Early contacts between Uralic and Yukaghir

1. Introduction

In this article I shall deal with the location of Pre-Proto-Uralic, a well-known interest of Professor Juha Janhunen both from a linguistic (Janhunen 2000; 2007b) and origin-orientated (Janhunen 2001; 2007a) point of view. The following survey is based on the article “Zu den uralisch-jukagirischen Sprachkontakten” by Károly Rédei (1999). Rédei lists 121 words common to Uralic and Yukaghir and argues that these common words are borrowings from Uralic into Yukaghir. He assigns the loanwords to three strata by their distribution: 1. Uralic words borrowed from Samoyed, 2. Finno-Ugrian words borrowed from Ugric and 3. Samoyed words.

There are a total of 121 words in Rédei’s analysis; he has already dropped some less plausible correspondences presented in the manuscript of Nikolaeva (Rédei 1999). As my intention is to go back to the very first contacts between the two language families, I first omitted all words which have cognates only in Samoyed (38 words). These words require closer scrutiny in some future article. I have chosen for a closer analysis the remaining 83 words with a wide distribution in Uralic, after which I have omitted all the words with uncertain intra-Uralic cognates, too narrow distribution, implausible semantics or unique vowel correspondences, 30 altogether.

The remainder, 53 words plus 3 additional words, I have separated into strata on the basis of phonology: both consonants and vowels consistently attest two separate layers, the older consisting of 32 words and the more recent of 24 words. This stratification is based on the phonological criteria presented elsewhere (Häkkinen, forthcoming). Finally, I will consider the time and place of the Uralic-Yukaghir contacts and Pre-Proto-Uralic. The wordlist can be downloaded at the web address <http://www.mv.helsinki.fi/home/jphakkin/UralicYukaghirWordlist.pdf>.
2. Uralic loanword strata in Yukaghir

Let us first examine the consonant correspondences. In the loanword strata Proto-Uralic *s has two different correspondences, Yukaghir Ø (loss) and *l. Károly Rédei proposed that the first reflex could represent borrowing from Pre-Proto-Samoyed *ϑ (IPA [θ]) at the stage between Proto-Uralic *s and Proto-Samoyed *t (Rédei 1999: 26–27).

However, there are arguments against this kind of substitution. There are Tungusic loanwords in which we also find a correspondence T *s ~ Y Ø (Nikolaeva 2006: 66–67). Because there are no reasons to assume *ϑ in Tungusic (*s there has remained unchanged), we must assume that at some point in Pre-Proto-Yukaghir the change *s > Ø occurred. Therefore, the most credible option is to assume that the Uralic *s was also substituted by a Pre-Proto-Yukaghir *s which was later lost (U *s → PreY *s > Y Ø). The Yukaghir layer of borrowings with Ø would then represent a very old Uralic stage, because *s changed to *l (a voiceless lateral fricative, IPA l̺) already in the East Uralic protodialect (> Hungarian Ø ~ Mansi *t ~ Khanty *L ~ Samoyed *t) (Häkkinen 2007: 71–73).

An additional argument is that only East Uralic *L explains *l, the second Yukaghir reflex for Uralic *s. In Tungusic loanwords there is no correspondence between T *s and Y *l, which also points to a conclusion that *l was not a substitute for original *s but rather for some other sound which Tungusic did not possess, namely East Uralic *L.

1. U *s → EPY *s > MPY *s > LPY Ø
2. IU *L → MPY *l > LPY *l
3. Sy. *t → LPY *t?

There are three possible correspondences and their relative chronology, although there are no words in which Samoyed *t < Proto-Uralic *s would correspond with Yukaghir *t. This may be due to chance, or it can be explained by assuming that Proto-Samoyed was the receiving part in the contact, or that it was in contact with some Para-Yukaghir language rather than Yukaghir itself, which at this point was probably already spoken further east. In the latter case there would be no Proto-Samoyed loanwords in Yukaghir proper.

