Soi. *nännä*, Fin. *nänni*. This word came from child language but it is probably rather old (SSA 2 : 252).

12. to burn (trans.) [жечь] Est. *põletada* [*põletama*] Vot. *pęłętta* L-L. *poltta*
 Soi. *polttaa* Fin. *polttaa*

Proto-Finnic **poltta-* 'to burn (trans.)' is an irregular causative derivative from Proto-Finnic **pala-* 'to burn (intrans.)' (Est. *põleda*, Vot. *pęłęssā*, L-L. *palla*, Soi. *pallaa*, Fin. *palaa*) (cf. SSA 2 : 392; EES 399). This pair of verbs goes back to Proto West Uralic **pala-* 'to burn (intrans.)' ~ **poltta-* 'to burn (trans.)' (cf. UEW 352). In Estonian and Votic the reflex of **poltta-* was replaced by a more regular causative from the same root.

13. cloud [облако] Est. *pilv* Vot. *pilvi* L-L. *pilvi*
 Soi. *pilvi* Fin. *pilvi*

Proto-Finnic **pilvi* 'cloud' goes back to Proto-Uralic **pilwi* 'cloud' (cf. SSA 2 : 367; EES 370; UEW 381).

14. cold [холодный] Est. *külm* Vot. *tšülmā* L-L. *külmä*
 Soi. *külmä* Fin. *kylmä*

Proto-Finnic **külmä* 'cold' goes back to Proto-Uralic **külmä* 'cold', attested in Finnic, Saami, Mordvin, Mari and Permic (cf. UEW 663). The wide distribution of this word and completely regular sound correspondences make the hypothesis of its borrowing from Baltic (Koivulehto 1983; SSA 1 : 462; EES 213) quite improbable.

15. to come [приходить] Est. *tulla* [*tulema*] Vot. *tullā* L-L. *tulla*
 Soi. *tulla* Fin. *tulla*

Proto-Finnic **tule-* 'to come' goes back to Proto-Uralic **tuli-* 'to come' (cf. SSA 3 : 324; EES 552—553; UEW 535).

16. to die [умирать] Est. *surra* [*surema*] Vot. *koollā* L-L. *kuolla*
 Soi. *koolla* Fin. *kuolla*

Proto-Finnic **koolē-* 'to die' goes back to Proto-Uralic **kali-* 'to die' (cf. SSA 1 : 440; UEW 173). In Estonian, it is observed only in dialects (EES 176). Estonian *surra* goes back to Proto-Finnic **surē-* < Proto-Uralic **šuri-* 'to die' (cf. EES 489; UEW 489) — certainly not the main synonym for this meaning in Proto-Uralic.

17. dog [собака] Est. *koer* Vot. *kojrā* L-L. *koira*
 Soi. *koira* Fin. *koira*

Proto-Finnic **koira* 'dog' goes back to Proto-Uralic **kojra* 'male' (cf. SSA 1 : 385; EES 168; UEW 168—169). The meaning 'male' is preserved in the Finnic derivative **koiras*. The original Finnic word for 'dog' was rather Proto-Finnic **peni* 'dog' (< Proto-Uralic **peni* 'dog'), replaced as the main word for this meaning everywhere except Livonian and South Estonian (cf. SSA 2 : 335—336; EES 361; UEW 371).

18. to drink [пить] Est. *juua* [*jooma*] Vot. *juuvvā* L-L. *juovva*
 Soi. *jõvvva* Fin. *juoda*

Proto-Finnic **joo-* 'to drink' goes back to Proto-Uralic **jixi-* 'to drink' (cf. SSA 1 : 249; EES 98; UEW 103).

19. dry [сухой] Est. *kuiv* Vot. *kujvā* L-L. *kuiva*
 Soi. *kuiva* Fin. *kuiva*

Proto-Finnic **kuiva* 'dry' lacks an acceptable etymology (cf. SSA 1 : 426; EES 187). The hypothesis of a Germanic origin is implausible (LÄGLOS II 114), and the comparison with Proto-Khanty **kujəm-* 'to fall, sink (of water)' is dubious (cf. UEW 196—197).

20. ear [ухо] Est. *kõrv* Vot. *kęrvā* L-L. *korva*
 Soi. *korva* Fin. *korva*

Proto-Finnic **korva* 'ear' is cognate with Proto-Saami **koarvē* 'oarlock'. Further etymological connections of this word are unclear (cf. SSA 1 : 408; EES 202–203; UEW 187–188). It is a replacement of Proto-Uralic **peljä* 'ear' (cf. UEW 370).

21. earth [земля]	Est. <i>muld</i>	Vot. <i>maa</i>	L-L. <i>maa</i>
	Soi. <i>maa</i>	Fin. <i>maa</i>	

Proto-Finnic **maa* 'earth' (cf. SSA 2 : 133; EES 268) goes back to Proto-Uralic **m̄xi* 'earth'. In Estonian, earth as a physical substance (i.e. earth vs sand, handful of earth, etc. (see Kassian, Starostin, Dybo, Chernov 2010)) is expressed by the word *muld*, which is a Germanic borrowing (EES 286; LÄGLOS II 270). In other idioms the word *multa* also exists but is more peripheral than in Estonian. However, there are deviations. For example, in Finnish, the test sentence "I don't know whether that site contains sand or earth" requires the word *multa*, since *maa* is a general term for both 'sand' and 'earth'.

22. to eat [есть]	Est. <i>süüa</i> [sööma]	Vot. <i>süüvvä</i>	L-L. <i>süüvvä</i>
	Soi. <i>šöövä</i>	Fin. <i>syödä</i>	

Proto-Finnic **söö-* 'to eat' goes back to Proto-Uralic **sewi-* 'to eat' (cf. SSA 3 : 235; EES 500–501; UEW 440).

23. egg [яйцо]	Est. <i>muna</i>	Vot. <i>muna</i>	L-L. <i>muna</i>
	Soi. <i>muna</i>	Fin. <i>muna</i>	

Proto-Finnic **muna* 'egg' goes back to Proto-Uralic **muna* 'egg' (cf. SSA 2 : 178; EES 287; UEW 285–286).

24. eye [глаз]	Est. <i>sil̄m</i>	Vot. <i>sil̄m̄</i>	L-L. <i>sil̄mä</i>
	Soi. <i>šilmä</i>	Fin. <i>sil̄mä</i>	

Proto-Finnic **sil̄mä* 'eye' goes back to Proto-Uralic **šilmä* 'eye' (cf. SSA 3 : 181; EES 472–473; UEW 479).

25. fat [жир]	Est. <i>rasv</i>	Vot. <i>razv̄</i>	L-L. <i>razva</i>
	Soi. <i>ražva</i>	Fin. <i>rasva</i>	

Proto-Finnic **rasva* 'fat' is possibly an early Germanic borrowing (SSA 3 : 53; EES 420; LÄGLOS III 132). The word replaces Proto-Uralic **waji* 'fat', whose Finnic reflex **voi* means 'butter' (cf. UEW 578–579). Cf. also Proto-Uralic **koja* 'fat, tallow', whose Finnic reflex **kuu* 'tallow' is preserved only in Finnish and Karelian (cf. UEW 195–196).

26. feather [перо]	Est. <i>sulg</i>	Vot. <i>sul̄k̄</i>	L-L. <i>sulka</i>
	Soi. <i>šulga</i>	Fin. <i>sulka</i>	

Proto-Finnic **sulka* 'feather' is possibly an irregular reflex of Proto-Uralic **tulka* 'feather' (cf. SSA 3 : 211; EES 487; UEW 535–536). Livonian *türḡz* 'feather' may be another irregular reflex of the same Uralic word. As suggested by Kirill Reshetnikov (p.c.), the Livonian word may instead be cognate with Finnish *turkki* 'fur, hair (of animals)', Ingrian *turkki* 'chicken feather'. However, we would expect Livonian *-rk-* as a reflex of Proto-Finnic **-rkk-*, so the etymology of Livonian *türḡz* remains moot.

27. fire [огонь]	Est. <i>tuli</i>	Vot. <i>tuli</i>	L-L. <i>tuli</i>
	Soi. <i>tuli</i>	Fin. <i>tuli</i>	

Proto-Finnic **tuli* 'fire' goes back to Proto-Uralic **tuli* 'fire' (cf. SSA 3 : 211; EES 553; UEW 535).

28. fish [рыба]	Est. <i>kala</i>	Vot. <i>kala</i>	L-L. <i>kala</i>
	Soi. <i>kala</i>	Fin. <i>kala</i>	

Proto-Finnic **kala* 'fish' goes back to Proto-Uralic **kala* 'fish' (SSA 1 : 282; EES 120; UEW 119).

29. to fly [летать] Est. *lennata* [*lendama*] Vot. *lentä* L-L. *lenttä*
 Soi. *lenittä* Fin. *lentää*

Proto-Finnic **lentä-* 'to fly' has no plausible etymology (cf. SSA 2 : 64; EES 236).

30. foot [нога] Est. *jalg* Vot. *jalčkä* L-L. *jalka*
 Soi. *jalga* Fin. *jalka*

Proto-Finnic **jalka* 'foot' goes back to Proto-Uralic **jalka* or **jilka* 'foot' (cf. SSA 1 : 234; EES 96–97; UEW 88–89).

31. full [полный] Est. *täis* Vot. *täinə* L-L. *täin*
 Soi. *täin* Fin. *täysi/täynnä*

Proto-Finnic **täüci* 'full' goes back to Proto-Uralic **täüdi* 'full' (cf. SSA 3 : 358; EES 566; UEW 518). Koivulehto's hypothesis of a Germanic borrowing does not withstand scrutiny (LÄGLOS III 331–332; Aikio 2002 : 31–34). In some languages, the lexicalized essive form **täün-nä* is used in the test contexts.

