



Universiteit  
Leiden  
The Netherlands

**Indo-Uralic, Indo-Anatolian, Indo-Tocharian**  
Peyrot, M.; Kloekhorst, A.; Pronk, T.

**Citation**

Peyrot, M. (2019). Indo-Uralic, Indo-Anatolian, Indo-Tocharian. In A. Kloekhorst & T. Pronk (Eds.), *Leiden Studies in Indo-European* (pp. 186-202). Leiden: Brill. doi:10.1163/9789004409354\_014

Version: Publisher's Version

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3304387>

**Note:** To cite this publication please use the final published version (if applicable).

# Indo-Uralic, Indo-Anatolian, Indo-Tocharian\*

Michaël Peyrot

In this paper, I intend to illustrate the relevance of the Tocharian branch of Indo-European for questions concerning the Indo-Anatolian hypothesis and the possible macro-relationship of Indo-European with Uralic. To this end, I discuss the relevance for the Indo-Anatolian hypothesis of the possibility that the Tocharian branch was the second to split off after Anatolian, which I call the “Indo-Tocharian” hypothesis; the case of the Anatolian and Tocharian verb for ‘drink’ and its possible Uralic cognate; and the Anatolian and Tocharian *m*-interrogative, which has a neat parallel in Uralic. As will become clear, the results of these discussions are of uneven value, and the main aim of this paper is principally of a methodological nature.

## 1 Indo-Uralic, Indo-Anatolian, Indo-Tocharian

For the important but still disputed question of Indo-Anatolian the relevance of the Tocharian branch is in my view not so much its informativeness on the original state of affairs in Proto-Indo-European, since it has undergone so many sound changes, and lost and replaced so much of the original lexicon. Rather, its relevance is due to its presumed position in the Indo-European family tree: it is often assumed to have been the second branch to split off after Anatolian, as in the tree reconstructed by Ringe, Warnow & Taylor (2002: 87; see figure 13.1, next page).

Several authors have argued that Tocharian was the second branch to split off, e.g. Carling (2005: 48–49), Jasanoff (2003: 204), Kim (2007), Kortlandt (e.g. 2016: 81–82), Schindler (*apud* Jasanoff 2003: 46), Schmidt (1992) and Winter (1997). Yet, the evidence is not overwhelming, and, strikingly, many authors strongly differ in the arguments they adduce for this position of Tocharian in

---

\* This research was first supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme (project number 626656) and then by the Netherlands Organisation for Scientific Research (NWO, project number 276-70-028). This article is an adaptation of a talk at *The precursors of Proto-Indo-European: The Indo-Hittite and Indo-Uralic hypotheses*. 9–11 July 2015, Leiden University. I am grateful to the editors Alwin Kloekhorst and Tijmen Pronk for valuable comments on an earlier version.

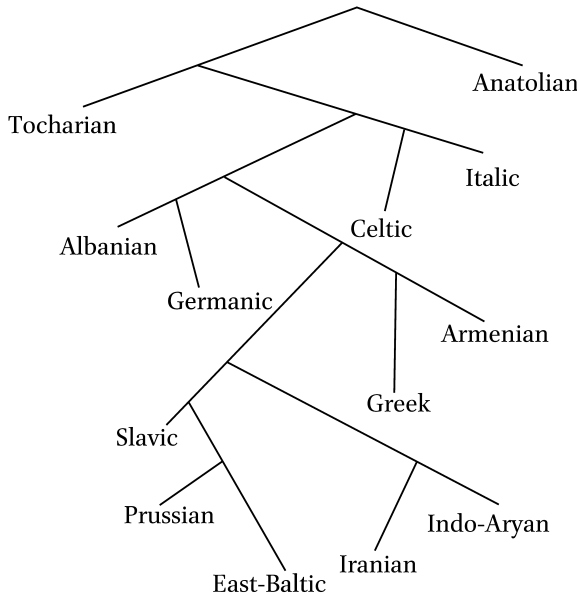


FIGURE 13.1 Tree of the Indo-European family; after Ringe, Warnow & Taylor 2002: 87; simplified

the family tree. There are also critical voices in the literature, for instance by Malzahn (2016), who argues that the lexical arguments of e.g. Schmidt (1992) and Winter (1997) are not strong and should not be used. I agree with Malzahn that we need more and better evidence before we can consider the early split-off of Tocharian proven, but that is no reason to discard the lexical evidence that we have. With Kloekhorst (2008; see below), I consider the potential of lexical evidence for subgrouping to be actually rather good. Although I admit that more work needs to be done, the hypothesis that Tocharian was the second to split off seems to me the most likely so far.

The possibility that Tocharian was the second branch to split off is relevant for the Indo-Anatolian hypothesis. Since the Indo-Anatolian node in the tree is defined by common innovations of the non-Anatolian branches, it is only of value in contrast to the next node in the tree. Thus, if the next node down from Indo-Anatolian is the node that we may term “Indo-Tocharian” (see figure 13.2, next page), Indo-Anatolian can only be established in contrast to Indo-Tocharian; if Tocharian evidence for a given common innovation of the non-Anatolian languages is lacking, the position of Tocharian for this feature is not clear, and what seems to be an argument for Indo-Anatolian could as well be in fact an argument for Indo-Tocharian. Obviously, if Tocharian was not the second branch to split off, the same argument applies to the branch that was in

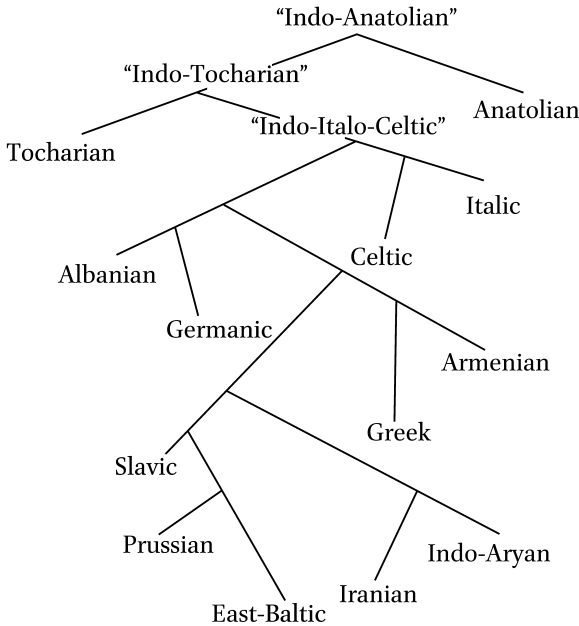


FIGURE 13.2 Tree of the Indo-European family with labels for the Indo-Anatolian, Indo-Tocharian and Indo-Italo-Celtic nodes

fact the first to split off after Anatolian, for instance Italo-Celtic. In light of the Indo-Tocharian hypothesis, therefore, evidence for Indo-Anatolian requires a systematic check against Tocharian data.

