Universiteit Leiden
The Netherlands

Like dust on the Silk Road: an investigation of the earliest Iranian loanwords and of possible BMAC borrowings in Tocharian
Bernard, C.B.A.S.

Chapter 2: Old Steppe Iranian loanwords in Tocharian

2.1 Introduction

2.1.1 Old Iranian loanwords and the history of their research

In this chapter I will discuss all the Iranian loanwords in Tocharian (A and B) imputable to an Old Iranian stage. The fact that there was a "mystery" Old Iranian layer of contact with Tocharian has been known since Schmidt 1985, but the quantity of words attributable to this stage has been increasing ever since. One important question to consider is whether the loanwords of this Old Iranian stage belong to one single language or to multiple varieties. I have collected all examples that I consider to be secure, adding a few (section 2.2). I will discuss those which are more problematic in 2.3, and those which are less likely or which could also come from other sources in 2.4. Finally, those which I rejected, or which cannot be considered as of Old Iranian provenance at this stage, are discussed in 2.5.

Since Schmidt (1985), a number of discoveries have been made concerning this layer of Old Iranian words. Specific features were established already by Schmidt (1985), such as Proto-Tocharian *t^s representing the reflex of Proto-Iranian *j and *$ć$ in the source language, Proto-Tocharian *e representing the reflex of Proto-Iranian *a, Proto-Tocharian *a representing the reflex of Proto-Iranian *$ā$, etc. Since his seminal study, new words have been found, and I have endeavoured to establish a complete list of all features, both features discovered by and after Schmidt, and newly found ones. All these features are discussed in detail in section 2.6 of this chapter.

Despite these features being better known than before, no study on loanwords in Tocharian or on Tocharian historical phonology has gathered them systematically. On the contrary, many studies done on Tocharian loanwords have ignored the systematicity and coherence of the earliest layer of Old Iranian loanwords in Tocharian. Tremblay (2005), notably, although he aimed at systematicity, has not been able to analyze all the data coherently, did not understand the sound correspondences, and confused various layers of contact. This has led to the inclusion of words which are not of Iranian

origin such as TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric', s.v. and TB *witsako* 'root', s.v.. Because this category of words was included in research on Iranian loanwords in Tocharian, I have dedicated a chapter to the discussion of another group of old loanwords which are not of an Iranian origin: the so-called substratum or BMAC loanwords in Tocharian.

In fact, once this layer of oldest words is separated from the other borrowings, it can inform us on Iranian languages. Armenian has yielded a trove of useful information, and one can think that this trove has not been exploited to its fullest extent: Iranian loanwords in Armenian provide insight into Middle Iranian languages, the chronology of changes, and dialectological data, alongside cultural and historical information. The present study aims to prove that similar results can be obtained with Iranian loanwords in Tocharian, in particular, that the unearthing of an otherwise unattested ancient language will be made possible.

As I discuss in more detail in the conclusion (section 4.1), I will depart from the assumption that all these ancient loanwords in Tocharian, sharing a set of features, are from one single language. This is more practical, in my view, than, for example, starting with the idea that these loanwords are from multiple languages. While the latter is possible, naturally, if one assumes that there is one single language with definite features, it will be easier to recognize when some of the features analyzed in those words do not fit, and if more than one language is to be assumed. On the other hand, assuming multiple languages makes us ask, with every word, the question whether this word belongs with that one, or with that other one. With a wide set of loanwords, this is thinkable, but there are only about 48 words that I deem of Old Iranian origin discussed in this chapter.

In this chapter I will thus discuss the etymology of each of the recognized Old Iranian loanwords in Tocharian. A discussion will ensue, where I will try to determine to which group of Iranian languages this Old Iranian language belonged, and if it is one single language. This discussion will be based on the isoglosses found in the words mentioned here, with, naturally, a greater emphasis on the first part of the list, the more secure loanwords, but occasionally using the second part of the list to support specific points. A conclusion will follow, although the determination of the dialectal affiliation of this Old Iranian language will remain tentative.

2.1.2 Naming the language

A new language – no matter how old – deserves a new name. If indeed this Old Iranian language is one single language, we cannot know how these Iranians called themselves or their language. Possibly, they called themselves 'Aryans', that is, Iranians, or 'Scythians', whatever reality this notion might have covered for them. Perhaps they called themselves 'men', as did the Ossetes (Digoron *ir*, Iron *iræ* 'Ossetic, Ossetian' ultimately derives from **u̯īrāh* 'men', cf. Bielmeier 1987: 103; Cheung 2002: 193), and so many other people. This we will never know for sure, until we have found a designation of that people in Tocharian A or B, or in another source.