32. to give [давать] Est. *anda* [*andma*] Vot. *anta* L-L. *anta*
 Soi. *antta* Fin. *antaa*

Proto-Finnic **anta-* 'to give' goes back to Proto-Uralic **amta-* or **jmta-* 'to give' (cf. SSA 1 : 77; EES 50; UEW 8) — one of two or three Proto-Uralic verbs of giving.

33. to go [идти] Est. *minna* [*minema*] Vot. *mennə* L-L. *männä*
 Soi. *männä* Fin. *mennä*

Proto-Finnic **mene-* 'to go' goes back to Proto-Uralic **meni-* 'to go' (cf. SSA 2 : 159; EES 282; UEW 272). In Estonian, this word is combined in one suppletive paradigm with the verb *lähe-* < Proto-Finnic **läkte-* < Proto-Uralic **läkti-* 'to go out, to go away' (cf. EES 262; UEW 239–240).

34. good [хороший] Est. *hea* Vot. *üvä* L-L. *hüvä*
 Soi. *hüvä* Fin. *hyvä*

Proto-Finnic **hüvä* 'good' has cognates in Saami and Mordvin languages (cf. SSA 1 : 201; EES 86; UEW 499). In Saami the word means 'to heal (of wound)', the Mordvin word means 'good', but it is not the main synonym for 'good' in Mordvin. Comparisons with words in other Uralic languages are hypothetical. The Estonian word is somewhat aberrant phonetically; still, it is cognate with words in other Finnic languages. The non-aberrant form *hüva* exists in Estonian dialects and in colloquial speech.

35. green [зеленый] Est. *roheline* Vot. *rohojn* L-L. *rohoin*
 Soi. *rohhojn* Fin. *vihreä*

Finnish preserves the reflex of Proto-Finnic **viherä* 'green'. Together with a morphological variant **vihanta*, this word goes back to a late dialectal Uralic protoform **wiša* 'green; poison', borrowed from Indo-Iranian (cf. SSA 3 : 438; UEW 823–824). In four idioms, the adjective 'green' is derived from Proto-Finnic **rooho* 'grass', possibly of Germanic origin (EES 432–433; LÄGLOS III 180).

36. hair [волос] Est. *juus* Vot. *ivuz* L-L. *hiuz*
 Soi. *hiuž* Fin. *hius*

Proto-Finnic **hiβus* 'hair' is a derivative based on a root borrowed from Germanic (cf. SSA 1 : 168; EES 102; LÄGLOS I 107–108). This word is a replacement of Proto-Uralic **ipti* 'hair' (cf. cf. UEW 14–15).

37. hand [рука] Est. *käsi* Vot. *tšäsi* L-L. *käsi*
 Soi. *käži* Fin. *käsi*

Proto-Finnic **käci* 'hand' goes back to Proto-Uralic **käti* 'hand' (cf. SSA 1 : 479; EES 209; UEW 140).

38. head [голова]	Est. <i>pea</i>	Vot. <i>pää</i>	L-L. <i>pää</i>
	Soi. <i>pää</i>	Fin. <i>pää</i>	

Proto-Finnic **pää* 'head' goes back to Proto-Uralic **päñi* 'head' — one of the two Uralic words for 'head' (cf. SSA 2 : 462; EES 357; UEW 365—366).

39. to hear [слышать]	Est. <i>kuulda</i> [<i>kuulma</i>]	Vot. <i>kuullā</i>	L-L. <i>kuulla</i>
	Soi. <i>kuulla</i>	Fin. <i>kuulla</i>	

Proto-Finnic **kuulge-* 'to hear' goes back to Proto-Uralic **kuwli-* 'to hear' (cf. SSA 1 : 456; EES 197; UEW 197—198).

40. heart [сердце]	Est. <i>süda</i>	Vot. <i>süä</i>	L-L. <i>süän</i>
	Soi. <i>šüän</i>	Fin. <i>sydän</i>	

Proto-Finnic **südän* 'heart' goes back to Proto-Uralic **śäd'am* 'heart' (cf. SSA 3 : 228; EES 501; UEW 477).

41. horn [рог]	Est. <i>sarv</i>	Vot. <i>sarvi</i>	L-L. <i>sarvi</i>
	Soi. <i>šarvi</i>	Fin. <i>sarvi</i>	

Proto-Finnic **sarvi* 'horn' goes back to the dialectal Uralic protoform **śarwi* 'horn', borrowed from Indo-Iranian (cf. SSA 3 : 159; EES 461—462; UEW 486—487).

42. I [я]	Est. <i>mina</i> (~ <i>ma</i>)	Vot. <i>miä</i>	L-L. <i>miä</i>
	Soi. <i>miä</i>	Fin. <i>minä</i> (~ <i>mä</i>)	

Proto-Finnic **minä* 'I' goes back to Proto-Uralic **min* 'I' (cf. SSA 2 : 168; EES 281—282; UEW 294). In Estonian and Finnish, there is variation between a long and a short form.

43. to kill [убивать]	Est. <i>tappa</i> [<i>tapma</i>]	Vot. <i>tappa</i>	L-L. <i>tappa</i>
	Soi. <i>tappaa</i>	Fin. <i>tappaa</i>	

Proto-Finnic **tappa-* 'to kill' goes back to Proto West Uralic **tappa-*, whose reflex in Mordvin languages means 'to break' (cf. SSA 3 : 269—270; EES 514—515; UEW 509—510).

44. knee [колени]	Est. <i>põlv</i>	Vot. <i>pɛlvi</i>	L-L. <i>polvi</i>
	Soi. <i>polvi</i>	Fin. <i>polvi</i>	

Proto-Finnic **polvi* 'knee' goes back to the Proto-Uralic word for 'knee', whose exact reconstruction is doubtful. Apparently it was a compound of two roots: **pu.xi* or **pu.wi* 'knee' (> Proto-Samoyed **puə* 'knee') and **l̥wi* 'bone' (cf. SSA 2 : 393; EES 400; UEW 393).

45. to know [знать]	Est. <i>teada</i> [<i>teadma</i>]	Vot. <i>täätä</i>	L-L. <i>tiitä</i>
	Soi. <i>tiittää</i>	Fin. <i>tietää</i>	

Proto-Finnic **teetä-* 'to know' is derived from **tee* 'road, path' (SSA 3: 289). The hypothesis of a Germanic origin (EES 519) is unacceptable (LÄGLOS III 292—293). This word replaces Proto-Uralic **tumti-* 'to know', whose Finnish reflex **tunte-* means rather 'to feel; to recognize' (cf. SSA 3 : 327; UEW 536—537).

46. leaf [лист]	Est. <i>leht</i>	Vot. <i>lehto</i>	L-L. <i>lehti</i>
	Soi. <i>lehti</i>	Fin. <i>lehti</i>	

Proto-Finnic **lehti* 'leaf' goes back to late dialectal Uralic (West Uralic and Mari) **lešti* 'leaf', apparently of Balto-Slavic origin (cf. SSA 2 : 58—59; EES 234; UEW 689).

47. to lie [лежать]	Est. <i>lamada</i> [<i>lamama</i>]	Vot. <i>ležiä</i>	L-L. <i>ležže</i>
	Soi. <i>leššä</i>	Fin. <i>maata</i>	

The Finnic languages usually use the verb 'to be' to denote the position of an object and do not express the difference between 'to lie' and 'to stand'. Therefore, the Proto-Finnic word for 'to lie' cannot be convincingly reconstructed. The Estonian word is derived from the Proto-Finnic noun/adjective **lama* 'lying', borrowed from Germanic (cf. SSA 2 : 42; EES 225–226; LAGLOS II 165), Votic and Ingrian borrowed this verb from Russian, and Finnish uses the reflex of Proto-Finnic **maka-* 'to sleep' (q.v.). In Estonian, there is another word for 'to lie', *lebada* [*lebama*], that is less general than *lamada* [*lamama*]. It goes back to Proto-Finnic **lepä-*, for which two mutually contradictory and phonetically problematic Germanic etymologies were proposed (cf. SSA 2 : 67–68; EES 232; LÄGLOS II 198–199). Proto-Uralic root **kuji-* 'to lie' has no reflexes in West Uralic (cf. UEW 197). Votic verbs *lammoa* and *lamota* 'lie about, to rest lying' are peripheral: they are very restricted dialectally (VKS 574) and are not known to contemporary speakers.

48. liver [печень] Est. *maks* Vot. *mahsā* L-L. *maksA*
 Soi. *leibä-liha* ~ *petšonka* Fin. *maksa*

Proto-Finnic **maksA* 'liver' goes back to Proto-Uralic **miksa* 'liver' (cf. SSA 2 : 142; EES 273; UEW 264). This is one of the most stable words in the Uralic basic lexicon. However in Soikkola Ingrian the word *maksa* means 'fish liver' or (in plural) 'internal apparatus'. The meaning 'liver' is expressed either by a descriptive compound *leibä-liha* (literally: bread meat) or by a Russian borrowing.

49. long [длинный] Est. *pikk* Vot. *pittšə* ~ *pittši* L-L. *pitkÄ*
 Soi. *pitkä* Fin. *pitkä*

Proto-Finnic **pitkä* 'long' goes back to Proto-Uralic **pid-kä* 'long', from the root **pidi* (cf. SSA 2 : 377; EES 368; UEW 377–378).

50. louse [вошь] Est. *täi* Vot. *täj* L-L. *täi*
 Soi. *täi* Fin. *täi*

Proto-Finnic **täi* 'louse' goes back to Proto-Uralic **täji* 'louse' (cf. SSA 3 : 353; EES 565; UEW 515).

51. man (male) [мужчина] Est. *mees* Vot. *meez* L-L. *mies*
 Soi. *męgž* Fin. *mies*

Proto-Finnic **mees* 'man' has no acceptable etymology (SSA 2 : 166). The hypothesis of a Germanic origin is not likely (EES 279; LÄGLOS II 263). The shape CVVC is anomalous from the point of view of Finnic phonotactics.