Below, I will systematically review Kloekhorst's seven arguments for Indo-Anatolian (2008: 8–10), because these are in my view well selected and clear evidence. The main purpose of this review is not to challenge Kloekhorst's arguments, but to illustrate the methodological case I want to make.

1) Hitt. *mer-zi* / *mar-* 'disappear' < *\*mer-*

The Hittite reflex of *\*mer-* means 'disappear', while in the other Indo-European languages the meaning 'die' is found. While the shift of meaning of 'disappear' to 'die' is commonplace according to a well established metaphor, the opposite semantic change is difficult to imagine.

In Tocharian, the verb for 'die' is *wäl-* in Tocharian A and *srəwka-* in Tocharian B; there is so far no reflex of *\*mer-*.<sup>1</sup>

1 Tocharian has the apparently related *\*mers-* 'forget' as A *mräsā-* 'forget', B *mərša-* 'id.'; this verb has not been influenced by the semantic shift of *\*mer-* and is therefore irrelevant here.

2) Hitt. *zīk* / *tu-* ‘you’ < PAnat. *\*tiH* / *\*tu-*

In the other Indo-European languages the vowel of the allomorphs *\*tiH* / *\*tu-* was levelled as *\*tuH* / *\*tu-*.

Tocharian goes together with the other non-Anatolian languages: Tocharian A *tu*, B *tuwe* < *\*tuH-om*.

3) Hitt. *šāh-i* ‘fill up, plug, stuff’ < *\*seh<sub>2</sub>-*

The Hittite reflex of *\*seh<sub>2</sub>-* means ‘fill up, stuff’, but those of the other Indo-European languages mean ‘satiated’. Again, a semantic change of ‘stuff’ to ‘satiated’ is more likely than the converse.

Tocharian goes together with the other non-Anatolian languages: Tocharian A *sāy(n)*- ‘satiated; be satiated, depressed’, B *soy-* ‘be sated’, *səyn-* ‘be satiated, depressed’.

4) HLuw. *tuwatra/i-* ‘daughter’ and Lyc. *kbatra-* ‘id.’ < *\*d<sup>h</sup>uegh<sub>2</sub>tr*, *\*d<sup>h</sup>ugh<sub>2</sub>tér<sup>m</sup>*

According to Kloekhorst, the Anatolian words for ‘daughter’ point to an original paradigm *\*d<sup>h</sup>uegh<sub>2</sub>tr*, *\*d<sup>h</sup>ugh<sub>2</sub>tér<sup>m</sup>*. After Anatolian split off, this paradigm was simplified to become the *\*d<sup>h</sup>ugh<sub>2</sub>tér*, *\*d<sup>h</sup>ugh<sub>2</sub>tér<sup>m</sup>* known from the non-Anatolian languages.

The Tocharian words for ‘daughter’ are A *ckācar* and B nom.sg. *tkācer*, obl.sg. *tkātār*. The initial *ck-* of Tocharian A *ckācar* is not regular. In theory, it could replace *\*cukācar* < *\*t<sup>w</sup>akacer* < *\*d<sup>h</sup>uegh<sub>2</sub>t-*, with elimination of the *\*-u-* after the non-nom.sg. cases with *\*tk-* < *\*takat-* < *\*d<sup>h</sup>ugh<sub>2</sub>t-* (this is what I suggested *apud* Kloekhorst 2011: 241). However, the evidence is weak, and *ckācar* more probably derives from earlier *\*tkācar* through distant assimilation. In this latter, more probable case, Tocharian goes together with the other non-Anatolian languages.

5) *ħarra-i* / *ħarr-* ‘grind, crush’ < *\*h<sub>2</sub>erh<sub>3</sub>-* and *ħārš-i* ‘harrow, till the soil’

Kloekhorst argues that Hittite *ħarra-i* ‘grind, crush’ preserves the older meaning, and that the meaning ‘plough’ of the same etymon in the non-Anatolian languages is a common, later development.

The verb *\*h<sub>2</sub>erh<sub>3</sub>-* is not attested as such in Tocharian. The regular expression for ‘plough’ seems to be TA *pātā-* ‘plough’, *pate* ‘ploughing’ < *\*b<sup>h</sup>od<sup>h</sup>h<sub>2</sub>-* (Lat. *fodiō*, *-ere* ‘pierce, dig’, Hitt. *padda-i*, *padd-* ‘dig (the ground, a pit)’, OCS *bodq* ‘stab’, Lith. *bedù* ‘stick, dig’). However, Tocharian A *äreñ* surely means

‘plough’. Apparently this is in origin the plural of an abstract noun *āre* ‘ploughing’ derived from a verb *\*ār-* or *\*ārā-*; this verb is obviously to be set up as *\*ārā-* < Proto-Tocharian *\*ara-* < *\*h<sub>2</sub>erh<sub>3</sub>-* (Peyrot 2018b: 262–263).

The replacement of the original verb for ‘plough’ by the verb for ‘dig’ is a little peculiar. Perhaps the reason is the phonological merger of *\*ara-* ‘plough’ with *\*ara-* ‘cease’ (Tocharian B *ara-* ‘cease’, A *arā-* ‘id.’).

6) Hitt. *mimma-i* / *mimm-* ‘refuse, reject’ < *\*meh<sub>1</sub>-*

According to Kloekhorst, the prohibitive negation *\*meh<sub>1</sub>* of the non-Anatolian languages derives from the imperative of the verb *\*meh<sub>1</sub>-* still preserved as such in Anatolian.

Tocharian clearly goes together with the non-Anatolian languages: Tocharian AB *mā* ‘not’.

7) The Anatolian words for ‘horse’ < *\*h<sub>1</sub>ékū-*

Kloekhorst reconstructs the Anatolian words for ‘horse’ as a *u*-stem, from *\*h<sub>1</sub>ékū-*. The thematic noun *\*h<sub>1</sub>ékūo-* found in the non-Anatolian languages is the result of a later development.

