

Article

The geohistorical stratification of Uralic-Indo-European contacts

In light of some Tocharian loanwords in Mordvin

Arnaud Fournet

Abstract: During their movements and splits Uralic languages, and foremost those of the Finnic and Volgaic groups which have the westernmost and southwesternmost locations, got in contact with different branches of the Indo-European family. This issue has been at the center of Uralic etymological and historical studies from the start. The paper first presents the principles for the stratification of Uralic-Indo-European loanwords and then examines a set of words that exhibit a very peculiar sound correspondence in the initial syllable between Finnic *va* ~ Mokša *u* and Erzia *vi*. These words can be shown to be loanwords of Tocharian origin. These words conflict with the usual paradigm that only the most widespread Indo-European loanwords into the Uralic family could have a Tocharian origin. In addition they offer new insight on the original location of proto-Tocharian.

Keywords: Uralic, Proto-Indo-European, Tocharian, Mordvin.

1. Introduction

Uralic languages bear ample testimony to intense and lasting contacts with Indo-European languages. The antiquity of the first potential loanwords is so high that even the status of some widespread Uralic words as cognates with Proto-Indo-European (PIE) or loanwords by a compact Proto-Uralic (PU) community is controversial. The PU word **wete* ‘water’ is a typical instance and it is quite difficult to assert with any real certainty whether it is a loanword or a cognate. There is no agreement on this issue and it can be noted that some Uralicists, among whom Björn Collinder, actually changed their minds shifting from cognate status to loanword status back to cognate status.

If many words in Uralic languages can be traced back to PIE roots and words attested in Indo-European languages, only a handful is shared by about all Uralic languages, from Samoyedic and Ugric in the east to Finnic and Saamic in the west. Most are limited to a particular branch of Uralic or even to a specific language or dialect. On the whole they are telltale clues about the scope, the nature and the duration of the contacts between Uralic and Indo-European languages, branches and possibly even their respective proto-languages. An example of widespread loanwords is UEW560 **waške* ‘metal (iron, copper, gold, etc.)’ which is supposed to reflect the Proto-Tocharian form underlying A *wās* and B *yasa* ‘gold’. An example of loanwords with a more limited extension is UEW815 **wašara* ‘axe, hammer’ present in Finno-Volgaic and supposed to reflect Indo-Iranian **važra*. It can be noted that the Mokša-Mordvin word *kšni* ‘iron’ even follows the sound changes of Avestic *kšni* < **kṛstni* ‘the black metal (iron)’, as compared to Mari-Cheremis *kərtñi-waž* ‘iron ore’.

During their westward movements and splits Uralic languages, and foremost those of the Finnic and Volgaic groups which have the westernmost and southwesternmost locations, got in contact with different branches of the Indo-European family. This issue has been at the center of Uralic etymological and historical studies from the start. The paper first presents the principles of the stratification of Uralic-Indo-European loanwords and then examines a set of words that exhibit a very peculiar sound correspondence in the initial syllable between Finnic *va* ~ Mokša *u* and Erzia *vi*. These

words can be shown to be loanwords of Tocharian origin and this sound correspondence probably originates in a Proto-Tocharian syllable like *wə with a schwa. This situation has a number of consequences for the general theory of Uralic-Indo-European loanwords and it sheds light on the location of Proto-Tocharian before it moved to Chinese Turkestan and to a lesser extent on its early phonology.

2. Principles for the stratification of Uralic-Indo-European loanwords

A historiographical survey of works on Uralic-Indo-European loanwords from the beginning of the 19th century until the 1960s can be found in Joki (1973). It is summarized by Rédei (1988:639-646) who concludes: Joki (1973) “ist bisher die ausgiebigste Fundgrube für die Erforschung der uralisch-indo-germanischen Berührungen”¹. Generally speaking the field has seen a change in its general tone and perspective. Originally it was often surmised a close genetic relationship (“Urverwandschaft”) of Uralic and Proto-Indo-European. This was gradually replaced by a more critical and cautious point of view that deals with waves of borrowings (“Entlehnungen”) caused by contacts (“Berührungen, Kontakten”) between speakers of Uralic and Indo-European languages. The (pre)historical scenario proposed by Rédei (1986 & 1988) distinguishes three groups of words:

- A first group is pan-Uralic and therefore potentially the most ancient. Rédei (1988:648) admits that four of them can be Tocharian: “dass bei vier Wörtern auch die Abstammung aus dem Tocharischen möglich sein kann”, among which UEW560 *waške ‘metal (iron, copper, gold, etc.)’,
- Another group is Finno-Ugric. Rédei describes it as Indo-Iranian (“urarische”) or Pre-Indo-Iranian (“vorarische oder frühurarische”), according to phonetic criteria coherent with the historical phonology of this branch,
- A third group is Finno-Volgaic or Finno-Permian. Rédei (1988:649) distinguishes two layers of Indo-Iranian and Iranian origin and dating.

In that scenario, or so to speak this model, accounting for contacts between Uralic and Indo-European languages and subbranches the large or restricted diffusion into Uralic is correlated and synchronous with a period of interactions with a specific Indo-European language or subbranch:

- Ancient period ~ Tocharian ~ pan-Uralic,
- Intermediary period ~ (Pre-)Indo-Iranian ~ Finno-Ugric,
- Recent period ~ (Indo-)Iranian ~ Finno-Volgaic or Finno-Permian.

It can be underlined that the geo-historico-linguistic consistency of this scenario is very high. As a matter of fact up to this day most scenarios that describe the lexical borrowings into individual Uralic languages follow the same canonical pattern, which would also involve (North) Germanic in the case of Saami for example. It can also be noted that the scenario developed by Rédei entails a sequence of splits and diffusions of Uralic, or to be precise of Finno-Ugric, languages from the east to the west, an underlying premise that Rédei does not acknowledge explicitly. Besides, Rédei resorts to the word *layer* (“Schicht”) with archeological undertones as a synonym for *group of words*.

3. An example of Tocharian loanword UEW560 *waške ‘metal’

This word is traced back by the UEW to Proto-Tocharian *wäs ‘gold’, itself from PIE *H₂weseH₂ ‘gold’ and cognate with Latin *aurum*, according to Adams (1999). There is little doubt that this item cannot be an Indo-Iranian word as this meaning is not attested in that branch of Indo-European, as underlined by Rédei (1988:653): “Das PU-Wort kann nicht aus dem Vorarischen gekommen sein, weil das Wort in keiner arischen Sprache nachzuweisen ist.”². This word is attested throughout all Uralic:

¹ Joki (1973) “is up to now the most abundant source of information for the study of Uralic-Indo-European contacts.”

² “This PU word cannot come from Pre-Indo-Iranian as it is not attested in any Indo-Iranian language.”

- Finnish *vaski* (gen. *vasken*) ‘ore, copper, bronze’, Estonian *vask* (gen. *vaske*) ‘copper, brass’,
- Saami *voei’ke* (Northern dialect) ‘copper’,
- Erzia-Mordvin *uške, viškä* ‘wire, chain’, Mokša-Mordvin *uške* ‘wire, cable’.

This word displays a very unusual correspondence: *va ~ u ~ vi* between these languages for which there is no instance among Uralic cognates. It is unclear whether the apparent formative *-k-* should be considered a real suffix or the adaptation of $H_2 > k$.

- Mari (Cheremis) *kərtñi-waž* ‘iron ore’, *ši-waž* ‘silver ore’,
- Udmurt (Votyak) *az-veš* (S K G) ‘silver’,
- Komi (Zyrian) *ez-iš* (S P) ‘silver’,

The following Ugric forms are also interesting because they seem to have kept traces of $*H_2$ as γ or *k* either word-initially or finally. In that case, the loanword was $*H_2wäsH_2$ at the time of borrowing.

- Khanty (Ostyak) *way* (V), *wax* (DN), *ox* (O) ‘iron, metal, gold’,
- Mansi (Vogul) *ät-küš* (TJ), *ööt-wəs* (KO), *at-wəš* (P), *at-wəs* (So), *oat-khwes* (K) ‘lead’,
- Hungarian *vas* ‘iron’,
- Nenets (Yurak) *ješe* (O), *weše* (Lj.) ‘iron, gold’,
- Enets (Yen) *bese* ‘iron’,
- Nganasan (Tavgi) *bása* (gen. *baja*) ‘iron, metal’,
- Selkup *kezi* (Ta.), *kēzä* (Tur.), *kwez* (Ke.), *kwezi* (Ty.) ‘iron’,
- Kamass *baza, waza* ‘iron’,
- Koibal *bazě*; Motor *baze*; Taigi *beiše* ‘iron’.