52. man (person) [человек] Est. *inimene* Vot. *inimin* ~ *inemin* L-L. *ihmin*
 Soi. *ihmiin* ~ *ilmihin* Fin. *ihminen*

The phonetic reconstruction of Proto-Finnic **inehminen* 'person' is tentative (cf. SSA 1 : 221; EES 92–93). Forms like Finnish *ihminen* are probably due to contamination with Proto-Finnic **imeh* 'miracle'. The word has no acceptable etymology; attempts to derive it from various Indo-European sources are unconvincing. At the same time, comparison with the Mordvin word for 'guest' (UEW 627–628) faces multiple irregularities. In Soikkola Ingrian, there are variants of this word; the choice depends on the particular idiolect.

53. many, a lot of [много] Est. *palju* Vot. *pal'lo* L-L. *paljo*
 Soi. *paljo* Fin. *paljon* ~ *monta*

The etymology of Proto-Finnic **paljo* 'many' remains disputed (cf. SSA 2 : 301; EES 350). Potential Uralic comparisons are dubious (cf. UEW 350–

351). The Germanic origin is not accepted in LÄGLOS III 22. Saarikivi (2009 : 146–147) suggests a Slavic etymology. In Finnish there is also a word *monta* (partitive of *moni*), that may be viewed as an archaism. It goes back to Proto-Finnic **moni* ‘many’ < Proto-Uralic **moni*, the reflex of which is preserved also in Permic. The Germanic origin of this word cannot be accepted (LÄGLOS II 265–266). It is not clear which word is more general in Finnish (both words sound good in the test sentences). In Estonian, Votic and Ingrian, the reflex of **moni* either has a different meaning or is not the main word for ‘many’.

54. meat [мясо]	Est. <i>liha</i>	Vot. <i>liha</i>	L-L. <i>liha</i>
	Soi. <i>liha</i>	Fin. <i>liha</i>	

Proto-Finnic **osa* ‘meat’, cognate with Proto-Saami **oańćē* ‘meat’, is preserved only in Livonian. In other languages, this word is replaced by Proto-Finnic **liha* (cf. SSA 2 : 72; EES 238–239), whose Livonian reflex preserved the original meaning ‘body; (human) flesh’.

55. moon [луна]	Est. <i>kuu</i>	Vot. <i>kuu</i>	L-L. <i>kuu</i>
	Soi. <i>kuu</i>	Fin. <i>kuu</i>	

Proto-Finnic **kuu* ‘moon’ goes back to Proto-Uralic **kjuwi* or **kiji* ‘moon’ (cf. SSA 1 : 455–456; EES 196–197; UEW 211–212).

56. mountain [гора]	Est. <i>mägi</i>	Vot. <i>mätši</i>	L-L. <i>mäki</i>
	Soi. <i>mägi</i>	Fin. <i>vuori</i>	

Proto-Finnic **voori* ‘mountain’, going back to Proto-Uralic **wari* ‘hill, mountain’ (cf. SSA 3 : 475; UEW 571), is preserved only in Finnish, where it is opposed to *mäki* ‘hill’. Other languages have lost the inherited word for ‘mountain’ and replaced it with the word for ‘hill’. Proto-Finnic **mäki* ‘hill’ goes back to Proto-Uralic **mäki*, also preserved in Khanty, where its reflex means ‘tussock’ (cf. SSA 2 : 191; EES 294; UEW 266).

57. mouth [рот]	Est. <i>suu</i>	Vot. <i>suu</i>	L-L. <i>suu</i>
	Soi. <i>šuu</i>	Fin. <i>suu</i>	

Proto-Finnic **suu* ‘mouth’ goes back to Proto-Uralic **šuwi* ‘throat, mouth’ (cf. SSA 3 : 223–224; EES 491; UEW 492–493).

58. nail [ноготь]	Est. <i>küüs</i>	Vot. <i>tšüinsi</i>	L-L. <i>künsi</i>
	Soi. <i>künž</i>	Fin. <i>kynsi</i>	

Proto-Finnic **künci* ‘claw, nail’ goes back to Proto-Uralic **küńci* ‘claw, nail’ (cf. SSA 1 : 464; EES 216; UEW 157).

59. name [имя]	Est. <i>nimi</i>	Vot. <i>nimi</i> ~ <i>imi</i>	L-L. <i>nimi</i>
	Soi. <i>imi</i>	Fin. <i>nimi</i>	

Proto-Finnic **nimi* ‘name’ goes back to Proto-Uralic **nimi* ‘name’ (cf. SSA 2 : 222; EES 313; UEW 305). Votic and Ingrian show variation between *nimi* and the variant *imi*, whose origin is not obvious (possibly it results from a contamination of *nimi* and Russian *имя* ‘name’). In Soikkola Ingrian, the variant *imi* is the most prevalent form; in Luuditsa Votic both variants are used; for Lower Luga Ingrian the variant *nimi* looks more typical.

60. neck [шея]	Est. <i>kael</i>	Vot. <i>kaglađ</i>	L-L. <i>kagla</i>
	Soi. <i>kagla</i>	Fin. <i>kaula</i>	

Proto-Finnic **kakla* ‘neck’ is borrowed from Baltic (SSA 1 : 331; EES 113). This word replaces Proto-Uralic **šepä* ‘neck, collar’, preserved in Finnic with the meanings ‘collar, front part of sledge, etc.’ (cf. SSA 3 : 169–170; UEW 473–474).

61. new [НОВЫЙ] Est. *uus* Vot. *uus(i)* L-L. *uusi*
 Soi. *uuž* Fin. *uusi*

Proto-Finnic **uuci* 'new' goes back to Proto-Uralic **wuδi* 'new' (cf. SSA 3 : 381; EES 581; UEW 587).

62. night [ночь] Est. *öö* Vot. *üü* L-L. *üö*
 Soi. *öǰ* Fin. *yö*

Proto-Finnic **öö* 'night' goes back to Proto-Uralic **üji* or **ejj* 'night' (cf. SSA 3 : 493; EES 633; UEW 72).

63. nose [нос] Est. *nina* Vot. *nenä* L-L. *nenä*
 Soi. *nenä* Fin. *nenä*

Proto-Finnic **nenä* ~ **nęna* ~ **nana* 'nose' is related to Proto-Saami **ñuonē* 'nose' (cf. SSA 2 : 213; EES 313–314).

64. not [не] Est. *ei* Vot. *eb* L-L. *ei*
 Soi. *ei* Fin. *ei*

Proto-Finnic negative verb **e-* goes back to the Proto-Uralic negative verb **e-* (cf. SSA 1 : 99; EES 59; UEW 68–70).

65. one [один] Est. *üks* Vot. *ühs(i)* L-L. *üks*
 Soi. *ükš* Fin. *yksi*

Proto-Finnic **ükci* 'one' goes back to the Proto-Uralic word for 'one', attested from Finnic to Mansi (cf. SSA 3 : 489; EES 635; UEW 81). However, the exact phonetic reconstruction of the Proto-Uralic form is difficult.

66. rain [дождь] Est. *vihm* Vot. *vihmδ* L-L. *vihma*
 Soi. *vihma* Fin. *sade*

Proto-Finnic **vihma* 'rain' is related to Proto-Saami **vęsmē* 'light snow' (cf. SSA 3 : 438; EES 601). In Finnish, *vihma* means 'drizzle' and a derivative from Proto-Finnic **sata-* 'to rain, to snow' (< Proto-Uralic **śada-* 'to rain') is used as the main word for 'rain' instead (cf. SSA 3 : 141, 160, EES 455–456).

67. red [красный] Est. *punane* Vot. *kauniz* L-L. *punnain*
 Soi. *puñnain* Fin. *punainen*

Proto-Finnic **punainęn* 'red' is derived from Proto-Finnic **puna* 'red colour' — a reflex of Proto-Uralic **puna* 'hair, fur' (cf. SSA 2 : 426–427; EES 137; UEW 402). The semantic development may look strange, but is actually understandable. The words for 'hair' in Eurasia frequently have an additional meaning 'colour'. An intermediate meaning 'hair colour (of animals)' is actually attested for reflexes of PU **puna* in Hill Mari and South Khanty. The following path of semantic development can be supposed in this case: 'hair, fur' > '(hair) colour' > 'red colour'. In Votic, the main word for 'red' is *kauniz*, going back to Proto-Finnic **kaunis* 'beautiful', borrowed from Germanic (LÄGLOS II 62). The semantic shift 'beautiful' > 'red' occurred under the influence of Russian *красный* 'red/beautiful'.⁸

68. road [дорога] Est. *tee* Vot. *tee* L-L. *tie*
 Soi. *tęę* Fin. *tie*

Proto-Finnic **tee* 'road' is apparently related to Komi *tuj* 'road', although the reconstruction of a common protoform is difficult (cf. SSA 3 : 288; EES 520; UEW 794).

69. root [корень] Est. *juur* Vot. *juuri* L-L. *juuri*
 Soi. *juuri* Fin. *juuri*

⁸ *Kauniz* did not preserve the original meaning 'beautiful' in Luuditsa Votic, but this meaning was observed in some Central Votic varieties (VKS 408).

Proto-Finnic **juuri* 'root' goes back to Proto West Uralic **juwri* 'root', attested also in Mordvin (cf. SSA 1 : 253; EES 102; UEW 639). This word replaces Proto-Uralic **wanča* 'root' (cf. UEW 548–549).

70. round [круглый] Est. *ümmargune* Vot. *ümmekajñ* L-L. *ümmekäin*
 Soi. *ümlerläin* Fin. *pyöreä*

Proto-Finnic **pööredä* 'round', reflected in Finnish, has cognates with the same meaning in Ob-Ugric languages and goes back to Proto-Uralic **penirä* 'round' (cf. SSA 2 : 455; EES 406; UEW 372–373). In other languages in our sample, the word 'round' is derived from Proto-Finnic **ümpärä*, a Germanic loanword with a Finnic suffix **-rä* (cf. SSA 3 : 491; EES 636–637; LÄGLOS III 426–427). There is no difference between '3D round' and '2D round'.