Tocharian clearly goes together with the non-Anatolian languages: Tocharian A *yuk* ‘horse’, B *yakwe* ‘id.’ < *\*h<sub>1</sub>ékūo-*.

Of these seven arguments for Indo-Anatolian, six easily stand the test because Tocharian goes together with the other non-Anatolian branches. Only number 1), ‘disappear / die’, is not attested in Tocharian and should therefore strictly speaking not be used as an argument for Indo-Anatolian because the change of ‘disappear’ to ‘die’ could theoretically also have taken place in the core Indo-European languages after Tocharian split off. Nevertheless, I do not think that we should discard this argument for Indo-Anatolian completely. It is in fact a good argument. We should just keep in mind that the position of Tocharian for this item cannot so far be decided.

## 2 Indo-Tocharian and Indo-Uralic? The Case of the Verb ‘Drink’

One of the frequently cited Hittite-Tocharian matches is Hitt. *eku-<sup>zi</sup>* / *aku-* ‘drink’ ~ Tocharian AB *yok-* ‘drink’ < *\*h<sub>1</sub>eg<sup>wh-</sup>* (Pinault 2006: 93). Although Anatolian and Tocharian are indeed the only two branches in which this verb is found, and most other branches have reflexes of the more common *\*peh<sub>3</sub>-* ‘drink’, this

etymon is difficult to use as an argument for the Indo-Tocharian hypothesis. Most importantly, even though Anatolian and Tocharian are the only branches in which *\*h<sub>1</sub>eg<sup>wh</sup>-* is attested, reflexes of this root are also found in Lat. *ēbrius* ‘drunk’ and Gr. *νήφω* ‘be sober’.

A lookalike of this Proto-Indo-European root is found in Uralic: compare among others Fi. *juo-* ‘drink’, Norw. Sa. *jukkâ-*, *-g-* ‘drink’, and Hu. *iv-* ‘drink’. In the *Uralisches Etymologisches Wörterbuch* (Rédei 1988–1991: 103), this etymon is reconstructed as *\*juye-* (*juke-*). The correspondence between PIE *\*h<sub>1</sub>eg<sup>wh</sup>-* and the reconstruction *\*juye-* (*juke-*) reminds of PIE *\*deh<sub>3</sub>-* ‘give’ ~ PU *\*toye-* ‘bring, get, give’ (Rédei 1988–1991: 529). Kortlandt (1989: 83) explained this correspondence assuming a Proto-Indo-Uralic preform *\*tagu-*, which developed through *u*-umlaut to PU *\*toye-* on the one hand, and with *\*gu* > *\*h<sub>3</sub>* to PIE *\*deh<sub>3</sub>-* on the other. In the same vein, one might explain PU *\*juye-*, *juke-* from PIU *\*eku-* with *u*-umlaut, and the initial *j-* perhaps as a result of breaking. PIE *\*h<sub>1</sub>eg<sup>wh</sup>-* would derive from this reconstructed PIU *\*eku-* through the change of *\*ku* to *\*g<sup>wh</sup>*.

Today, the reconstructions of the *Uralisches Etymologisches Wörterbuch* are generally viewed as outdated. However, the more recent reconstructions of Sammallahti (1988: 543, 550), PFU *\*toxi-* ‘bring’ and PFU *\*juxi-* ‘drink’, respectively, have not changed this picture in any essential way. On the basis of these reconstructions, it would still be possible to argue that PFU *\*toxi-* and PIE *\*deh<sub>3</sub>-* < PIU *\*tagu-*, and PFU *\*juxi-* and PIE *\*h<sub>1</sub>eg<sup>wh</sup>-* < PIE *\*eku-* or *\*egu-*.<sup>2</sup>

A more radical new reconstruction of the Uralic verb for ‘drink’ has been proposed by Aikio (2002: 38–40): *\*jixi*.<sup>3</sup> Obviously, the relevance of this revision is that no labial vowel is reconstructed anymore, which weakens the comparison between the Proto-Indo-European and the Proto-Uralic roots considerably. Revised reconstructions for Proto-Indo-Uralic are conceivable, for instance *\*igu* with *-u* > *-i* in Uralic, but the number of unproven sound laws that has to be assumed increases, so that such reconstructions are hardly falsifiable.

- 
- 2 Obviously, this derivation has to cope with the difficulty that PU *\*x* would correspond to PIE *\*h<sub>3</sub>* in ‘give’ but to *\*g<sup>wh</sup>* in ‘drink’. However, a more serious problem, in my view, is discussed directly below. Kortlandt later offered a revised explanation for this etymon: “The rounded laryngeal *\*q<sub>3</sub>* of Indo-European *\*deq<sub>3</sub>-* < *\*toqi-* suggests that the non-initial vowel was rounded as a result of Indo-Uralic vowel harmony in this root.” (2002: 217–227). I interpret this to mean that PIU (= PU) *\*toxi-* > *\*toxu-* > *\*texu-* > *\*deq<sup>w</sup>-* > PIE *\*deh<sub>3</sub>-*. In my view, this revision only brings Proto-Indo-Uralic closer to Proto-Uralic. Whether this is the right route to take is questionable, and I find his earlier derivation more convincing.
- 3 In the following, the symbols “*ɨ*” and “*ɛ*” denote a high unrounded back vowel (alternatively sometimes “*ɨ*”, IPA “*u*”) and a mid unrounded back vowel (alternatively sometimes “*ɛ*”, IPA “*ɜ*”), respectively.

Aikio's new reconstruction is based on 1) the inclusion of Proto-Samoyedic \**ε*- among the cognates, and 2) a different interpretation of the Permic and Hungarian vocalism (on which see below).