Vowel correspondences also testify for this kind of relative chronology. Those words which by the *s-criterion are the oldest (Ø-words) seem to have different vowel correspondences than those words which are younger (*l-words):

1st layer:

<table>
<thead>
<tr>
<th>U *s</th>
<th>~ Y Ø</th>
</tr>
</thead>
<tbody>
<tr>
<td>U *sala- ‘to steal’</td>
<td>~ Y *olo- ‘to steal’</td>
</tr>
<tr>
<td>U *sala- ‘to load’</td>
<td>~ Y *el- ‘to load’</td>
</tr>
<tr>
<td>U *s/šoja ‘sleeve, wing’</td>
<td>~ Y *uuje ‘wing’</td>
</tr>
<tr>
<td>U *sula- ‘to melt’</td>
<td>~ Y *alaa- ‘to melt’</td>
</tr>
</tbody>
</table>
Early contacts between Uralic and Yukaghir

2nd layer:
U *särä ‘fiber, root’ ~ Y *larq- ‘root’
U *sevi- ‘to eat’ ~ Y *leγ- ‘to eat’
U *saxi- ‘to come’ ~ Y *laq- ‘to come’

In the first layer the vowels are of different height, while in the second layer they are of the same height or even identical. Consequently, we may suppose that even in those words, which do not have consonantal evidence to support the stratification, the words with similar vowel correspondences are younger than the words with different vowel correspondences. However, details concerning the phonology of the contact languages and the patterns of substitution must be left to another article.

3. Uralic-Yukaghir contacts in time and space

56 old Uralic loanwords of high quality in Yukaghir, assigned to two different strata, are sufficient to prove mutual contacts. When and where did the Uralic-Yukaghir contacts take place? Yukaghir is spoken in Eastern Siberia, while the easternmost Uralic languages have only recently crossed the Yenisei from Western to Central Siberia: Nganasan and Enets in the north and Sayan Samoyed languages in the south. Nearly 2 000 km – almost the whole of Central Siberia – lie between the two language families. However, Yukaghir can be derived from the west, as it was spoken earlier near the Lena. We may assume that Yukaghir at some point in the past migrated down the Lena, just as Yakut did later, and that Early Proto-Yukaghir was spoken somewhere near the Upper Lena and the region of Lake Baikal, the watershed area between the Lena and Yenisei river systems.

However, the further back in time we go, the further west the easternmost Uralic language moves. 2 000 years ago Proto-Samoyed was spoken near the Sayan area, but 4 000 years ago Proto-Uralic was spoken in Europe, west of the Urals (Kallio 2006; Häkkinen 2009). The geographical distance seems to be smallest between East Uralic (already spread close to the Yenisei) and Middle Proto-Yukaghir (still spoken near the watershed area). Even though the East Uralic developments without any reasonable doubt occurred on the European side of the Urals, for Mansi and Khanty were still recently spoken west of the Urals, the East Uralic dialect seems to have spread to near the Sayan region before East Uralic *L changed into Samoyed *t (Häkkinen 2009: 49–50).

Unfortunately, there are also Tungusic loanwords in Yukaghir which seem to have been borrowed at this stage (T *s ~ Y Ø; see Nikolaeva 2006: 67). Tungusic contacts pull the Yukaghir area eastward, as Tungusic is considered the easternmost of the language families representing the Altaic typology; Juha Janhunen locates Proto-Tungusic in Manchuria (Janhunen 1996: 216). We could try
to explain this contradiction by assuming that Middle Proto-Yukaghir moved after the East Uralic contacts from the Upper Lena to the Middle Lena bend, where contact with Proto-Tungusic could have soon begun.

The greatest problem concerns the older, Proto-Uralic loanword layer. As Proto-Uralic itself was spoken on the European side of the Urals (Kallio 2006; Häkkinen 2009), it cannot have had any direct contacts with Yukaghir. Furthermore, as the East Uralic changes occurred already on the European side of the Urals, it is not possible to assume that loanwords belonging to the oldest layer were borrowed from Pre-East Uralic (= the form of Proto-Uralic that spread to Siberia). We therefore have only three possibilities: either the older layer represents vocabulary inherited from a common Uralo-Yukaghiric protolanguage, or it is connected to Pre-Proto-Uralic or to Para-Uralic.