71. sand [песок] Est. *liiv* Vot. *liivâ* L-L. *liiva*
 Soi. *liiva* Fin. *hiekkä*

Proto-Finnic **liiva* 'sand' may be a Baltic or Germanic loan (SSA 2 : 205; EES 240; LÄGLOS II 207). Although now Ingrian is the only North Finnic language that has this word for 'sand', the Proto-Finnic status of the word is confirmed by the fact that it was borrowed from a lost North Finnic idiom into Permic languages: Komi *lja*, Udmurt *luo* 'sand' (Saarikivi 2006 : 36). In Finnish, a specific word *hiekkä* is used instead (SSA 1 : 160).

72. to say [сказать] Est. *ütelda ~ öelda* [üttema] Vot. *juteľľäđ* L-L. *sanno*
 Soi. *šainoa* Fin. *sanoa*

Proto-Finnic **sano-* ~ **seno-* 'to say' is derived from **sana* ~ **sena* 'word' (cf. SSA 3 : 155; EES 494). In Estonian and Votic, this word is replaced by the reflexes of Proto-Finnic **jutta-* 'to talk; to tell, narrate', going back to Proto-Uralic **jupta-* 'to tell, narrate' (cf. SSA 1 : 252; EES 102, 627; UEW 104; Aikio 2002 : 48). The Estonian verb demonstrates an irregular change **ju-* > *ü-*. The original anlaut is preserved in the Estonian noun *jutt* 'story; talk'. The Estonian reflex of **seno-* has a clearly secondary meaning 'to scold'.

73. to see [видеть] Est. *näha* [nägema] Vot. *nähhâ* L-L. *nähä*
 Soi. *näh(h)ä* Fin. *nähdä*

Proto-Finnic **näke-* 'to see' goes back to Proto-Uralic **näki-* 'to see' (cf. SSA 2 : 249; EES 326–327; UEW 302).

74. seed [семя] Est. *seeme* Vot. *seemene* L-L. *siemen*
 Soi. *šęęmen* Fin. *siemen*

Proto-Finnic **seemen* 'seed' is a Baltic borrowing (SSA 3 : 173; EES 464).

75. to sit [сидеть] Est. *istuda* [istuma] Vot. *issua* L-L. *isto*
 Soi. *ištua* Fin. *istua*

Proto-Finnic **istu-* 'to sit' goes back to Proto West Uralic **isa-* 'to sit', which may be an Indo-European borrowing (cf. SSA 1 : 229; EES 94; UEW 629).

76. skin [кожа] Est. *nahk* Vot. *nahkâ* L-L. *nahka*
 Soi. *nahka* Fin. *iho*

Proto-Finnic **iho* 'skin' goes back to Proto-Uralic **iša* 'skin, surface' and is preserved in Finnish (cf. SSA 1 : 222; EES 89; UEW 636–637). Other idioms use Proto-Finnic **nahka* 'skin, hide', borrowed from Germanic (SSA 2 : 202; EES 306; LÄGLOS II 287–288). In Finnish, there is a word *nahka* but it has a more specific meaning (mainly it is 'a skin of an animal, fur' but in colloquial speech it can be easily used in the test contexts).

77. to sleep [спать] Est. *magada* [magama] Vot. *magatâ* L-L. *maata*
 Soi. *maada* Fin. *nukkua*

The Germanic etymology of Proto-Finnic **maka-* 'to sleep', *pace* LÄGLOS, does not seem convincing to us (SSA 2 : 136; EES 270; LÄGLOS II 237–238). In Finnish, this word means 'to lie' (see above) and the meaning 'to sleep' is expressed by the reflex of Proto-Finnic **nukku-* 'to doze, to drowse' (SSA 2 : 237), cognate with Proto-Saami **nokke-* 'to doze, to drowse' (SSA 2 : 237). The Proto-Uralic word for 'to sleep' was **adi-* (cf. UEW 334; Aikio 2015 : 51).

78. small, little [маленький] Est. *väike* Vot. *peen(i)* L-L. *pienu* (~ *pikkarain*)
Soi. *pikkarain* Fin. *pieni*

Proto-Finnic **peeni* 'small' is preserved in Votic and Finnish (cf. SSA 2 : 348; EES 358). The Germanic etymology of this word is not convincing (LÄGLOS III 55). This word exists in Estonian but rather means 'thin, fine'. In Soikkola Ingrian, the word *pikkarain* (that also exists in Finnish) predominates, but in Lower Luga it is not the most prevalent variant. In Votic, *pikkegrajn* is less common than *peen(i)*. According to SSA 2 : 361, this word is a hypocoristic byform of **peeni*. In Estonian, the main word for 'small' is derived from Proto-Finnic **vähä* 'small', which may go back to Proto West Uralic (cf. SSA 3 : 478; EES 618–619; UEW 818–819). Germanic etymologies, proposed for this word, are dubious (LÄGLOS III 420). The semantic difference between **peeni* and **vähä* on the Proto-Finnic level remains elusive.

79. smoke [дым] Est. *suits* Vot. *savvu* L-L. *savvu*
Soi. *šavvu* Fin. *savu*

Proto-Finnic **savu* 'smoke' goes back to Proto West Uralic **sjiwi* 'smoke' (cf. SSA 3 : 163; UEW 754). The Estonian word is a reflex of Proto-Finnic **suiccu* 'smoke', with potential cognates in Saami meaning 'to rise' (cf. SSA 3 : 208; EES 486). This word is also attested in Finnish dialects. Since reflexes of **savu* are the main words for 'smoke' in Livonian and South Estonian, there can be no doubt that the main Proto-Finnic word for 'smoke' was **savu*.

80. to stand [стоять] Est. *seista* [*seisma*] Vot. *sejssâ* L-L. *seissa*
Soi. *šeišša* Fin. *seisoa*

Proto-Finnic **saisa-* 'to stand' goes back to Proto-Uralic **saŋša-* 'to stand' (cf. SSA 3 : 164–165; EES 466; UEW 431–432).

81. star [звезда] Est. *täht* Vot. *tähti* L-L. *tähti*
Soi. *tähti* Fin. *tähti*

Proto-Finnic **tähti* 'star' is related to Saami and Mordvin words for 'star' and the Mari word for 'sign' (cf. SSA 3 : 353; EES 565; UEW 793–794). However, irregular sound correspondences between these forms suggest that the word was borrowed from an unknown substrate separately in already differentiated branches of West Uralic (Aikio 2015 : 43–47). This word replaced Proto-Uralic **kuńsi* 'star' (cf. UEW 210–211).

82. stone [камень] Est. *kivi* Vot. *tšivi* L-L. *kivi*
Soi. *kivi* Fin. *kivi*

Proto-Finnic **kivi* 'stone' goes back to Proto-Uralic **kiwi* 'stone' (cf. SSA 1 : 378; EES 163–164; UEW 163–164).

83. sun [солнце] Est. *päike* Vot. *päivüd* L-L. *päivükkäin*
Soi. *päivüd* Fin. *aurinko*

Proto-Finnic **päivä* 'sun, day' goes back to Proto-Uralic **päjvä*, whose reflexes mean 'sun, day' in Saami and 'heat, warm' in Samoyed (cf. SSA 2 : 456; EES 403; UEW 360). Finnish *päivä* means 'day' only. In the meaning

'sun', the word is replaced by *aurinko*, which has no acceptable etymology (SSA 1 : 90).

84. to swim [плыть, плавать] Est. *ujuda* [*ujuma*] Vot. *ujjua* L-L. *ujjo*
Soi. *ujjua* Fin. *uida*

Proto-Finnic **ui-* 'to swim' goes back to Proto-Uralic **uji-* 'to swim' (cf. SSA 3 : 368; EES 576–577; UEW 542).

85. tail [хвост] Est. *saba* Vot. *äntä* L-L. *händä*
Soi. *händä* Fin. *häntä*

Proto-Finnic **hätä* 'tail' has no acceptable etymology: supposed cognates in other branches of Uralic show irregular correspondences (cf. SSA 1 : 208; EES 85; UEW 56). In Estonian, it also exists but the main word for 'tail' was borrowed from the Baltic languages (EES 455). The Proto-Uralic word for 'tail' was **ponči*.

86. that [тог] Est. *see* Vot. *see*⁹ L-L. *see*
Soi. *šee* Fin. *tuo*

Both Proto-Finnic **se* 'that' (SSA 3 : 163; EES 463–464; UEW 33–34) and Proto-Finnic **too* 'that' (cf. SSA 3 : 327–328; EES 538; UEW 526–528) have Uralic pedigree. However, it is difficult to reconstruct the Proto-Finnic demonstrative system. Finnic dialects have different systems: monopartite, bipartite or tripartite. Standard Estonian has a formally bipartite system *see* ~ *too* but it functions rather as a monopartite system where *see* means 'this/that' and in the contrastive contexts the word *teine* 'other' is usually used. Finnish has a tripartite system *tämä* ~ *tuo* ~ *se*, and in the test contexts *tuo* is preferable. Votic and Ingrian have bipartite systems but *see* is often used in the contexts for 'this'.

87. this [этот] Est. *see* Vot. *kase* L-L. *tämä*
Soi. *tämä* Fin. *tämä*

According to Laanest (1982 : 196), Votic *kase* results from the merging of some interjection with *se*. *Tämä* is a Uralic word (SSA 3 : 355; UEW 513–515). Estonian *tema* and Votic *tämä* are 3Sg personal pronouns but not demonstrative pronouns. Since the typical path of diachronic development leads from demonstrative pronouns to personal pronouns, but not vice versa, we can suppose that Proto-Finnic **tämä* 'this' was a demonstrative (see comments on the previous word).