The Proto-Samoyedic root \**ε*- 'drink' is set up by Aikio on the basis of \**ε**r*- 'drink' and \**ε**kəl*- 'drink avidly'.<sup>4</sup> According to him, the longer root \**ε**kəl*- must be a derivative, and since the base is apparently \**ε*-, the other root \**ε**r*- must be a derivative as well. Indeed, so-called "augmentative" derivatives in *-r* are quite well attested: they are frequently found in the individual languages and for Proto-Samoyedic Janhunen (1977) lists *r*-augmentatives for \**əm*- 'eat', \**cin3-/cɨn3*- 'smell', \**jatə*- 'go', \**jäcə*- 'forge', \**ko*- 'see' (as well as \**kontʰə*- 'see'), \**kot*- 'cough', \**kunə*- 'run away', \**mej*- 'make' (\**mir*-), \**nət*- 'scrape', \**ńensə*- 'glide', \**ńim*- 'suck', \**pe*- 'seek', \**pit*- 'tan', \**sänä*- 'play', \**talä*- 'steal', \**təmta*- 'trade' (\**təmtajr*-) and \**wit3*- 'drink (water)'. Although the suffix *-kəl* is much less frequent, it has parallels too. Janhunen reconstructs derivatives in *-kəl* for \**u*- 'swim', \**ü*- 'drag', \**je* 'heel' (\**je-kəl*- 'step'), \**nek*-/*nek*- 'pull', \**ńic*- 'tear',<sup>5</sup> \**pən*- 'plait', \**pət*- 'dive', \**so*- 'scoop', \**tək*- 'hide'. Aikio then notes that Proto-Samoyedic roots of the structure \*(C)V usually continue PU \*(C)V*xi* and connects the Finno-Ugric etymon previously set up as \**juxi* by Sammallahti (1988: 543).

The problem with this connection is that PSam. \**ε*- suggests \**ixi* instead of PU \**juxi*. As a solution, Aikio proposes that the protoform was \**jixi* with sporadic loss of \**j*- in Samoyedic and a sporadic change of \**i* to \**u* in Finno-Ugric. The assumed loss of \**j*- in Samoyedic is in need of an exact conditioning, but Aikio adduces a parallel that is convincing in itself and this development seems acceptable. However, I have difficulties accepting his sporadic change \**i* > \**u* in Finno-Ugric. The reason is that his \**jixi* invites to reconsider the reconstruction of a number of other etyma, which in turn suggests that the change \**i* > \**u* in Finno-Ugric was not sporadic.

If Samoyedic needs \**ixi* or \**jixi* and Finno-Ugric \**juxi*, it may at first glance seem simpler to keep the PU reconstruction \**juxi* and assume a change of \**uxi* to \**ixi* in Samoyedic. The reason why this is not possible is that there are good examples for the correspondence PSam. \**-u* : PFU \**-uxi*, \**-uji*, which reflect PU \**-uxi*, \**-uji*:

4 For Proto-Samoyedic a weak vowel is reconstructed that is noted with "ə" in Janhunen 1977, "ø" in Sammallahti 1988 and "ø" in Janhunen 1998; here it is noted with "ə".

5 4 out of 8 are monosyllabic roots ending in a vowel. \**ńic-kəl*- 'tear' may have been formed after \**ü-kəl*- 'drag', cf. the variant \**ńüc*- of \**ńic*-, probably through influence from \**ü*-. Perhaps the suffix arose in roots in *-k*, of which there are two, and then spread to mostly monosyllabic roots ending in a vowel.



- PSam. \**tu*- ‘row’, PFU \**suxi*- < PU \**suxi*-
- PSam. \**u*- ‘swim’, PFU \**uji*- < PU \**uji*- (Aikio 2002: 44, who reconstructs \**j* instead of Sammallahti’s \**x*)

The following example of the correspondence PSam. \**-u* : PFU \**-uwi* < PU \**-uwi* is more problematic, according to Aikio (2012: 247), because the \**j* should not have disappeared in the Fi. cognate *puo* ‘anus’.

- PSam. \**puə*- ‘behind’, PFU ?\**puji* < PU ?\**puwi*

A similar correspondence is that between PSam. \**-o* and PFU \**-uxi*, \**-uwi*, \**-uji*. To explain this correspondence, PU \**o* is reconstructed with raising to \**u* in an open syllable before \**i* in Proto-Finno-Ugric (Sammallahti 1988: 486):

- PSam. \**ńo*- ‘pursue’, PFU \**ńuxi*- < PU \**ńoxi* (Aikio 2014a: 53)
- PSam. \**so* ‘mouth’, PFU \**śuwi* < PU \**śowi*- (Aikio 2002: 35, who reconstructs \**w* instead of Sammallahti’s \**x* because of South Sami *tjovve*)
- PSam. \**to* ‘lake’, PFU \**tuxi* < PU \**toxi*

A further etymon to be added here is probably PSam. \**tə*- ‘bring’ and PFU \**tuxi*- ‘bring’ (Sammallahti \**toxi*-). This root is not listed by Janhunen (1981) because the correspondence is not regular. According to the established basic sound correspondences (Janhunen 1981), there is no possible PU source form that could yield the PSam. root. I assume that the PU form was \**tuxi*-, which yielded \**tu* in Samoyedic. The \**u* of \**tu* was then weakened to \**ə* before \**a* in the second syllable in the derivative \**təta*-, as in PSam. \**kəpta*- ‘extinguish’ < PU \**kupsa*- (Sammallahti 1988: 484). Afterwards, the phonologically regular \**ə* of the derivative spread to the underived root. Since only \**u*, not \**o*, is weakened to \**ə*, PSam. \**tə*-, \**təta*- suggests \**tuxi*- for Proto-Uralic, not \**toxi*-.

The correspondence between PSam. \**-e* and PFU \**-uxi*, \**-uwi*, \**-uji* is also attested in more examples than just ‘drink’:

- PSam. \**e*- ‘drink’, PFU \**juxi* < PU \**jixi*
- PSam. \**le* ‘bone’,<sup>6</sup> PFU \**luwi* < PU \**liwi* (Aikio 2002: 35, who reconstructs \**w* instead of Sammallahti’s \**x* because of Mordvin *lovaža* with *o* and *v*)
- PSam. \**je* ‘tree’,<sup>7</sup> PFU \**juxi* < PU \**jixi*<sup>8</sup>

6 The preservation of the initial \**l* in Samoyedic is irregular. The expected outcome is rather \**je*. Conceivably, the \**l* was reintroduced from compounds, in which it was regularly preserved; cf. \**kuykälä* ‘shin bone’ from \**kuykə* ‘bend’ + \**le* and \**puajls* from \**puaj*- ‘knee’ + \**le* (Janhunen 1977). Obviously, if \**lempara* ‘breast’ contains \**le* as its first member, it must be a later formation. Another option is that loss or preservation of initial \**l*- depends on the following vowel. Aikio (2014c: 86) argues that \**l*- is regularly preserved before PU \**i*, and adduces three additional examples fitting this conditioning.