There are many arguments against the first of these possibilities. First of all, the morphological features common with Yukaghir are shared only by Samoyed, not the other Uralic languages (Rédei 1999: 12–13). As the Samoyed branch no longer seems to have split first and directly from Proto-Uralic, but instead split from an East Uralic dialect together with the Ugric branches (Häkkinen 2007; 2009: 11–16), we cannot reconstruct these features even for Proto-Uralic; we have no other choice than to consider these common Samoyed-Yukaghir features contact-induced.

Secondly, excluding sibilant correspondences, Uralo-Yukaghir cognates are amazingly similar looking, compared to the very different morphological endings. It does not seem very probable that the languages could have developed from a common protolanguage into such different directions morphologically but still retained the inherited words almost unchanged. Therefore, it seems more probable that the cognates reflect mutual contacts, not inheritance from a common protolanguage.

The term Para-Uralic denotes a sister language to Proto-Uralic, descended from the same ancestor (Pre-Proto-Uralic). It is theoretically possible that there was such an eastern language which retained the Pre-Proto-Uralic *s and that this language could be responsible for the loanwords of Proto-Uralic appearance in Yukaghir. However, so far there are absolutely no traces that any such language ever existed. Furthermore, such an eastern Para-Uralic language would also require that Pre-Proto-Uralic was spoken further east than Proto-Uralic, so this hypothesis does not provide any advantage over the Pre-Proto-Uralic hypothesis.

Pre-Proto-Uralic is the language stage preceding Proto-Uralic. If Pre-Proto-Uralic was spoken further east than Proto-Uralic, perhaps near the watershed area between the Yenisei and Lena, it could be responsible for the loanwords of Proto-Uralic appearance in Yukaghir. I shall now survey the evidence for an eastern or western origin for Pre-Proto-Uralic.
Pre-Proto-Uralic – from the west or east?

Asko Parpola has recently presented a convincing sketch in which he connects the stages before and after Proto-Uralic to the archaeological record (Parpola, forthcoming). Parpola suggests that Pre-Proto-Uralic was spoken in the Upper Volga region and that its eastward expansion reached the Kama region only during the Volosovo culture, when Volosovo influence can be detected in the Garino-Bor (Turbin) culture: “The Cembra pine and the Siberian fir grow in this area, which had been in contact with the Lyalovo culture already, but now became part of the Volosovo culture, and was particularly important as its source of metal.”

The names for the Cembra pine (*Pinus cembra: U *seksi*) and Siberian fir (*Abies sibirica: U *ńulka*) can be reconstructed for Proto-Uralic, although they have no cognates in West Uralic (Häkkinen 2009: 35–36). The words could have disappeared from western dialects in the area where the trees were not known, but it is equally possible that Proto-Uralic had already split into daughter dialects at the time when these words entered the language. Consequently, it may well be that the names for the two eastern conifers spread westward from the Garino-Bor culture, just as Parpola suggests.

However, on the basis of the present study I propose that not only the names of these trees, but also the Proto-Uralic language as a whole, spread westward from the Garino-Bor culture. Even though the archaeologically perceivable influence at this stage seems to go eastward, it could possibly be explained as a “gift in return”, because the rich metal sources probably gave major prestige to the Garino-Bor people, even though they were connected to the Volosovo culture. Wealth and prestige were precisely what the Proto-Uralic speakers possessed, as can be interpreted from the concepts for which they borrowed words from the Aryans: U *asira ‘lord, prince’ ← A *asura; U *serńa ‘gold’ ← Ir., cf. Avestan zaranya; U *šeća ‘100’ ← A *ćata- or IA *šata-; U *šosra ‘1000’ ← A *žasra (Häkkinen 2009: 20–25), together with the terms for metal U *wāśka ‘copper/bronze’ and *āsa-wāśka ‘tin/lead’ which still lack an etymology (Häkkinen 2009: 25–28). The Proto-Samoyed cognate *wesä (< U *wāśka) seems to have been borrowed further into Proto-Tocharian: Tocharian A wās ~ B yasa ‘gold’ (Kallio 2004).