88. tongue [язык] Est. *keel* Vot. *tšeele* L-L. *kieli*
Soi. *keeli* Fin. *kieli*

Proto-Finnic **keeli* 'tongue' goes back to Proto-Uralic **käli* 'tongue' (cf. SSA 1 : 353; EES 140; UEW 144–145).

89. tooth [зуб] Est. *hammas* Vot. *amməz* L-L. *hammäz*
Soi. *hammaž* Fin. *hammas*

Proto-Finnic **hambas* 'tooth' is a Baltic loanword (SSA 1 : 136; EES 68–69). This word replaced Proto-Uralic **piŋi* 'tooth', whose Finnic reflex **pii* means 'tooth in a saw, rake etc.' (cf. SSA 2 : 352; UEW 382).

90. tree [дерево] Est. *puu* Vot. *puu* L-L. *puu*
Soi. *puu* Fin. *puu*

Proto-Finnic **puu* 'tree' goes back to Proto-Uralic **pawi* 'tree' (cf. SSA 2 : 443–444; EES 396–397; UEW 410–411).

⁹ The spelling of this word in Votic and Ingrian is approximate as there is significant variation in the length of this vowel. We spell it with long *ee*.

91. two [два]	Est. <i>kaks</i>	Vot. <i>kahs(i)</i>	L-L. <i>kaks</i>
	Soi. <i>kakš</i>	Fin. <i>kaksi</i>	

Proto-Finnic **kakci* 'two' goes back to the Proto-Uralic numeral 'two', whose exact phonetic shape is hard to reconstruct (cf. SSA 1 : 282; EES 120; UEW 118–119).

92. warm [теплый]	Est. <i>soe</i>	Vot. <i>soojā</i>	L-L. <i>sooja</i>
	Soi. <i>lämmää</i>	Fin. <i>lämmiin</i>	

Proto-Finnic **lämbin* 'warm' goes back to Proto-Uralic **lämpi* 'warm' (cf. SSA 2 : 124; EES 263; UEW 685; Aikio 2002 : 13). The word *lämmi* exists in Estonian dialects (EES 263), and the same root is known in Votic (mostly through the word *lämmittä(ä)* 'to stoke', VKS 657). Other idioms use the root **sooja* 'shelter; warm', borrowed from an Iranian word for 'shade' (cf. SSA 3 : 214; EES 478; UEW 748–749). In Finnish, there is a word *suoja* but it is not the main word for 'warm' (it is used when speaking about above-zero weather). In Ingrian, the same root is observed only in the Lower Luga dialect (Nirvi 1971 : 542).

93. water [water]	Est. <i>vesi</i>	Vot. <i>vesi</i>	L-L. <i>vesi</i>
	Soi. <i>veži</i>	Fin. <i>vesi</i>	

Proto-Finnic **veci* 'water' goes back to Proto-Uralic **weti* 'water' (cf. SSA 3 : 429; EES 599; UEW 570).

94. we [мы]	Est. <i>meie</i> ~ <i>me</i>	Vot. <i>müü</i>	L-L. <i>müö</i>
	Soi. <i>mõõ</i>	Fin. <i>me</i>	

Proto-Finnic **me(k)* 'we' goes back to Proto-Uralic **me(-)* 'we' (cf. SSA 2 : 156; EES 279; UEW 294–295). In Estonian, there is variation between a long and a short form.

95. what [что]	Est. <i>mis</i>	Vot. <i>mikä</i>	L-L. <i>mikä</i>
	Soi. <i>migä</i>	Fin. <i>mikä</i>	

Proto-Finnic **mi(kä)* 'what' goes back to Proto-Uralic **mi* ~ **mj* 'what' (cf. SSA 2 : 164; UEW 296). In Estonian, the formative *-s* originates from a demonstrative pronoun *see* (EES 282–283).

96. white [белый]	Est. <i>valge</i>	Vot. <i>valka</i>	L-L. <i>valke</i>
	Soi. <i>valkkia</i>	Fin. <i>valkoinen</i>	

Proto-Finnic **valkeḁa* 'white' goes back to Proto-Uralic **wylki* 'light' (cf. SSA 3 : 399–400; EES 588; UEW 554–555; Aikio 2015 : 59). In Finnish, the derivate with the adjectival suffix *valkoinen* looks more natural in the test contexts than *valkea* 'white'.

97. who [кто]	Est. <i>kes</i>	Vot. <i>tšen</i>	L-L. <i>ken</i>
	Soi. <i>ken</i>	Fin. <i>kuka</i>	

Proto-Finnic **ken* 'who' goes back to Proto-Uralic **ke(-)* 'who' (cf. SSA 1 : 342–343; EES 145–146; UEW 140–141). The Estonian word has the formative *-s* absent from three of the idioms, however the variant *ken* is observed in the Estonian dialects (EES 145–146). In Finnish, the main word for 'who' is *kuka* < Proto-Uralic interrogative stem **ku(-)*, used in words for 'where', 'which', etc. (SSA 1 : 423–424; UEW 191–192), but the word *ken* also exists as a poetic variant.

98. woman [женщина]	Est. <i>naine</i>	Vot. <i>najn</i>	L-L. <i>nain</i>
	Soi. <i>nain</i>	Fin. <i>nainen</i>	

Proto-Finnic **nainen* 'woman' (cf. SSA 2 : 202; EES 306) is derived from a root **naa-*, seen also in *naaras* 'female' (SSA 2 : 200–201). This root goes back to Proto-Uralic **näxi* 'woman' (Janhunen 1981 : 245–246).

99. yellow [желтый] Est. *kollane* Vot. *keļtejn* L-L. *keltain*
 Soi. *kelttain* Fin. *keltainen*

Proto-Finnic **keltainen* 'yellow' consists of the root borrowed from Baltic, and an adjectival suffix (SSA 1 : 342; EES 172–173).

100. you (thou) [ты] Est. *sina* (~ *sa*) Vot. *siä* L-L. *siä*
 Soi. *šiiä* Fin. *sinä* (~ *sä*)

Proto-Finnic **cinä* 'thou' goes back to Proto-Uralic **tin* 'thou' (cf. SSA 3 : 184; EES 473–474; UEW 539). In Estonian and Finnish, there is variation between a long and a short form.

101. far [далеко] Est. *kaugel* Vot. *kaukallā* L-L. *kaukall*
 Soi. *ettää/ettäl* Fin. *kaukana*

Proto-Finnic **kauka-* 'far' is a Germanic loanword (Aikio 2000; EES 137). Supposed cognates in Mordvin and Khanty (SSA 1 : 330–331; UEW 132) are phonetically incompatible with the Finnic word. Soikkola Ingrian uses a reflex of Proto-Finnic **etä-* 'far', going back to Proto West Uralic **ečä-* 'far' (cf. SSA 1 : 109–110; UEW 624). This word is the main word for 'far' also in Veps. It is hard to say which of these two words was the main Proto-Finnic word for 'far'.

102. heavy [тяжелый] Est. *raske* Vot. *rankkā* L-L. *rankka*
 Soi. *raškaž* Fin. *raskas*

There are two different words: Proto-Finnic **rankka* 'heavy', apparently borrowed from Germanic (cf. SSA 3 : 47; EES 419, 445; LÄGLOS III 124–125), and Proto-Finnic **raskas* 'heavy' (cf. SSA 3 : 52; EES 419–420). The former became dominant in Votic and Lower Luga Ingrian, the latter in three other idioms. Estonian *ränk* is more bookish than *raske*. It is difficult to tell which word was the main word for 'heavy' in Proto-Finnic.

103. near [близко] Est. *lähedal* Vot. *litši* L-L. *liki*
 Soi. *ligi* Fin. *lähellä*

Proto-Finnic **lähe-* 'near', going back to Proto-Uralic **läsi* 'near', is preserved in Estonian and Finnish (cf. SSA 2 : 122; EES 262; UEW 687; Aikio 2002 : 48). Votic and Ingrian use another root, Proto-Finnic **liki* 'near', cognate with Proto-Saami **leke* 'near' (cf. SSA 2 : 76; EES 238). Estonian *ligidal* and Finnish *liki* ~ *likellä* are synonymic forms but are less general or neutral. The original semantic difference between **lähe-* and **liki* in Proto-Finnic is not clear.

104. salt [соль] Est. *sool* Vot. *soolā* L-L. *suola*
 Soi. *šõõla* Fin. *suola*

Proto-Finnic **soola* 'salt' is borrowed from an Indo-European language, most probably from Baltic (cf. SSA 3 : 214–215; EES 480). Similar loanwords exist in other Uralic languages (UEW 750–751), but the phonetic shape of the Finnic word (long vowel in an *a*-stem) shows that it was borrowed independently.

105. short [короткий] Est. *lühike* Vot. *lühüd* L-L. *lühüd*
 Soi. *lühüd* Fin. *lyhyt*

Proto-Finnic **lühüt* 'short' has no acceptable etymology (cf. SSA 2 : 117; EES 266).

106. snake [змея] Est. *uss* Vot. *mato* L-L. *mato*
 Soi. *mado* Fin. *käärme*

Although Proto-Finnic **küü* 'viper, snake', going back to Proto-Uralic **küji* 'snake', retains the meaning 'snake' in Karelian, Veps, and Livonian dialects (cf. SSA 1 : 467; UEW 154–155), these are hardly the main words for 'snake'

in the respective idioms. Proto-Finnic **mato* 'snake, worm' was perhaps the main word for 'snake' already in the proto-language. It is possibly a Germanic borrowing. According to an alternative etymology, **mato* is cognate with Proto-Saami **muocē* 'moth' (cf. SSA 2 : 154; EES 270; LÄGLOS II 255). In Finnish, this word means 'worm' (see below), and another word, ultimately borrowed from Baltic, is used for 'snake' (SSA 1 : 484). In Estonian, the word *madu* means 'snake', but a more neutral word is *uss* 'snake, worm'. The etymology of *uss* is not clear but it is possibly a Russian borrowing (EES 580).