7 Janhunen (1977: 42) reconstructs this word as \**je<sup>l</sup>* = \**je*, \**je*.

8 The reconstruction of PFU \**juxi* follows Sammallahti (1988: 537), who sets up PU \**joxi* instead.

Since these examples include all three “weak” consonants *\*x*, *\*w*, *\*j* as well as the PSam. reflexes *\*-u*, *\*-o* and *\*-e*, there seems no other option than to reconstruct different vowels for Proto-Uralic, i.e. *\*u*, *\*o* and *\*i*, respectively, and to assume merger of all three into *\*u* in Proto-Finno-Ugric. An alternative hypothetical development of, for instance, *\*uxi* to *\*ixi* > *\*e* in PSam. (which could be a case of assimilation) is contradicted by PSam. *\*tu-* ‘row’, PFU *\*suxi-* < PU *\*suxi-*.

Aikio’s interpretation of the correspondence between unrounded vowels in Samoyedic and rounded vowels in Finno-Ugric is that sporadic rounding took place in several words in Finno-Ugric, and at several stages. This is unlikely. In view of the correspondences above, the rounding in Finno-Ugric was more probably a regular development, which occurred at an early stage in Finno-Ugric, not several times in the separate branches. In support of his idea that the rounding of the vowel of *\*jixi* ‘drink’ was a late development in this word in particular, he adduces Hu. *iszik*, *iv-* ‘drink’ and Hu. *új* ‘bow’ < *\*jijysi* (Aikio 2002: 40). However, since both words have initial *\*j-*, it seems best to return to the earlier explanation that the Hungarian vocalism in these words is due to a secondary development caused by the initial, and derive them from PFU *\*joxi* and *\*joysi*, respectively. I will not discuss the Permic evidence for survival of *\*i* into Finno-Ugric that Aikio adduces, because, according to him, it is uncertain.

The word for ‘bow’ is one of another small group of etyma in which Samoyedic shows unrounded vowels for rounded vowels in Finno-Ugric:

- PSam. *\*jintə* ‘bow’, PFU *\*joysi* < PU *\*jijysi* (Aikio 2002: 39)<sup>9</sup>
- PSam. *\*kĩnsV-* ‘star’, PFU *\*kunśa* < PU *\*kĩnśa*
- PSam. *\*kij* ‘moon, month’, PFU *kuxi*<sup>10</sup> < PU *\*kixji*

The last two words are further evidence against Aikio’s idea of a survival of old *\*i* in Hu. *iv-* and *új*, since Hungarian shows reflexes of rounded vowels in both cases: Hu. *húgy* ‘star’ and *hó*, *hava-* ‘moon’.

If indeed Samoyedic preserves the contrast between PU *\*o*, *\*u* and *\*i* so much better than Finno-Ugric, this further confirms that the primary split in the Uralic family was between Samoyedic on the one hand and Finno-Ugric on the other: the merger of these vowels is then a common innovation of the Finno-Ugric languages. Nevertheless, it must be noted that, apart from the position before *\*xi*, *\*wi*, *\*ji*, the exact conditions of these vowel changes are not clear; that the total number of examples is modest by all standards; and that

9 In a later article, Aikio reconstructs PU *\*joysi* (2014b: 11), perhaps for PFU *\*joysi*.

10 Aikio (2002: 39) notes that the reconstruction of *\*x* in this word is uncertain. Other options are *\*kuwi* and *\*kuji*.

the Proto-Samoyedic words are so short that not all relevant conditions are necessarily clear at this point. Another problem is that with the small number of accepted Uralic etymologies for Samoyedic it is largely unclear which root structures were absent for structural reasons and which happen to be unattested by chance. For instance, all roots with \*x are reconstructed with final \*i, while with other root types final \*a is found as well; it is theoretically possible that contrasts now seen as archaisms of Samoyedic are to be attributed to an earlier contrast between roots in \*-xi and \*-xa.<sup>11</sup>

With the revised reconstruction of the Uralic verb for ‘drink’ by Aikio, the comparison with PIE \*h<sub>1</sub>eg<sup>wh</sup>- loses much of its initial appeal. If Aikio’s revision turns out to be mistaken, or if more evidence for phonological correspondences between Proto-Indo-European and Proto-Uralic is found, the connection may eventually be revived—the semantic side, at least, is good. But at this point the more detailed reconstruction within Uralic overrules the macro-comparison with Indo-European.

For the subgrouping of Indo-European the verb \*h<sub>1</sub>eg<sup>wh</sup>- is difficult to use since the etymon is not confined to Anatolian and Tocharian, as mentioned above, even though it is best attested there. Lat. *ēbrius* ‘drunk’ and Gr. *νήφω* ‘be sober’ (see also Weiss 1994) simply prove that the root survived into core Indo-European. The only possible way to use this root for subgrouping is to assume that the original meaning was ‘drink’, which acquired the meaning ‘get drunk’ after Anatolian and Tocharian had left the speech community.

### 3 The *m*-interrogative

Next to the widespread PIE interrogative stem \*k<sup>w</sup>i-, \*k<sup>w</sup>e-, \*k<sup>w</sup>o-, there is another interrogative in \*m- (Dunkel 2014: 518–523). Although the existence of this interrogative stem is recognised in the literature, it is not well known, and not mentioned in standard introductions such as Beekes (2011: 227–231), Fortson (2004: 130) and Meier-Brügger (2003: 227–229). This is certainly due to the fact that this stem is clearly attested only in Anatolian and Tocharian. In Anatolian the following forms are found:

- Hitt. *maši-* ‘how many; however many’
- Hitt. *mān* ‘if, how, when, like’, possibly also the modal particle *man*
- Hitt. *mānḫanda* ‘just as’ (Kloekhorst 2010)
- Pal. *maš* ‘as much as’

11 According to Janhunen (2007: 216–217), the lack of roots in \*-xa may be due to a sound change of \*-ki to \*-xi: while roots in \*-ka are well attested, roots in \*-ki are extremely rare.

In addition, Hitt. =*ma* ‘and, but’ may be related, but the semantics are not compelling; several steps of development would have to be assumed, e.g. ‘how’ > ‘as’ > ‘as well as’ > ‘and’. All forms are apparently based on a stem *\*mo-*, e.g. Hitt. *maši-* < *\*mo-s-i-* and Hitt. *mān* < *\*mó-n* (Kloekhorst 2008: 552, 564).