The compound word *āsa-wāśka proves that Proto-Uralic speakers knew the role of tin/lead in bronze metallurgy and could thus produce valuable bronze artifacts. As the Proto-Uralic speakers knew bronze metallurgy and valuable metals, it seems to me that the people of the Garino-Bor culture, who possessed a source of metal, fit the criteria for Proto-Uralic better than the people of the Volosovo culture; and as metals were the key to wealth and prestige, it seems to me that people of the Garino-Bor culture, who possessed the metal sources, had a much greater possibility to gain wealth and prestige than people of the Volosovo culture. Therefore, a language of the Garino-Bor culture seems more suitable and probable to spread than a language of the Volosovo culture.
Consequently, I find it more probable to connect Early Proto-Uralic to the Garino-Bor culture than to the Volosovo culture, although Late Proto-Uralic may well be connected to the Volosovo culture as well; cultural continuity may remain the same while the language spoken shifts to another. Prestige produced by wealth in metals would be a motive for the expansion of Proto-Uralic, and the widespread Volosovo culture, which absorbed the Garino-Bor culture, would be the area in which Proto-Uralic expansion took place, at the expense of earlier language(s) of the cultural area.

Of course this kind of reasoning alone cannot prove that Pre-Proto-Uralic did not spread to the Kama area from the west. However, there are other arguments either against a western origin or for an eastern origin.

Why was Pre-Proto-Uralic not spoken in the west?

The first argument is that there is a high probability that other languages were spoken west of the Volga-Kama fork. In the Finno-Permic vocabulary there appear words with consonant clusters consisting of a sibilant followed by a nasal, alien to Proto-Uralic phonotactics: *šm, *šn, *šń. Further west (judging from a narrower distribution in Finno-Mordvin or Finno-Saamic) there appear yet more words of this kind (*šn, *šm), and words with other foreign features, such as geminate nasals (*mm, *nn) and word-final *s (Häkkinen 2009: 37–40, 47–48).

The number of foreign phonotactic features in vocabulary grows as the Uralic dialects advance westward from the Volga-Kama fork. This would not be the case if only languages related to Uralic were spoken in the west. Of course it is possible that both languages related to Uralic and unrelated languages were spoken; after all, there were many different cultures in the same area (Volosovo, Fatyanovo-Balanovo, Abashevo). However, as all the other cultures except the Volosovo are connected to Indo-European-speaking groups by Parpola, and as the words borrowed into Uralic with alien phonotactics are not of Indo-European origin, we should leave the Volosovo culture for these unknown Palaeo-European languages.

A second argument is a lack of Early Proto-Indo-European loanwords in Uralic (Kallio 2006: 10). However, there seem to have been contacts between Early Proto-Uralic and Late Proto-Indo-European (or some form of archaic Indo-European) from at least 3000 BC on, as the Indo-European laryngeals have two different reflexes in Proto-Uralic (here only one eastern and one western language are presented):

1. IE *bhēh₁-(ye/o-)* ‘to cook’ → U *pexi- ‘to cook’ > Samoyed *pi- ~ Erzya pije- (Koivulehto 1991: 55)
2. IE *bhēh₁-(ye/o-)* ‘to cook’ → U *peša- ‘to cook’ > North Saami bassi- ~ Mansi *piit- (Koivulehto 1991: 85)
Early contacts between Uralic and Yukaghir

I have earlier argued that the Uralic *x*-substitution is older and can be connected to Early Proto-Uralic, when *x* was still something like a voiceless velar spirant /χ/, while in Late Proto-Uralic it would already have changed intervocally into voiced *γ*, and therefore another voiceless fricative, namely the sibilant *š*, was used to substitute for the Indo-European laryngeal (Häkkinen 2007: 42–44). As the new dating for Late Proto-Uralic suggests a time around 2000 BC (Kalilio 2006; Häkkinen 2009), even Early Proto-Uralic seems to be later than Late Proto-Indo-European. Consequently we have no reason to assume the beginning of contacts between Uralic and Indo-European before 3000 BC, and even the oldest Indo-European loanwords would be later than Late Proto-Indo-European.