107a. thin (2D) [тонкий] Est. *õhuke* Vot. *hojkkê* L-L. *hoikka*
 Soi. *hoikka* ~ *hoikkain* Fin. *ohut*

Two words can be reconstructed: Proto-Finnic **ohut* 'thin' (cf. SSA 2 : 260; EES 625) and Proto-Finnic **hoikka* 'thin' (cf. SSA 1 : 169). The former goes back to Proto-Uralic **vokši* 'thin' (Решетников 2011 : 110; Luobbal Sámmol Sámmol Ánte (Aikio) 2014 : 10–11), the latter has no known etymology. It is hard to reconstruct the semantic difference between these words on the Proto-Finnic level. The word *hoikka* also exists in Finnish but is not predominant there, while the reflexes of **ohut* are not predominant in Votic and Ingrian. In Soikkola Ingrian, there is a variant with an adjectival suffix.

107b. thin (1D) [тонкий] Est. *peenike* Vot. *hojkkê* L-L. *hoikka*
 Soi. *hoikka* Fin. *ohut*

In Votic, Finnish and Lower Luga Ingrian there is no difference between '2D thin' and '1D thin'. In Soikkola Ingrian, the variant with the suffix is not typical in the test contexts. In Estonian, the derivate from *peen* 'small' (see above) is more typical in the test contexts (the form *peen* without a suffix is also possible in the test contexts but *peenike* looks more neutral).

108. wind [ветер] Est. *tuul* Vot. *tuuli* L-L. *tuuli*
 Soi. *tuuli* Fin. *tuuli*

Proto-Finnic **tuuli* 'wind' goes back to Proto-Uralic **tjwli* 'wind' (cf. SSA 3 : 340; EES 558–559; UEW 800).

109. worm [червь] Est. *uss* Vot. *matokkejn* ~ *mato* L-L. *matokkain* ~ *mato*
 Soi. *madokkain* ~ *mado* Fin. *mato*

The distinction snake vs worm is not typical for Finnic languages. Among the five analysed idioms only Finnish distinguishes these two notions, while in the other languages this distinction is not relevant. Thus, we can reconstruct Proto-Finnic **mato* 'snake, worm'. In Votic and Ingrian, a derivate with the diminutive suffix can be used to stress that it is a worm but not a (big) snake. (See comments to the word for 'snake', #106.)

110. year [год] Est. *aasta* Vot. *voosi* L-L. *vuosi* ~ *aastaika*
 Soi. *võõž* Fin. *vuosi*

Proto-Finnic **vooci* 'year' goes back to Proto-Uralic **j̥di* 'year' (cf. SSA 3 : 476; EES 612–613; UEW 335–336). In Estonian, this word means 'harvest' and another word is used for 'year' (etymologically a compound **aiyasta-aika* built from the forms of **aika* 'time', see EES 42). In Lower Luga Ingrian, both words are used; the choice depends on the particular idiolect.

3. Discussion

In the current section we formulate some observations on the compiled wordlists. These are preliminary observations that do not purport to be a comprehensive analysis of the data.

3.1. The analysed set of five languages is rather homogeneous. Among 111 items, 77 (69%) have the same word in all five varieties. There are no items where all five idioms use different roots neither are there items with four different roots. There are only three items in the list where three roots appear: #47 'to lie' Est. *lamada* vs Fin. *maata* vs Vot. *ležiä*, L-L. *ležže*, Soi. *leššiä*, #106 'snake' Est. *uss* vs Fin. *käärme* vs Vot. *mato*, N L. *mato*, Soi. *mado*, #107b 'thin' Est. *peenike* vs Fin. *ohut* vs Vot. *hojkkê*, L-L. *hoikka*, Soi. *hoikka*. For all three items opposition is organized in the same way: Estonian opposes Finnish and they both oppose three minor varieties, which have the same root.

In all other cases, either one language has a root that is different from the other languages (24 items) or two languages differ from the other three (7 items).

3.2. The three minor varieties are rather uniform; the two major languages are often different from the minor ones.

The three minor Finnic varieties do not demonstrate significant diversity. Only in 8 cases (i.e. 7%), the roots were not the same. Ingrian is opposed to all other varieties in #48 'liver' and #101 'far'; Votic is different from all other varieties in #67 'red' (and this difference would not hold if we take other Votic varieties into account); in two cases Votic is uniform only with Estonian (#72 'say' and #87 'this'); in one case Votic and Lower Luga Ingrian are different from the other varieties including Soikkola Ingrian (#102 'heavy'), in another Soikkola Ingrian and Finnish differ from the other varieties (#92 'warm'), and there is also a specific Estonian word which exists as one of the two variants for Lower Luga Ingrian (#110 'year'). Summing up, the number of cases where a minor variety does not have the same root as the two other minor varieties is the following: Votic — 3 items, Soikkola Ingrian — 4 items, Lower Luga Ingrian — 1 item.

However, the situation with major languages is quite different. There are 11 cases where Estonian has a root that differs from all other languages (#4 'belly', #7 'to bite', #16 'to die', #21 'earth', #47 'to lie', #78 'small', #79 'smoke', #85 'tail', #106 'snake', #107b 'thin (1D)' and #109 'small, little') and 5 cases where the Estonian root is found in one other variety but where they are opposed to the other three varieties (#72 'to say', #87 'this', #103 'near', #107a 'thin (2D)', and #110 'year'). In Finnish, a root is opposed to all other varieties in 16 cases (#3 'bark', #5 'big, large', #35 'green', #47 'to lie', #53 'many, a lot of', #56 'mountain', #66 'rain', #77 'to sleep', #70 'round (3D)', #71 'sand', #76 'skin', #83 'sun', #86 'that', #97 'who', #106 'snake' and #107b 'thin (1D)'), and there are 3 cases where a Finnish root is the same as in one of the other varieties but different from all others (#92 'warm', #103 'near', #107a 'thin (2D)').

3.3. Since none of the five varieties is isolated from all others, the compiled lists should be considered also from the point of view of language contact. The most typical directions of borrowing for these varieties are the following:

(a) Votic borrowed many words from Ingrian. Usually it is difficult to define whether it was a borrowing from Soikkola Ingrian adapted to Votic phonetics or a borrowing from Lower Luga Ingrian. Also there are two types of borrowings: regular borrowings (e.g. Vot. *hüü* 'they', *kärkkü* 'cone') and recent "double-layer" borrowings where the Ingrian pronunciation of a word

replaced the original Votic variant (e.g. *auki* 'pike', *hiili* 'coal', *haapezikko* 'aspen forest' cf. proper Votic *autši*, *iili*, *aapezikko*).

In the compiled Swadesh lists, we did not notice obvious borrowings from Votic into Ingrian.¹⁰ If a word, which has specific phonetic differences between Votic and Ingrian, is borrowed from Ingrian into Votic, it usually keeps the Ingrian phonetic shape (e.g. the initial [h], or [k] before a front vowel). However, for all such pairs which appear in our Swadesh lists, Votic has its original phonetic shape so we cannot assume that these words were borrowed: cf. #14 'cold' Soi. *külmä*, Vot. *tšülmə*, #34 'good' Soi. *hüvä*, Vot. *üvä*, #36 'hair' Soi. *hiuž*, Vot. *ivuz*, #37 'hand' Soi. *käži*, Vot. *tšäsi*, #49 'long' Soi. *pitkä*, Vot. *pittšə~pittši*, #56 'mountain' Soi. *mägi*, Vot. *mätši*, #58 'nail' Soi. *künž*, Vot. *tšünsi*, #82 'stone' Soi. *kivi*, Vot. *tšivi*, #85 'tail' Soi. *händä*, Vot. *äntə*, #88 'tongue' Soi. *keeli*, Vot. *tšeei*, #89 'tooth' Soi. *hammaž*, Vot. *amməz*, #97 'who' Soi. *ken*, Vot. *tšen*, #99 'yellow' Soi. *kelttain*, Vot. *keatejn*, #103 'near' Soi. *ligi*, Vot. *litši*.

Based on this, we can state that the Swadesh list is stable from the point of view of new borrowings.

(b) As Lower Luga Ingrian is a convergent language on the basis of Votic and Ingrian, it could have taken many words from Votic. However, among 111 words of the core lexicon, there is only one possible candidate for such a borrowing: the word #102 *rankkA* 'heavy' (Vot. *rankkə*). In the three other varieties, another root is observed. We do not have solid evidence that this word came from Votic and was not some dialectal variant in Ingrian.

(c) One can also expect some borrowings from Finnish via the Ingrian Finnish dialect into Votic or into Lower Luga Ingrian. However, we did not notice such candidates in the compiled lists. The same concerns the borrowings from Estonian into Lower Luga Ingrian: usually, they are not from the core lexicon (e.g. *kleit* < Est. *kleit* 'dress').

3.4. Diversity in the core lexicon is explained by different reasons. Among the 34 items where the five varieties were not uniform, several groups of words are distinguished.

a. The biggest group appeared because of quasi-synonymic words that existed in Proto-Finnic.¹¹ It happened (usually without obvious reason) that one word became predominant in one language and its synonym became predominant in another language. This situation is observed with the following items. Estonian: #4 'belly' Est. *kõht* vs Fin. *vatsa*¹², #78 'small, little' Est. *väike* vs Fin. *pieni*; Estonian and Finnish: #103 'near' Est. *lähedal*, Fin. *lähellä* vs Vot. *litši*, #107a 'thin(2D)' Est. *õhuke*, Fin. *ohut* vs Vot. *hojkkə*; Finnish: #5 'big, large' Fin. *iso* vs Est. *suur*, #53 'many, a lot of' Fin. *monta* vs Est. *palju* (as well as the alternative Finnish variant *paljon*), #70 'round(3D)' Fin. *pyöreä* vs Est. *ümmargune*, #76. 'skin [кожа]' Fin. *iho* vs Est. *nahk*, #86 'that' Fin. *tuo* vs Est. *see*, #107b 'thin(1D)' Fin. *ohut* vs Soi. *hoikka*; Finnish and Soikkola Ingrian: #92 'warm' Fin. *lämmän*, Soi. *lämmää* vs Est. *soe*; Soikkola Ingrian: #101 'far' Soi. *ettää/ettäl* vs Fin. *kaukana*; Votic and Lower Luga Ingrian: #102 'heavy' Vot. *rankkə*, L-L. *rankkA* vs Fin. *raskas*.