The relationship of these Anatolian forms with a number of interrogatives, relatives and indefinites in Tocharian has been discussed by Hackstein (2004), Pinault (2010) and myself (Peyrot 2018a). Since the relevant formations are treated in some detail in Peyrot (2018a), I will here give only a brief presentation of the material.

The basic elements found in the Tocharian interrogatives, relatives and indefinites are:

- PToch. *\*kʷə-* < PIE *\*kʷi-*
- PToch. *\*mə-* < PIE *\*mo-*
- PToch. *\*ən-* < PIE *\*mo-*
- The PToch. demonstrative stem nom.sg.m. *\*se*, obl.sg.m. *\*ce*, nom.sg.f. *\*sa*, etc. < PIE *\*so*, *\*seh<sub>2</sub>*, *\*tod*
- The PToch. clitic markers *\*-nə*, *\*-w*, probably from PIE *\*nu*, *\*u*, and the emphatic particle *\*kə*

About PToch. *\*mə-* < *\*mo-* it should be noted that the reconstruction of the vowel *\*o* is based on Anatolian. As far as the Tocharian evidence is concerned, *\*mu* would also be possible, or, with loss of the expected palatalisation of *\*m*, also *\*mi* or *\*me*. This needs to be stressed, since the *\*m*-interrogative may have had different stem variants, e.g. *\*mi-*, *\*me-*, *\*mo-*, parallel to *\*kʷi-*, *\*kʷe-*, *\*kʷo-*, as pointed out to me by Tijmen Pronk. The derivation of *\*mə-* < *\*mo-* requires a special reduction of the vowel, since the normal reflex of *\*o* is PToch. *\*e*. A parallel for this reduction is found in the demonstratives, where next to *\*se* < *\*so* and *\*te* < *\*tod* we also find *\*sə* and *\*tə*.

In my analysis, PToch. *\*ən-* is a further development of *\*mə-*. Since *\*ən-* is always followed by a demonstrative element, the *\*n* can be explained from assimilation of original *\*m* to a following *\*t-* (or *\*c-*), which then spread throughout the paradigm, that is, to forms with a demonstrative in *\*s-*. Although the details remain to be settled, the rise of forms of the type *\*ən-te* < *\*m-te* < *\*mə-te* < *\*mo-tod* through syncope of the *\*ə* of *\*mə-* is probably due to the accent or the syllable structure. PToch. *\*ən-* is a unique initial, since no other word in the language begins with *\*ə-*. In Toch.A, this situation is preserved, with the unique initial *än-* < *\*ən-*, while in Toch.B we find it changed to *in-*.

It is not clear at which stage the demonstrative elements have been added, but since all three interrogative, relative and indefinite elements are uninflected, it is very likely that the inflected demonstrative could compensate for the loss of the inflexion of the elements *\*kʷə-*, *\*mə-* and *\*ən-*. The demonstra-

tives ending in *\*-e* have reduced variants in *\*-ə*: nom.sg.m. *\*sə* for *\*se* < *\*so* and sg.n. *\*tə* for *\*te* < *\*tod* (see also above).

Most of the complex formations listed below were probably found in Proto-Tocharian, but not in Proto-Indo-European. The PIE reconstructions are only meant to illustrate the derivation of the separate elements.

With the formative *\*mə-* we find:

- Toch.B *mäksu* ‘which’ < *\*mə-k<sup>w</sup>ə-sə-w* < *\*mo-k<sup>w</sup>i-so-u*
- Toch.B *mäkte* ‘how’ < *\*mə-k<sup>w</sup>ə-te* ‘what’ < *\*mo-k<sup>w</sup>i-tod*
- Toch.A *mänt* ‘how’ and Toch.B *mant* ‘so’ < *\*mə-ən-tə* ‘how’ < ‘what’ < *\*mo-mo-tod*
- Toch.B *mantsu* ‘some’ (?) < *\*mə-ən-sə-w* < *\*mo-mo-so-u*

In addition, Toch.B *manta* /*mánta*/ ‘never’ may derive from ‘ever’ and reflect *\*mə-ən-ta* < *\*mo-mo-teh<sub>2</sub>m*.

With the formative *\*ən-* we find:

- Toch.B *intsu* ‘which’ < *\*ən-sə-w* < *\*mo-so-u*
- Toch.A *äntsam* ‘which’ < *\*ən-se-nə* < *\*mo-so-nu*
- Toch.A *äntā* ‘where’ (with a reduced variant *tā*) < *\*ən-ta* < *\*mo-teh<sub>2</sub>m*

And further *\*ən* is found in second position in Toch.A *mänt* and Toch.B *mant*, *mantsu* and *manta* (see above).

With the formative *\*k<sup>w</sup>ə-* we find:

- Toch.B *k<sub>i</sub>se* ‘who, what’ and Toch.A *kus*, id. < *\*k<sup>w</sup>ə-se* < *\*k<sup>w</sup>i-so*
- Toch.B *ksa*, indefinite pronoun, a reduced form of *k<sub>i</sub>se* < *\*k<sup>w</sup>ə-se* < *\*k<sup>w</sup>i-so*

And further with *\*k<sup>w</sup>ə* in second position we find Toch.B *mäksu* and *mäkte* (see above).

As argued by Hackstein (2004: 280–283), in most of these formations the first element must have been originally interrogative, and if there was a second interrogative-relative element it was relative or “connective”. As I see it, the demonstrative element provided the necessary inflexion. Thus, the interrogative value of *\*mə-* is preserved in Toch.B *mäksu* ‘which’, *mäkte* ‘how’ and Toch.A *mänt* ‘how’; for *\*ən-* it is preserved in Toch.B *intsu*, Toch.A *äntsam* ‘which’, and in Toch.A *äntā*, *tā* ‘where’; and for *\*k<sup>w</sup>ə-* it is preserved in Toch.B *k<sub>i</sub>se*, Toch.A *kus* ‘who, what’.

In my view, Tocharian thus provides a solid piece of evidence for the reconstruction of the interrogative pronoun stem *\*mo-*. Interestingly, the co-existence of PIE *\*k<sup>w</sup>i-*, *\*k<sup>w</sup>e-*, *\*k<sup>w</sup>o-* and *\*mo-* is parallel to the two interrogative pronouns found in Uralic; cf. Fi. *kuka* ‘who’, Hu. *ki* ‘who’ and Fi. *mikä* ‘what, which’, Hu. *mi* ‘what’.<sup>12</sup> The comparison of the Proto-Indo-European and Proto-

12 Yet the vocalism is difficult to reconstruct. For instance, Rédei reconstructs PU *\*m3* ‘what’ with “3” as an “unbestimmbarer Vokal” (1988: 296).