In the absence of a clear ethnonym, or auto-ethnonym, I propose to name this Old Iranian language "Old Steppe Iranian". This is based on the supposition that the speakers of this language came from the eastern Eurasian steppes, possibly as part of the culture associated with roller pottery, which massively replaced previous Andronovo cultures (cf. Parpola 2022). In these Eurasian steppes, they were possibly in contact with speakers of Proto-Tocharian, perhaps in the plains to the north of the Tarim Basin. I dedicate a section of the conclusion of this thesis (section 4.2) to thoughts on the prehistory of this people, and on possibilities surrounding their contact with Tocharians.

If Old Steppe Iranian belonged to the Old Iranian linguistic stage, and was probably spoken at some point in the steppic areas, this does not necessarily mean that it is identical to all the other Old Iranian varieties spoken in the Steppes, and in particular not to those spoken in the Western part of the Steppes (cf. section 2.6). I hope the reader will forgive the vagueness of this name of convenience, which is the best I could offer. I also hope the reader will enter thus the story of this language through the traces it left, like dust on the Silk Road.

2.2 Old Steppe Iranian borrowings: plausible cases

2.2.1 TB *entse*, TA *eṃts* 'envy, greed'

Tocharian B *entse*, Tocharian A *eṃts* 'envy, greed' go back to Proto-Tocharian **entse*, with Pre-A **antsa* > **aynts* > *eṃts* (cf. Hilmarsson 1986: 282). Their etymology is debated. In the following, I will first discuss Hil-

marsson's etymological proposal, and then put forward my own proposal, which is largely similar to that of Isebaert (2019).

According to Hilmarsson (1986: 282), TB *entse* and TA *eṃts* are inherited and derive from Proto-Indo-European **h₂emǵʰos-*, compare Sanskrit *áṃhas-* 'fear, anxiety', Av. *ązah-* 'constriction; distress, peril'. In addition, Hilmarsson reconstructs an Indo-European "holokinetic" masculine *s*-stem **h₂emǵʰōs*, "preserved in Lat. *angor* (*-ōris*) "anguish; compression of the throat"." For more cognates, s.v. *ñyātse*. More exactly, Hilmarsson proposes that TB *entse* and TA *eṃts* go back to an *o*-stem derivative of the *s*-stem, i.e. **h₂omǵʰ-s-o-*:

> "[a]n abstract *o*-stem derivation with accentuated *o*-vocalism of the root […] i.e. I.-E. **h₂omǵʰso* [my notation], would in Tocharian yield **enkse* [my notation], which would result in the attested B *eṃtse*, A *eṃts* […]." (1986: 282).

There are three main difficulties linked to this etymology: first, it presupposes a Proto-Indo-European *o*-stem derivative **h₂omǵʰ-s-o-* formed from the *s*-stem **h₂emǵʰos-*, of which there is no trace. Second, it is based on a sound law, namely Pre-Proto-Tocharian **nks* > Proto-Tocharian **n(t)s*, that has no parallel.[1] Third, the meaning of this word would, in any case, be far removed from that of 'greed'. The root, which means 'to tie (a knot), to restrain', often takes the meaning of 'distress', due to the idea of narrowness, of a tight throat and of difficulty to breathe (cf. *ñyātse*), but it does not normally designate 'greed'.

Because of these three difficulties with Hilmarsson's derivation from Proto-Indo-European, it seems much more probable that Proto-Tocharian **entse* was directly borrowed from Old Steppe Iranian **anǯa(h)*, cf. Avestan *ązah-* 'constriction; distress, peril' < PIr. **anǰah-*, compare also Ved.

---

[1] While the sound change **nkt* > *nt* occurred in Tocharian A (e.g. PT **p'ənkte* 'fifth' > TA *pänt* but TB *piṅkte* 'id.'), there is no trace of it in Tocharian B (cf. DTB: 644). The example TB *wäntalyi* 'bow(-string)' adduced by Blažek & Schwarz (2017: 97) can be explained much more easily by a connection to TB *wənta-* 'to cover' (cf. DTB: 644; Peyrot 2013: 538f.) than by a relationship to Lithuanian *vìngis* 'bow, bending' (Van Windekens 1976: 556; Blažek & Schwarz 2017: 97).

áṃhas- 'id.',[2] as was also proposed independently by Isebaert (2019). The assumption of a borrowing from Old Steppe Iranian accounts neatly for the form of the Tocharian word, with the regular correspondences between OSIr. *a and PT *e; OSIr. *ĝ < PIr. *j and PT *ts; and between final *-a(h) and PT *-e. For the last correspondence, cf. TB *perne* < PT *perne, borrowed from OSIr. *farna(h).

However, there is a discrepancy in meaning between PIr. *anjah- 'constriction, distress, peril' and PT *entse 'greed'. Possibly, the Old Iranian word at the source of this borrowing developed the meaning 'greed', perhaps under the influence of the similar-sounding and perhaps related *āzi- 'greed'. Otherwise, the shift of meaning would have to have occurred within Tocharian. In both cases, the semantic change may be conceived of as 'distress' > 'need' > 'want' > 'greed'. Here, the difference in the last steps could be explained as the notion of "need", motivated by external circumstances, changing to that of "greed", caused by an internal (usually mental) motivation.