Few Uralic words are so richly documented in about all varieties. As mentioned before the Mordvin and Balto-Finnic languages have an unusual sound correspondence in that word: *va ~ u ~ vi*. It can be compared with the sound correspondence attested in and Indo-Iranian word like $*vasa$ ‘calf’:

- Finnish *vasa*, Estonian *vasik* ‘calf’,
- Erzia-Mordvin *vazñe* ‘calf’, Mokša-Mordvin *vaz* ‘heifer’, *vazñe* ‘calf’

The word ‘calf’ is never attested as $**uzñe$ or $**vizñe$ in Mordvin. Two types of correspondences can therefore be identified:

- a sound correspondence of Tocharian origin: *va ~ u ~ vi* (< $*wə$)
- a sound correspondence of Indo-Iranian origin: *va ~ va ~ va* (< $*wa$).

The description of the vowel <ä> in Pinault (1992:37) has it that <ä> is a “voyelle fermée, analogue au *jer* (voyelle ultra brève) du slave (ĩ, ů), fonctionnant comme le schwa (e caduc du français)”³ or in a more recent work in Pinault (2008:414): “cette voyelle, transcrite <ä>, est une voyelle brève et fermée ; elle est le produit diachronique des voyelles fermées du système vocalique indo-européen ($*e, *i, *u$)”⁴. A similar point of view is provided in Fortson (2004:355): “It has been proposed that it represents the mid high vowel [i].” The Finno-Volgaic correspondence: *va ~ u ~ vi* is clearly coherent with the description of the vowel <ä> $*[ə]$ of Tocharian. This correspondence is attested in some other words: Mokša-Mordvin *узерь* [užəf] ‘ax’, Mokša-Mordvin *yumop* [uštər] ‘maple’, Mokša-Mordvin *укупей* [uskirej] ‘wagtail’, Mokša-Mordvin *укомс* [uskəms] ‘to bear, drag’ that will be examined below.

4. The words ‘ax’ and ‘maple’

These items have much fewer representatives in Uralic than the word ‘metal’. They are only Finno-Volgaic and do not exist in either Permic or Ugric:

³ “high vowel, similar to the *jer* (ultra short vowel) of Slavic (ĩ, ů), functioning like a *schwa* (French mutable *e*).”

⁴ “this vowel transcribed <ä> is short and high; it is the diachronic reflex of PIE high vowels ($*e, *i, *u$).”

- Finnish *vasara* ‘hammer’, *vaahtera* ‘maple’ (*Acer platanoides*), Estonian *vasar* ‘hammer’, *vaaher* ‘maple’,
- Saami *vaeččer* (N) ‘ax’,
- Erzia-Mordvin *užer’, vižir’* ‘ax’, *ukštor(o)* ‘maple’, Mokša-Mordvin *užər’* ‘ax’, *uštər* ‘maple’.

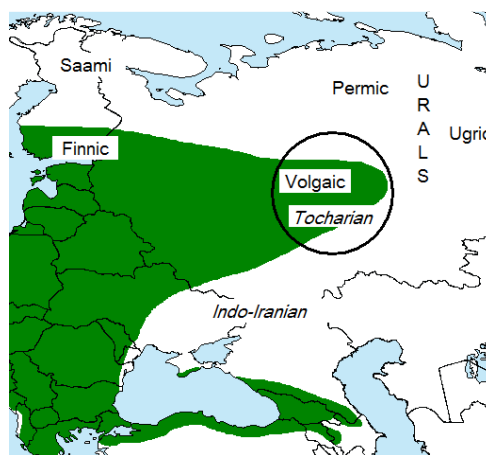
According to the UEW (815-816), the word ‘hammer, ax’ is a loanword of Indo-Iranian **vajra*, the well-known weapon of Indra. This traditional explanation is nevertheless highly doubtful because Erzia and Mokša cannot derive from a word containing **va* as shown by another Indo-Iranian loanword like **vasa* ‘calf’:

- Finnish *vasa* ~ *vaahtera* ~ *vasara* ~ *vaske*
- Estonian *vasik* ~ *vaaher* ~ *vasar* ~ *vask*
- Erzia-Mordvin *vazhe* ~ *ukštor(o)*⁵ ~ *užer’/vižir’* ~ *uške/viškä*
- Moksha-Mordvin *vazhe* ~ *uštər* ~ *užər’* ~ *uške*

These words owe their particular correspondences to their different origin: Proto-Tocharian instead of Indo-Iranian. In spite of their limited diffusion within Uralic, the words ‘ax’ and ‘maple’ display what is expected from a Tocharian initial **wä*: the proto-words can be reconstructed as **wäzär* and **wäkšter* when they were borrowed into Proto-Finno-Volgaic, as **wä* [wə] regularly becomes Mordvin [u] or [vi] but **wa* becomes Mordvin [va]. The Tocharian words **wäzär* and **wäkšter* are not attested in the vocabulary listed by Adams (1999). **Wäzär* may be indirectly attested in *wašir* ‘thunderbolt, diamond’ which Adams (1999) considers to be a Prakrit borrowing. **Wäkšter* ‘maple’ is not attested in Tocharian, as is logical because maples do not grow in areas to the east of Ural mountains and are mostly found in Europe up to the western bank of the Volga river. This feature probably also explains why it is not present in Permic and Ugric. It can be noted that a peculiarity of Tocharian is that plants do not grow, but augment: the Tocharian verb *auks-* ‘to sprout, grow up’ in Adams (1999:131) has several derivatives: AB *oko* ‘fruit’ and A *okar* ‘plant’ in the same semantic field, and Adams (1999:109) also lists two tree-names: *okaro* ‘*Acorus calamus* Linn’ according to Filliozat or possibly ‘*aloès, Aquilaia agallocha*’ according to Pinault. The proto-form underlying Finno-Volgaic words **wäkšter* ‘maple’ is based on **auks-ter* ‘that which grows’. This is a regular and well-attested word-formation. It is unclear why this word was specialized for naming maples.

This suggests that at some time in the past Proto-Tocharian must have been spoken in an area where *Acer platanoides* maples existed, that this tree was unknown to Uralic speakers and that the Tocharian name of this tree was borrowed by Finno-Volgaic speakers. The situation that results from this analysis is figured on the map.

This entails that Proto-Tocharian originates in the easternmost area where *Acer platanoides* maples used to exist, which seems to be the water-basin of the Kama river, a tributary of the Volga river. Finno-Volgaic speakers, coming from the east, met speakers of an Indo-European language there probably for the first time.



Present-day distribution of *Acer Platanoides* maple

5. The word ‘wagtail’

The white wagtail *Motacilla Alba* is widespread throughout all Eurasia and it cannot provide any indication about the previous location of the giver or receiver languages. The strangest feature is that

⁵ There is no dialectal form with initial *vi*.

Uralic languages should have a loanword to describe such a widespread and commonplace bird. We have no idea whether its distribution was originally limited to Western Europe. This item is not listed as an Indo-European loanword by the UEW but this is one more Proto-Tocharian item. The set of words is listed under UEW562 ‘white wagtail’:

- Finnish *västäräkki* (variants *västi*, *västäkä*, *västäkäinen*, *väästäjä*, *vestrikka*, *vestrikki*),
- Estonian *västrik* (gen. *västrikku*), dial. *väster*.

The formatives *-ter* ~ *-trik*, also attested in the word **wäkster* ‘maple’, clearly and easily identify an Indo-European word.

- Erzia-Mordvin *uškiřej⁶*, *uškeřej*, *veškořej*, Mokša-Mordvin *uškiřej*,
- Khanty (Ostyak) *wěrcək* (Sog.) *würśək* (Ni.),
- Mansi (Vogul) *väršex*.