¹⁰ The examples given in the previous paragraph are not from our Swadesh lists.

¹¹ Of course, in such cases one of the quasi-synonyms must have been "basic" in Proto-Finnic. Additional research is needed to determine the precise semantic difference between such quasi-synonyms at the Proto-Finnic level.

¹² In cases where several languages have the same root, we give examples only from one of these languages. In Section 2 one can find words with this root in other varieties.

b. Some words appeared in the list because of a semantic shift.¹³ They already existed in Proto-Finnic but in some language(s) they changed their meaning and became predominant for the corresponding item in the Swadesh list. In some cases, the semantic shift happened in a majority of the varieties, so that only one language preserves the original Proto-Finnic root while the others use another root for the item in the list. This is the case, for example, with #56 'mountain' where only Finnish retains the original Finnic root for 'mountain'.

The words that have a different root due to a semantic shift specific to Estonian are #16 'to die' *surra*, #21 'earth' *muld*, #79 'smoke' *suits*, and #107b 'thin(1D)' *peenike*. Specific to Estonian and Votic are the words #72 'to say' Est. *ütelda/öelda*, Vot. *jutʃaːaː*. In Votic, the word #67 'red' *kauniz* shifted its meaning from 'beautiful' to 'red'. The aforementioned word #56 'mountain' underwent a semantic shift in all varieties except Finnish: Est. *mägi*, Vot. *mätši* L-L. *mäki* Soi. *mägi*. Specific to Finnish are also the words #3 'bark' *kaarna*, #47 'to lie' *maata*, #77 'to sleep' [спать] *nukkua*, #97 'who' *kuka*, and #106 'snake' *käärme*.¹⁴

c. In rare cases a new derivative from the old root traced to Proto-Finnic or earlier becomes a predominant word in a language. In Estonian, such words are #47 'to lie' *lamada* and #110 'year' *aasta*. The latter word also appears in Lower Luga Ingrian: *aastaika* is one of the variants for 'year' (see Section 2). In all varieties except Finnish, the word #35 'green' is an adjective derived from the noun with the meaning 'grass': Est. *roheline*, Vot. *rohojn*, L-L. *rohoin*, Soi. *rohhojn*. In Finnish, the noun #66 'rain' *sade* is derived from the original verb. Possibly, a Soikkola Ingrian compound #48 'liver' *leibä-liha* built from two Finnic roots should be placed in this group too.

d. In spite of the fact that the core lexicon is relatively stable, new (post-Proto-Finnic) loan words can replace the original words. In Estonian, the word *saba* (#85 'tail') was borrowed from the Baltic languages, and the word *uss* (both #106 'snake' and #109 'worm') was possibly borrowed from Russian. In all three minor varieties, the word for #47 'to lie' was borrowed from Russian: Vot. *ležiä*, L-L. *ležže*, Soi. *leššiä*. In Soikkola Ingrian, one of the variants for #48 'liver' is also a Russian loanword: *petšonka*.

e. In addition to the described groups, there are two Finnish words with unclear etymology: #71 'sand' *hiekkä* and #83 'sun' *aurinko*. Also, the word *imi* (#59 'name'), which is predominant in Soikkola Ingrian and is present in Votic as one of two variants, does not belong unambiguously to one of the proposed groups: it could be either a borrowing from Russian or a contamination (see Section 2).

The distribution of the divergent part of the core lexicon among the discussed groups and varieties is summarized in Table 1.¹⁵

¹³ By "semantic shift" we mean not only a proper change of meaning but also finer modifications, e. g. stylistic changes.

¹⁴ It is unlikely that *käärme* is a new borrowing, because the Northern Finnic languages did not have contact with the Baltic languages since the Proto-Finnic period.

¹⁵ Note that a word of Finnic origin that did not change its meaning and was not a derivative was counted only in group "a" and only in cases where this word was not predominant for most of the varieties under discussion. In general, this table analyses only the words where these varieties demonstrate diversity while changes (e.g. semantic shifts) that happened in all five varieties are not studied here.

Table 1

Causes of lexical innovations

	Est.	Vot.	L-L.	Soi.	Fin.	Overall
a. Synonyms	4	1	1	2	8 + (1)	16 + (1)
b. Semantic shift	6	3	1	1	5	16
c. New derivatives	3	1	1 + (1)	1 + (1)	1	7 + (2)
d. New borrowings	3	1	1	1 + (1)		6 + (1)
e. Other		(1)		1	2	3 + (1)
Overall	16	6 + (1)	4 + (1)	6 + (2)	16 + (1)	

3.5. The distribution of words in the core lexicon does not correlate with borders between Finnic sub-groups.

One might expect that many of the analysed words would oppose southern Finnic languages (Estonian and Votic) and northern Finnic languages (Finnish and the two dialects of Ingrian). In fact, only two items demonstrate such an opposition: #72 'to say' and #87 'this' (the latter case is not pure since Votic uses a more complicated morphological form than Estonian: *kase* vs *se*). Even if we take into account the fact that Lower Luga Ingrian was heavily influenced by Votic and possibly should not be unambiguously considered a northern Finnic language, the situation would not change: only one word opposes Finnish and Soikkola Ingrian to the other varieties: #92 'warm'. This fact has two theoretically possible interpretations: (a) the difference between the two Finnic branches is not considerable enough to be reflected in the core lexicon represented in the Swadesh list; (b) in a contact zone between closely related languages, convergent processes can play a part (e.g. one of the existing basic words becomes predominant under the influence of the neighbouring idiom). Both interpretations can only be confirmed through a thorough analysis of individual words, and this task is beyond the scope of the current paper.

Table 2 presents pairwise comparisons of the Swadesh lists. In the upper-right part of the table, the percentage of the common roots is given. In the lower-left part of the table, the number of words that have different roots is indicated. Rare cases where a language has two roots for the same item (e.g. Finnish *paljon* ~ *monta* 'many, a lot of' or Lower Luga Ingrian *voosi* ~ *aastaikA*) but the second language in the pair only has one of these roots were counted as 0.5 instead of 1.

Table 2

Lexicostatistical distances

	Est.	Vot.	L-L.	Soi.	Fin.
Est.		86 %	86 %	83 %	75 %
Vot.	16		97 %	94 %	80 %
L-L.	16	3.5		96 %	82 %
Soi.	19	7	4.5		82 %
Fin.	27.5	22.5	20	19.5	

The closest varieties are Votic and Lower Luga Ingrian, which formally belong to different Finnic branches. In general, the distance between all three minor languages is small. The major languages demonstrate a greater diversity, and the largest distance is between Estonian and Finnish. It can be clearly seen that the distances between the analysed varieties do not obviously correlate with their genetic affiliation. Thus, we may conclude that a lexicostatistical analysis of the minimal depth (i.e. made for closely related languages) should not be seen as demonstrating a linear correlation with the genetic distance. Changes in the core lexicon happen due to different reasons including convergent processes that are not always transparent. In spite of the fact that the analysed Finnic varieties do not have obvious borrowings from each other, it is evident that the three minor varieties located in the compact area in Western Ingria are less diverse than geographically peripheral major languages.

4. Conclusions

The Swadesh lists for five Finnic varieties were compiled following an elaborated methodology that makes them transparent and discussable.

The difference between minor languages (Votic and two Ingrian dialects) is small: 94% or more of their core lexicon coincides. The major languages (Estonian and Finnish) demonstrate a greater difference both from minor languages (80–86%) and from each other (75%).

There are various reasons why the lexical diversity between languages increases: semantic shifts, the existence of synonymic pairs in the proto-language, new borrowings, and new derivatives, among other reasons.

The lexicostatic difference between closely related languages does not have a strong correlation with their genetic distance.

Acknowledgments

We are very grateful to our colleagues and native speakers of Finnic languages who we consulted in the course of our work on the wordlists, in particular, Alevtina Fedotova and Galina Samsonova on Soikkola Ingrian, Nikolai Pöder on Lower Luga Ingrian, Zinaida Saveljeva on Votic, Terhi Honkola on Finnish, Pärtel Lippus and Ellen Niit on Estonian.

We would like to thank the anonymous reviewer and Kirill Reshetnikov for the many valuable comments on the article.

The research of F. Rozhanskiy has been supported by the University of Tartu, grant PHVEE18904.

Appendix

It is obvious that Swadesh lists compiled by different researchers on the basis of different methods cannot be identical. However, a priori the degree of the diversity is not evident. For this reason, we give a short comment on the differences between the Swadesh lists for Estonian and Finnish compiled in the current article and those presented in Tillinger (2014). Tillinger's lists were chosen because they do not give synonyms and return exactly one word for each item (unlike the lists in Hofirková, Blažek 2012, and Syrjänen, Honkola, Korhonen, Lehtinen, Vesakoski, Wahlberg 2013).

For both Estonian and Finnish, we found four cases when we propose a word different from Tillinger's (2014), see Table 3.