Uralic *m*-interrogatives has been made a.o. by Pedersen (1938: 71–72), Collinder (1965: 113, 127, 149) and Greenberg (2000: 229–231, № 62). Although the equation is missing from Kortlandt's list of 27 Proto-Indo-Uralic grammatical elements (2002), it seems solid enough to be added.

Nevertheless, the neat contrast observed between Hu. *ki* 'who' and *mi* 'what' is not found in Indo-European. In view of Hitt. *maši-* 'how many' and Pal. *maš* 'as much as', Hackstein (2004: 281–282) suggests that *\*mo-* was a quantifying interrogative, 'how many'. However, I find this unattractive for the more basic meanings attested in Tocharian. If any more precise meaning should be reconstructed for *\*mo-*, it would rather be something like 'which', a more specific, restrictive interrogative; cf. Hackstein (2004: 281–282), "TB *mäksu* is best described as an adjectival interrogative for restricting reference, "which one of a given class or group."

For the internal subgrouping of Indo-European, the Tocharian-Anatolian match in the *m*-interrogative is of limited value at most. First of all, in the words of Pedersen, "Es genügt hervorzuheben, dass Verlust des Alten (in diesem Falle des *m*-Pronomens), der allmählich in jedem Sprachzweige für sich eingetreten sein kann, nicht als eine gemeinsame Neuerung gewertet werden darf." (1938: 72). This statement can be relativised a little, since it is obviously more economical to assume loss of an archaic feature once, as a common innovation of the languages that have not preserved it, than several times, for each branch independently. This is all the more true in the case of a grammatical element such as the *m*-interrogative. At the same time, it is clear that a common innovation that involves a loss is a much weaker argument for subgrouping than a common innovation consisting of an indisputably traceable "positive" change.

In the case of the *m*-interrogative, the evaluation of its usefulness for subgrouping is further complicated by the existence of further possible traces in Celtic, where we find OIr. *má* 'if' and related forms.<sup>13</sup> Although these apparently have no interrogative value, original interrogative value may be suggested by the Breton and Cornish local conjunction *ma* (Pedersen 1913: 230). If these conjunctions derive from the stem *\*mo-*, the only possible common innovation that remains for the Indo-European languages including Celtic and excluding Anatolian and Tocharian is a development of the original interrogative to a relative or a local relative. Since shifts of this kind are commonplace, as

13 Dunkel lists many more cognates of the *\*mo*-stem, including reflexes as indefinite in Vedic and Armenian, but these are in my view not convincing (Dunkel 2014: 518–523). The possible sources of indefinites are manifold (Haspelmath 1997) and not each and every *-m* can be derived from the same interrogative *\*mo-*.

shown by the Tocharian interrogatives and relatives listed above, but also by similar developments in other branches of Indo-European and in non-Indo-European languages, this would make the evidence of the *m*-interrogative for Indo-European internal subgrouping practically useless.

#### 4 Conclusions

In this paper, I have treated a number of independent problems that illustrate some of the methodological points relevant to the questions of the phylogenetic structure of the Indo-European language family and its supposed genealogical relationship with the Uralic language family.

I have argued that the Indo-Anatolian hypothesis requires the next node down to be meaningful. Assuming that the second branch to split off the Indo-European protolanguage was Tocharian, all evidence for the Indo-Anatolian node needs to be checked against this presumed "Indo-Tocharian" node in order to see whether also Tocharian goes together with the other non-Anatolian languages.

Since the Anatolian-Tocharian isogloss for *\*h<sub>1</sub>eg<sup>wh</sup>*- 'drink' is not exclusive, it can only be used as an argument for subgrouping if the meaning 'get drunk' found in the Greek and Latin reflexes is a common innovation of the other branches. The resemblance of the root *\*h<sub>1</sub>eg<sup>wh</sup>*- to Proto-Finno-Ugric *\*joxi* 'drink' turns out to be only superficial because of Aikio's addition of the Samoyedic cognate *\*e-*, which suggests a Proto-Uralic reconstruction *\*jixi*. Apart from the meaning, this revised reconstruction shows no obvious similarity with *\*h<sub>1</sub>eg<sup>wh</sup>*-.

The Proto-Indo-European interrogative stem in *\*m-* next to the well known *\*k<sup>w</sup>i-*, *\*k<sup>w</sup>e-*, *\*k<sup>w</sup>o-* has not received due attention in the literature, but can nevertheless safely be reconstructed, and especially Tocharian provides strong evidence for it, with for instance TB *mäksu* 'which', *intsu* 'which', TA *äntsam* 'which', etc. The reconstruction of a closely parallel set of interrogative stems for Proto-Uralic is uncontroversial; cf. for instance Fi. *kuka* 'who' next to *mikä* 'what, which' and Hu. *ki* 'who' next to *mi* 'what'. Especially taken together with other parallels in grammatical elements, the neat correspondence of two interrogative stems in Proto-Indo-European and Proto-Uralic is further evidence in favour of a genealogical relationship between the two families.