In this example, the correspondence is not *va* ~ *u* ~ *vi* but *vä* ~ *u* ~ *vi* because the presence of *-i-* in the second syllable causes front vowel harmony in the first. The underlying Proto-Tocharian form is **wäsktriki*/**wäskter* and the cluster of consonants has been emended in Uralic languages in different ways, including metathesis and deletion. The PIE root is **weǵh* ‘to move, shake’ and is well attested in Tocharian A *wāsk-*, B *wāsk-* / *wāsk-* ‘to move, shake’. Cf. Adams (1999:589): (*mā*) *wāskäntrā* ‘he (does not) move(s)’ and Adams (1999:584): *waskamo* ‘mobile’. The conveyed meaning is precisely that of the English name *wagtail*. The UEW lists some other words which describe other birds. They probably do not belong here for phonetic and semantic reasons.

6. The word ‘to carry, to drag’

Another intriguing set of words meaning ‘to lead, to draw, to pull’ is listed under UEW569. These words are attested in a much larger set of languages:

- Finnish *vetä-* ‘to draw, to pull’, Estonian *veda-* ‘to lead, to draw’,
- Mari (Cheremis) *wiđe-* (KB) *wüđe-* (U B) ‘to lead’,
- Hungarian *vezet-* ‘to lead, to drive’,
- Erzia-Mordvin *vedja-* ‘to lead’, Mokša-Mordvin *vädja-* ‘to transport, to draw, to train’,

Some Samoyedic words with the same meaning are rejected by the UEW because of their back vocalism:

- Nenets (Yurak) *wādā-* ‘to draw, to lead’,
- Enets *bará-* (Ch), *badadda-* (B) ‘to draw’.

These words can easily be compared with PIE **wedh*, which the UEW erroneously cites as being **H₂wed*, but the next issue is which precise proto-forms can be reconstructed out of this set of words. The main problem is the vocalic scheme of the root. The UEW rejects Samoyedic forms which reflect **wod-* with a back vowel but this objection could also apply to the Mari forms with *ü*. Obviously, these words originate in one than one proto-form. One is **wedh-* and the other one is the causative **wodhey-*, where *-ey-* in the second syllable has caused a front harmonisation of **o*, as in Mokša *vädja*. It can be noted that these words, which are attested in Indo-Iranian, nevertheless have traces of a phoneme **e* as in PU **wete* ‘water’. Indo-Iranian normally changes PIE **e* and **o* to **a/ā* but these words seem to predate that sound change.

The UEW (569) also lists a subset of words with a voiceless dental: Erzia-Mordvin *viti-* ‘to lead’, Moksha-Mordvin *väte-* ‘to transport, to draw, to train’. The Erzian form *viti-* with *#vi-* is the next issue. The vocalic scheme *#vi* looks like another possible Tocharian item with underlying **wä*. As a

⁶ This is the form of the standard literary language. Quite strangely, it is the only one not listed in the UEW.

matter of fact the UEW does not list an additional set of words: Mokša *utjams* ‘to draw’, Erzia and Mokša *uskoms ~ uskəms* ‘to transport, to draw’ with the expected *#ut-sk < *wät-sk-*, with a typically Indo-European formative **sk*, which is attested in Celtic for this root. Again the alternation is *vä ~ u ~ vi*. These Mordvin words are not listed in the UEW but they belong here. It must nevertheless be noted that Adams (1999) does not seem to contain any Tocharian entry which may reflect **wad*, **wäsk* or **wäd*.

Another point is the voiceless *-t-* of *viti-* and *utja* which can only originate in a sequence *-Ht-* with a laryngeal. A simple stop like in PU **wete* becomes voiced in Mordvin: *ved* ‘water’. This is another reason to reject an Indo-Iranian origin for this subset of lexemes. This quasi minimal pair between **wet* ‘water’ and **wəHt-* ‘to draw’ suggests that early Proto-Tocharian preserved a contrast between at least two series of stops. The historical documents of the first millenium AD do not provide any indication of a contrast.