Differences in versions of Swadesh lists for Estonian and Finnish

N	Meaning	Tillinger	Rozhanskiy, Zhivlov
Finnish			
3	bark	<i>kuori</i>	<i>kaarna</i>
5	big, large	<i>suuri</i>	<i>iso</i>
53	many, a lot of	<i>moni</i>	<i>paljon ~ monta</i>
76	skin	<i>nahka</i>	<i>iho</i>
Estonian			
10	bone	<i>kont</i>	<i>luu</i>
12	earth	<i>maa</i>	<i>muld</i>
86	that	<i>too</i>	<i>see</i>
106	snake	<i>madu</i>	<i>uss</i>

The reasons behind these differences are obvious: either our variant corresponds better to the context ('bark' and 'earth'), or it was chosen as more general and/or more neutral by a consultant ('big', 'skin', 'bone' and 'snake'). In case of 'many, lots of' we were not able to choose a single variant (but *monta* and *moni* have the same root); *too* 'that' looks more formal and is peculiar to written language so *see* 'this, that' was chosen as a more neutral variant.

In two cases, Tillinger (2014) does not have an exact correspondence to the words from our list. These are the items #12 *burn* (we use a transitive verb and Tillinger lists an intransitive verb) and #107 'thin' that is not mentioned by Tillinger.

Item #92 'warm' does not have an exact correspondence in Tillinger's Swadesh list but can be found in another wordlist (Tillinger 2014 : 183).

We conclude that in spite of the different methods of compiling the Swadesh lists, the differences between the versions do not look dramatic.

Addresses

Fedor Rozhanskiy
University of Tartu
Institute for Linguistic Studies of the Russian Academy of Sciences
E-mail: handarey@yahoo.com

Mikhail Zhivlov
Russian State University for the Humanities
National Research University Higher School of Economics
E-mail: zhivlov@gmail.com

Abbreviations

EVS — Eesti-vene sõnaraamat I–V, Tallinn 1997–2009; **GLD** — <http://starling.rinet.ru/new100/main.htm>; **LÄGLOS I** — A. D. Kylstra, S.-L. Hahmo, T. Hofstra, O. Nikkilä, Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen. Bd. I: A–J, Amsterdam–Atlanta 1991; **LÄGLOS II** — A. D. Kylstra, S.-L. Hahmo, T. Hofstra, O. Nikkilä, Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen. Bd. II: K–O, Amsterdam–Atlanta 1996; **LÄGLOS III** — Kylstra, A. D., Hahmo, S.-L., Hofstra, T., Nikkilä, O., Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen. Bd. III: P–Ä, Amsterdam–New York 2012; **VKS** — Vadja keele sõnaraamat. Toimetanud S. Grünberg, Tallinn 2013; **БФРС** — И. Вахрос, А. Щербakov, Большой финско-русский словарь, Москва 2007; **НБРФС** — М. Э. Куусинен, В. М. Оллыккайнен, Ю. Э. Сюръялайнен, Новый большой русско-финский словарь: в 2-х томах, Москва 1999.

REFERENCES

- Aikio, A. 2000, Suomen *kauka*. — Vir. 104, 612–614.
 — — 2002, New and Old Samoyed Etymologies. — FUF 57, 9–57.
 — — 2015, The Finnic ‘secondary e-stems’ and Proto-Uralic Vocalism. — JSFOu 95, 25–66.
- Chang, W., Cathcart, C., Hall, D., Garrett, A. 2015, Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis. — Language 91 (1), 194–244.
- Ernits, E. 2005, Vadja keele varasemast murdeliigendusest ja hilisemast hääbumisest. — Piirikultuuri ja -keeleg. Konvörents Kurgjärvel, 21.–23. rehekuu 2004, Võro, 76–90.
- Heinsoo, H. 2015, Vad’da sõnakopittõja, Tartu–Helsinki.
 — — 2018, Suuri päive, Tartu.
- Hofírková, L., Blažek, V. 2012, Ke klasifikaci ugrofinských jazyků. — Linguistica Brunensia 60 (1/2), 87–126.
- Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., Bakker, D. 2008, Explorations in Automated Language Classification. — Folia Linguistica 42, 331–354.
- Honkola, T. 2016, Macro- and Microevolution of Languages: Exploring Linguistic Divergence with Approaches from Evolutionary Biology, Turku (Turun Yliopiston Julkaisuja — Annales Universitatis Turkuensis. Ser. A II osa — tom. 311. Biologica — Geographica — Geologica).
- Janhunen, J. 1981, Uralilaisen kantakielen sanastosta. — JSFOu 77, 219–274.
- Kassian, A., Starostin, G., Dybo, A., Chernov, V. 2010, The Swadesh Wordlist. An Attempt at Semantic Specification. — Journal of Language Relationship 4, 46–89.
- Koivulehto, J. 1983, Suomalaisten maahanmuutto indoeurooppalaisten lainasanojen valossa. — JSFOu 78, 107–132.
 — — 2008, Frühe slawisch-finnische Kontakte. — Evidence and Counter-Evidence. Essays in Honour of Frederik Kortlandt. Vol. 1. Balto-Slavic and Indo-European Linguistics, Amsterdam–New York, 309–321.
- Laanest, A. 1982, Einführung in die ostseefinnischen Sprachen, Hamburg.
- Luobal Sámmol Sámmol Ánte (Aikio, A.) 2014, Studies in Uralic Etymology II: Finnic Etymologies. — LU L, 1–19.
- Markus, E., Rozhanskiy, F. 2012, Votic or Ingrian. New Evidence on the Kukkuzi Variety. — Finnisch-Ugrische Mitteilungen 35, 77–95.
- Nirvi, R. E. 1971, Inkeröismurteiden sanakirja, Helsinki (LSFU XVIII).
- Rozhanskiy, F., Markus, E. 2014, Lower Luga Ingrian as a Convergent Language. — On the Border of Language and Dialect. FINKA Symposium. University of Eastern Finland, Joensuu, 4–6 June, 2014, Joensuu, 36–37.
 — — 2015, Dialectal Variation in Votic: Jögöperä vs. Luuditsa. — ESUKA 6 (1), 23–39.
- Saarikivi, J. 2006, Substrata Uralica. Studies on Finno-Ugrian Substrate in Northern Russian Dialects, Tartu.
 — — 2009, Itämerensuomalais-slaavilaisten kontaktien tutkimuksen nykytilasta. — The Quasiquicentennial of the Finno-Ugrian Society, Helsinki (MSFOu 258), 109–160.
- Suhonen, S. 1985, Wotisch oder Ingrisch? — Dialectologia Uralica. Materialien des ersten Internationalen Symposions zur Dialektologie der uralischen Sprachen 4–7. September 1984 in Hamburg, Wiesbaden, 139–148.
- Swadesh, M. 1952, Lexicostatistic Dating of Prehistoric Ethnic Contacts. — Proceedings of the American Philosophical Society 96, 452–463.
 — — 1971, The Origin and Diversification of Language, Chicago.
- Syrjänen, K., Honkola, T., Korhonen, K., Lehtinen, J., Vesakoski, O., Wahlberg, N. 2013, Shedding More Light on Language Classification Using Basic Vocabularies and Phylogenetic Methods. — Diachronica 30 (3), 323–352.
- Taagepera, R. 1994, The Linguistic Distances between Uralic Languages. — LU XXX, 161–167.

- Tadmor, U. 2009, Loanwords in the World's Languages: Findings and Results. — Loanwords in the World's Languages. A Comparative Handbook, Berlin, 55–75.
- Tillingier, G. 2014, Samiska ord för ord. Att mäta lexikalt avstånd mellan språk, Uppsala (Studia Uralica Upsaliensia 39).
- Tsvetkov, D. 1995, Vatjan kielen Joenperän murteen sanasto, Helsinki (LSFU XXV).
- Грунтов И. А., Мазо О. М. 2015, Классификация монгольских языков по лексикостатистическим данным. — Journal of Language Relationship 13 (3), 205–255.
- Лаанест А. 1966, Ижорские диалекты. Лингвогеографическое исследование, Таллин.
- 1993, Ижорский язык. — Языки мира. Уральские языки, Москва, 55–63.
- Маркус Е. Б., Рожанский Ф. И. 2017, Современный водский язык. Тексты и грамматический очерк. 2-е издание, исправленное и дополненное, Санкт-Петербург.
- Муслимов М., Кузнецова Е., Николаева Е., Гореликов А., Ефимов С., Ефимова Т. 2003, *Vad̄da kaazgad*. Водские сказки, Санкт-Петербург.
- Решетников К. Ю. 2011, Новые этимологии для прибалтийско-финских слов. — Урало-алтайские исследования 2 (5), 109–112.
- Рожанский Ф. И., Маркус Е. Б. 2013, Ижора Сойкинского полуострова: фрагменты социолингвистического анализа. — Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies IX (3), St. Petersburg, 261–298.

ФЕДОР РОЖАНСКИЙ (Гарту—Санкт-Петербург), *МИХАИЛ ЖИВЛОВ* (Москва)

**БАЗОВАЯ ВОДСКАЯ И ИЖОРСКАЯ ЛЕКСИКА
НА ПРИБАЛТИЙСКО-ФИНСКОМ ФОНЕ:
СПИСКИ СВОДЕША ПЯТИ РОДСТВЕННЫХ ИДИОМОВ**

Основной целью статьи является построение списков Сводеша для пяти прибалтийско-финских идиомов: водского, эстонского и финского языков, а также сойкинского и нижнелужского диалектов ижорского языка. Принципиальное внимание уделяется методике построения списков: это применяемая в Московской школе компаративистики методика, при которой семантика включаемого в список слова задается не только переводом, но и контекстом употребления. Слова для списка отбирались во взаимодействии с носителями языка, что позволило находить наиболее подходящее слово из существующих в языке синонимов. Составленные таким образом 111-словные списки сопровождаются этимологическими комментариями к каждому слову. В статье делаются и предварительные наблюдения, касающиеся базового лексического состава рассматриваемых языков, в частности, исследуются сходства и различия лексических списков и рассматриваются механизмы проникновения в них новых слов. Одним из выводов оказывается отсутствие линейной зависимости между схожестью лексических списков и генетической близостью в случае близкородственных языков. Так, между малыми языками наблюдается больше сходства, чем при сравнении малого языка с финским или эстонским, у которых базовая лексика содержит больше конкретно-языковой специфики.