## References

- Aikio, A. (Luobbal Sámmol Sámmol Ánte) 2002. New and Old Samoyed etymologies. *Finnisch-Ugrische Forschungen* 57, 9–57.
- Aikio, A. (Luobbal Sámmol Sámmol Ánte) 2012. On Finnic long vowels, Samoyed vowel sequences, and Proto-Uralic \*x. In: T. Hyytiäinen (ed.), *Per Urales ad Orientem. Iter polyphonicum multilingue. Festschrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Helsinki: Finno-Ugrian Society, 227–250.
- Aikio, A. (Luobbal Sámmol Sámmol Ánte) 2014a. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62, 7–76.
- Aikio, A. (Luobbal Sámmol Sámmol Ánte) 2014b. Studies in Uralic etymology 11: Finnic etymologies. *Linguistica Uralica* 50, 1–19.
- Aikio, A. (Luobbal Sámmol Sámmol Ánte) 2014c. Studies in Uralic etymology 111: Mari etymologies. *Linguistica Uralica* 50, 81–93.
- Beekes, R.S.P. 2011. *Comparative Indo-European Linguistics. An Introduction*. Second edition. Revised and corrected by Michiel de Vaan. Amsterdam–Philadelphia: John Benjamins.
- Carling, G. 2005. Proto-Tocharian, Common Tocharian, and Tocharian. On the value of linguistic connections in a reconstructed language. In: K. Jones-Bley; M.E. Huld; A. Della Volpe; M. Robbins Dexter (eds.), *Proceedings of the Fifteenth Annual UCLA Indo-European Conference*. Washington: Institute for the Study of Man, 47–70.
- Collinder, B. 1965. Hat das Uralische Verwandte? Eine sprachvergleichende Untersuchung. *Acta Universitatis Upsaliensis* 1/4: 108–180.
- Dunkel, G.E. 2014. *Lexikon der indogermanischen Partikeln und Pronominalstämme. Band 2: Lexikon*. Heidelberg: Winter.
- Fortson, B.W., IV 2004. *Indo-European Language and Culture. An Introduction*. Oxford: Wiley-Blackwell.
- Greenberg, J.H. 2000. *Indo-European and its Closest Relatives: the Eurasiatic Language Family. Vol. 1: Grammar*. Stanford: Stanford University Press.
- Hackstein, O. 2004. From discourse to syntax: The case of compound interrogatives in Indo-European and beyond. In: K. Jones-Bley; M.E. Huld; A. Della Volpe; M. Robbins Dexter (eds.), *Proceedings of the Fifteenth Annual UCLA Indo-European Conference*. Washington: Institute for the Study of Man, 257–298.
- Haspelmath, M. 1997. *Indefinite Pronouns*. Oxford: Clarendon.
- Janhunen, J. 1977. *Samojedischer Wortschatz*. Helsinki: Finno-Ugrian Society.
- Janhunen, J. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77, 219–274.
- Janhunen, J. 2007. The primary laryngeal in Uralic and beyond. In: J. Ylikoski; A. Aikio (eds.) *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahtii miessemánu 21. beaivve* 2007. Helsinki: Finno-Ugrian Society, 203–227.



- Jasanoff, J.H. 2003. *Hittite and the Indo-European Verb*. Oxford–New York: Oxford University Press.
- Kim, R.I. 2007. The Tocharian subjunctive in light of the  $h_2e$ -conjugation model. In: A.J. Nussbaum (ed.), *Verba Docenti. Studies in Historical and Indo-European Linguistics Presented to Jay H. Jasanoff by Students, Colleagues, and Friends*. Ann Arbor: Beech Stave, 185–200.
- Kloekhorst, A. 2008. *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden–Boston: Brill.
- Kloekhorst, A. 2010. Hittite *mān*, *maḥḥan*, *māḥḥan*, *māḥḥanda* and *mānḥanda*. In: R. Kim; N. Oettinger; E. Rieken; M. Weiss (eds.), *Ex Anatolia Lux: Anatolian and Indo-European Studies in Honor of H. Craig Melchert on the Occasion of his Sixty-Fifth Birthday*. Ann Arbor–New York: Beech Stave Press, 217–226.
- Kloekhorst, A. 2011. The accentuation of the PIE word for ‘daughter’. In: T. Pronk; R. Derksen (eds.), *Accent Matters. Papers on Balto-Slavic Accentology*. Amsterdam–New York: Rodopi, 235–243.
- Kortlandt, F.H.H. 1989. Eight Indo-Uralic verbs? *Münchener Studien zur Sprachwissenschaft* 50, 79–85.
- Kortlandt, F.H.H. 2002. The Indo-Uralic verb. In: R. Blokland; C. Hasselblatt (eds.), *Finno-Ugrians and Indo-Europeans: Linguistic and Literary Contacts*. Maastricht: Shaker, 217–227.
- Kortlandt, F.H.H. 2016. Baltic, Slavic, Germanic. *Baltistica* 51, 81–86.
- Malzahn, M. 2016. The second one to branch off? The Tocharian lexicon revisited. In: B. Simmelkjær Sandgaard Hansen; B. Nielsen Whitehead; T. Olander; B.A. Olsen (eds.), *Etymology and the European Lexicon. Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*. Wiesbaden: Reichert, 281–292.
- Meier-Brügger, M. 2003. *Indo-European Linguistics. In Cooperation with Matthias Fritz and Manfred Mayrhofer*. Berlin–New York: Walter de Gruyter.
- Pedersen, H. 1913. *Vergleichende Grammatik der keltischen Sprachen. Zweiter Band*. Göttingen: Vandenhoeck & Ruprecht.
- Pedersen, H. 1938. *Hittitisch und die anderen indoeuropäischen Sprachen*. København: Munksgaard.
- Peyrot, M. 2018a. Interrogative stems in Hittite and Tocharian. *Indogermanische Forschungen* 123, 65–90.
- Peyrot, M. 2018b. Tocharian agricultural terminology: Between inheritance and language contact. In: G. Kroonen; J.P. Mallory; B. Comrie (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013*.
- Pinault, G.-J. 2006. Retour sur le numéral “un” en tokharien. *Indogermanische Forschungen* 111, 71–97.

- Pinault, G.-J. 2010. Le pronom d'ipséité en tokharien. In: I. Choi-Jonin; M. Deval; O. Soutet (eds.), *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Leuven–Paris: Peeters, 351–365.
- Rédei, K. 1988–1991. *Uralisches etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.
- Ringe, D.; T. Warnow; A. Taylor 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100, 59–129
- Sammallahti, P. 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric, and Permic. In: D. Sinor (ed.), *The Uralic Languages. Description, History and Foreign Influences*. Leiden–New York: Brill, 478–554.
- Schmidt, K.T. 1992. Archaismen des Tocharischen und ihre Bedeutung für Fragen der Rekonstruktion und der Ausgliederung. In: R.S.P. Beekes; A. Lubotsky; J.J.S. Weitenberg (eds.), *Rekonstruktion und relative Chronologie. Akten der VIII. Fachtagung der Indogermanischen Gesellschaft, Leiden, 31. August–4. September 1987*. Innsbruck: IBS, 101–114.
- Weiss, M. 1994. On the non-verbal origin of the Greek verb *νήφειν* 'to be sober'. *Historische Sprachforschung* 107, 91–98.
- Winter, W. 1997. Lexical archaisms in the Tocharian languages. In: H.H. Hock (ed.), *Historical, Indo-European, and Lexicographical Studies. A Festschrift for Ladislav Zgusta on the Occasion of his 70th Birthday*. Berlin: Mouton de Gruyter, 183–193.