7. Conclusions and perspectives

The phonetic correspondence *va ~ u ~ vi* between Finno-Baltic and Mordvin is not only very unusual, never attested in Uralic cognates, but it distinctly identifies Tocharian loanwords in these languages. Some of them are shared with other branches of Uralic and they have been duly listed among the oldest loanwords of Indo-European origin. Some others, which have not been clearly identified as Tocharian so far, provide new and valuable insight into the prehistory of Proto-Tocharian and Proto-Finno-Volgaic. Our analysis suggests that Proto-Tocharian was once spoken in the easternmost extension of the *Acer Platanoides* maple around the Kama River and that Proto-Tocharian speakers remained in contact with Proto-Mordvin speakers for a longer period than is usually assumed. The general stratification of Indo-European loanwords in Uralic certainly needs to be emended.

As noted by Fortson (2004:352) exchanges and influences between Uralic and Tocharian happened in both directions: “Some structural features [of Tocharian], such as the large number of cases in the noun and the limited stop inventory (only voiceless stops), are not typical of IE languages but are found in Uralic, Turkic, and Mongolian languages of western and central Asia. It has been suggested that the Tocharians picked these features up from contact with those languages.” It is possible that lasting interactions with Finno-Volgaic and Mordvin are responsible for the un-Indo-European features of Tocharian. In addition one word tends to show that early Proto-Tocharian had preserved a contrast between at least two series of stops.

References

- Abondolo, Daniel M. (ed). 1998. *The Uralic Languages*. London and New York: Routledge.
- Adams, Douglas Q. 1999. *A Dictionary of Tocharian B*. Leiden Studies in Indo-European 10. Amsterdam-Atlanta: Rodopi.
- Collinder, Björn. 1955. *Fenno-Ugric Vocabulary: An Etymological Dictionary of the Uralic Languages*. Stockholm: Almqvist & Wiksell. (1977 Second, revised edition: Hamburg: Helmut Buske Verlag.)
- Collinder, Björn (ed). 1957. *Survey of the Uralic Languages*. Stockholm: Almqvist & Wiksell.
- Collinder, Björn. 1960. *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.
- Collinder, Björn (ed). 1965. *An Introduction to the Uralic Languages*. Berkeley: University of California Press.

- Décsy, Gyula. 1990. *The Uralic Protolanguage: A Comprehensive Reconstruction*. Bloomington, Indiana.
- Fortson IV, Benjamin W. 2004. *Indo-European Language and Culture, an Introduction*. Oxford: Blackwell.
- Hajdu, Péter. 1963. *Finnugor népek és nyelvek*. Budapest: Gondolat kiadó.
- Hajdu, Péter. 1975. *Finno-Ugrian Languages and Peoples*, translated by G. F. Cushing. London: André Deutsch.
- Joki, A. J. 1973. *Uralier und Indogermanen*. Helsinki (MSFOu 151).
- Korenchy, É. 1972. *Iranische Lehnwörter in den obugrischen Sprachen*. Budapest.
- Künnap, A. 2000. *Contact-induced Perspectives in Uralic Linguistics*. LINCOM Studies in Asian Linguistics 39. München: LINCOM Europa.
- Laakso, Johanna. 1992. *Uralilaiset kansat [Uralic Peoples]*. Porvoo-Helsinki-Juva.
- Munkácsi, B. 1901. *Árja és kaukazusi elemek a finn-magyar nyelvekben*. Budapest.
- Pinault, Georges-Jean. 1992. *Introduction au tokharien, Laxies 11*. Paris:Ulm.
- Pinault, Georges-Jean. 2008. *Chrestomathie Tokharienne, Textes et Grammaire*, Paris-Leuven:Peeters.
- Rédei, Károly (ed). 1986-88. *Uralisches etymologisches Wörterbuch [UEW]*, 3 vol. Budapest.
- Rédei, Károly. 1986. *Zu den indogermanisch-uralischen Sprachkontakten*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Sammallahti, Pekka. 1988. Historical phonology of the Uralic Languages. In Sinor, Denis (ed) *The Uralic Languages*, 478-554. Leiden: E.J. Brill.
- Sauvageot, Aurélien. 1930. *Recherches sur le vocabulaire des langues ouralo-altaïques*. Paris.
- UEW: see Rédei, Károly (ed). 1986-88.
- Vershinin, Valerij Ivanovich. 2004, 2005, 2009. *Etimologičeskij slovarj mordovskih (erzjanskogo i mokshanskogo) jazykov*, 4 vols. Joškar-Ola: String.