Archaeologically we find penetration of intrusive cultures into the southern taiga zone close to the Volga–Kama fork: the Fatyanovo-Balanovo culture from the west (connected to arcaic Northwest Indo-European) and the Poltavka Culture from the south (probably connected to Pre- or Early Proto-Aryan), both about 2800 BC (Carpelan & Parpola 2001). From this point on we can explain all the old Indo-European loanword layers in Uralic (modified from Häkkinen 2009: 42):

<table>
<thead>
<tr>
<th>Stage</th>
<th>Period</th>
<th>Borrowings into Early Proto-Uralic</th>
<th>Borrowings into Late Proto-Uralic</th>
</tr>
</thead>
<tbody>
<tr>
<td>I stage</td>
<td>2800 BC</td>
<td>Northwest IE borrowings into Early Proto-Uralic: <em>bheh₁-(ye/o-)</em> → U <em>pexi-</em> ‘to cook’</td>
<td>Pre-Proto-Aryan borrowings into Early Proto-Uralic: PreA <em>gugb'ew-</em> → U <em>juxi-</em> ‘to drink’</td>
</tr>
<tr>
<td>II stage</td>
<td>2300 BC</td>
<td>Northwest IE borrowings into Late Proto-Uralic: <em>bheh₁-(ye/o-)</em> → U <em>peša-</em> ‘to cook’</td>
<td>Proto-Aryan borrowings into Late Proto-Uralic: A <em>argha-</em> → U <em>arwa</em> ‘value, price’</td>
</tr>
<tr>
<td>III stage</td>
<td>1800 BC</td>
<td>Northwest IE borrowings into Uralic dialects: <em>bheh₁-tó-</em> → WU <em>lešti</em> ‘leaf, blade’</td>
<td>Iranian borrowings into Uralic dialects: (No examples of similar source consonants)</td>
</tr>
</tbody>
</table>

The criterion here is the presence or absence of the Early Proto-Uralic *x* in the loanword layers: the layers in which it appears are older, and the layers in which it does not appear are younger. The Pre-Proto-Aryan loanwords also have *x* as a substitute for *gʰ*, but Proto-Aryan *gʰ* is no longer substituted by *x* but *w*, so we can consider these layers contemporaneous with the respective Northwest Indo-European layers. Northwest Indo-European seems to have remained quite archaic up to the second millennium BC, at least at its northern end (Mallory 2001: 354; Kallio 2006: 11).

This all seems to suggest that Pre-Proto-Uralic was not spoken anywhere near Early Proto-Indo-European – it is not even necessary to locate it anywhere near Late Proto-Indo-European. As most scholars hold that the Indo-European Urheimat was located in or near the Pontic-Caspian steppes (Mallory & Adams 2006: 442–), it would be justifiable to locate Pre-Proto-Uralic somewhere else and not in the vicinity – the Upper Volga seems to be too close to the Pontic-
Caspian steppes, considering the lack of Early Proto-Indo-European loanwords in Uralic. Of course it would still be possible that Pre-Proto-Uralic was spoken in the forest zone of Northeastern Europe if Early Proto-Indo-European were spoken in Caucasus or Anatolia (the contacts with Uralic would no longer link Proto-Indo-European to the steppes), so this particular piece of evidence cannot be seen as totally exclusive. All things considered, however, the further from the Pontic-Caspian steppes we locate Pre-Proto-Uralic, the easier it is to explain the lack of Early (and possibly even Late) Proto-Indo-European loanwords in Uralic.

Why was Pre-Proto-Uralic spoken in the east?

Juha Janhunen has repeatedly argued that the Ural-Altaic typological complex is an areally distinct unit with clear-cut boundaries in every direction against languages of different typology, and that the Ural-Altaic typology must have had one original centre of expansion. As the earliest protolanguages of the Altaic language families (Proto-Turkic, Proto-Mongolic and Proto-Tungusic) can be traced back to Greater Manchuria (up to Mongolia; Janhunen 1996: 216), this view requires that Pre-Proto-Uralic must also have been present somewhere adjacent to them. According to Janhunen there is no significant chance that Pre-Proto-Uralic could have developed a structural typology so similar to the Altaic languages without being in close contact with them. (Janhunen 2001; 2007a.)

The second argument has been scrutinized in the present article: the oldest layer of Uralic loanwords in Yukaghir represents a reconstruction level similar to Proto-Uralic. However, Proto-Uralic itself was spoken too far west for it to be responsible for this loanword layer. The East Uralic layer is the second oldest layer in Yukaghir, so it cannot explain the oldest layer. Neither can Pre-East Uralic explain it, because East Uralic changes seem to have occurred on the European side of the Urals, and only East Uralic proper spread to Siberia. Some lost archaic Uralic or Para-Uralic language could explain the layer, but there are no traces that any such language ever existed, and it would nevertheless require that Pre-Proto-Uralic was spoken further east. Because Pre-Proto-Uralic is already a logical necessity (every protolanguage has predecessors), it is more economical to explain the layer of loanwords of Proto-Uralic appearance by assuming that Pre-Proto-Uralic is responsible for this layer – there is no reason to conjecture a Para-Uralic.

Why was Pre-Proto-Uralic spoken in the west?

Asko Parpola (forthcoming) presents two arguments for this option. The first is the conservativeness of Finnic, which could point to a situation in which the language was spread at the expense of related languages. However, there are after all traces of unrelated languages in the Uralic west (see above).
His second argument is archaeological plausibility: all Uralic languages can be explained by the Volosovo culture, which has its roots in the Upper Volga region. This argument cannot be seen as absolute either, because retrospective archaeological method cannot always find a real-world cognate for linguistic expansion (Mallory & Adams 2006: 451–452). Therefore, it is possible that none of the most plausible options is actually responsible for certain linguistic expansion.

Another point related to this argument is the phenomenon that usually the language of a community shifts in spite of a continuity that we can follow back in time archaeologically; this is due to the fact that archaeological cultures can have – and usually do have – many roots, any of which may or may not be connected to a certain language.

Furthermore, there are other archaeologically plausible options as well: the expansion of Proto-Uralic could be connected to the Sejma-Turbino network, or a chain consisting of the Chirkovo, Kama and Krotovo cultures, or something else entirely (Häkkinen 2009: 49–50). The Uralic expansion has earlier been explained archaeologically in a different way (Fodor 1976).

## Conclusion

I agree with Asko Parpola on post-Proto-Uralic matters, but I conclude that the evidence as a whole points to an eastern rather than western origin for Pre-Proto-Uralic. Consequently, on the basis of the earliest, seemingly Proto-Uralic (and definitely not post-Proto-Uralic) loanword layer in Yukaghir, I can see no other choice than to agree with Juha Janhunen: Pre-Proto-Uralic must have been spoken further east, probably somewhere in the Sayan region near both Mongolia (due to a common Ural-Altaic typology) and the watershed area between the Yenisei and the Lena (due to the oldest layer of Uralic loanwords in Yukaghir), possibly as recently as 3000 BC. From here the language spread/moved westward to the European side of the Urals, where Early Proto-Uralic came into contact with archaic Northwest Indo-European and Pre-Proto-Aryan.

## Appendix

<table>
<thead>
<tr>
<th>PreY = Pre-Proto-Yukaghir</th>
<th>S = Proto-Samoyed</th>
</tr>
</thead>
<tbody>
<tr>
<td>(all stages before Y)</td>
<td></td>
</tr>
<tr>
<td>EY = Early Proto-Yukaghir</td>
<td>T = Proto-Tungusic</td>
</tr>
<tr>
<td>MY = Middle Proto-Yukaghir</td>
<td>PreA= Pre-Proto-Aryan</td>
</tr>
<tr>
<td>Y = Late Proto-Yukaghir</td>
<td>A = Proto-Aryan</td>
</tr>
<tr>
<td>U = (Pre-)Proto-Uralic</td>
<td>IA = Indo-Aryan</td>
</tr>
<tr>
<td>WU = West Uralic</td>
<td>Ir. = (Proto-)Iranian</td>
</tr>
<tr>
<td>EU = East Uralic</td>
<td>&gt; = was developed into</td>
</tr>
<tr>
<td></td>
<td>→ = was borrowed as</td>
</tr>
</tbody>
</table>
Literature


— (forthcoming): Phonology of Pre-Proto-Yukaghir.