An alternative etymology of Proto-Tocharian *entse has been suggested to me by Georges-Jean Pinault (p.c.), whereby the word goes back to (PIE) *n̥-d(h₃)-o- 'absence de don, qui ne donne pas'. This etymology is possible formally, if *d developed to *ts before *H, but no Indo-European cognate comes to confirm the existence of the projected *n̥-d(h₃)-o-.

In conclusion, Tocharian B *entse*, Tocharian A *eṃts* 'envy, greed' are probably of Iranian origin and borrowed from OSIr. *anĝa(h).

Interestingly, Tocharian B *entse* is probably the word in which the Tocharian B suffix *-sse* has arisen. This suffix is found in *śāmpasse* 'with haughtiness', *waikesse* 'with falsehood', *werässe* 'with hate', and similar words, according to Winter denoting 'provided with the negative property X' (cf. Winter 1979: 991).

---

[2] According to Morgenstierne (1942: 265), the derivation of New Persian *hōšāz* 'excessive thirst of cattle' from *hau̯ša-anĝah- 'drought distress' is slightly problematic. Indeed, *an- is not expected to yield *ā- in Persian. A compound *hau̯ša-āĝi- would mean 'drought desire; drought greed' which at first does not seem very convincing. However, one could interpret the compound as meaning 'greed (caused by) drought', which does not seem out of place here. I thus believe an etymon *hau̯ša-āĝi- to be more fitting than Morgenstierne's 'drought distress'.

Winter (1979: 991-92) argues that this suffix spread from *entsetstse 'greedy, of greed', where -tstse was dissimilated to -sse after the -ts- of the base word. It spread both to inherited words such as waike (DTB: 666) and to other loanwords, such as wer, which is borrowed from Prakrit, and śām-po*, which is borrowed from Khotanese (cf. Dragoni 2022: 191-193). Thus, indirectly, the Tocharian B suffix -sse arose due to borrowing of the Old Steppe Iranian word *anɟa(h).

2.2.2 TB epiyac, TA opyāc 'memory'

Tocharian B epiyac and Tocharian A opyāc, both meaning 'memory', go back to Proto-Tocharian *epiyac. The Tocharian A o- of opyāc is the result of the p/m-umlaut of Pre-TA *a < PT *e, as in TA porat (< PT *peret) 'axe' s.v. TB peret, TA porat. TB epiyac and TA opyāc have been recognized as Iranian loanwords for a long time (e.g. Hansen 1940: 151).

The Iranian source of the borrowing clearly belongs to the group of words Khotanese byāta 'memory', and Middle Persian ayyād, New Persian yād 'memory', etc. These words reflect a formation with PIr. *abi-, and the root is set up as *Hi̯aH- by Cheung (EDIV: 175-76), i.e. PIr. *abi-Hi̯aHta- (or *-ti). Cheung translates *Hi̯aH- as 'to remember' and supposes that it is related to Skt. yā- 'drive'. We could perhaps think of *abi-Hi̯aHta- as 'the thing that came around (the mind)', that is the memory.

The phonological correspondences to be observed between the Iranian words cited and PT *epiyac point to borrowing from Old Steppe Iranian, for which a form *abii̯āta- can be set up. The correspondences are then mostly regular: OSIr. *a is rendered as PT *e and OSIr. *ā is rendered as PT *a. The prefix *abí- presumably had fixed accent on the second syllable (as in Indic), which is reflected in Tocharian, i.e. TB epiyac /epə́yac/.

However, the final -c of the Tocharian forms can absolutely not derive from OSIr. *-ta, and it was already mentioned as problematic by Hansen (1940: 151). Nonetheless, most researchers have presented it as being the result of a palatalization of *t in front of *i, as if from *abi-i̯āti-, e.g. Isebaert (1980: 103); Hilmarsson (1986: 56); Klingenschmitt (2000: 199); Tremblay (2005: 424); Pinault (2008: 451); etc.

Adams observes, correctly in my view, that nothing permits us to posit a proto-form *abi-i̯āti- at any stage of Iranian (DTB: 95). Nevertheless, he too

suggests that a "Middle Iranian" *abi-i̯āti- was borrowed early enough for it to undergo the Tocharian reflexes of PIE *ti*-stems, with an alternation -*t* ~ -*c*.

This is completely *ad hoc*, as there is not solid evidence that *ti* is reflected as *c* in Tocharian words borrowed from Old Steppe Iranian. If one believes my tentative derivation of TB *epastye*, TA *opäśśi* (s.v.) < PT *epəstəye* from OSIr. *abi-st(H)-ii̯a-, this would show that Old Steppe Iranian *-ti- was rather reduced to *-tə-, at least after *-s-.

A few hypotheses can be set up to explain the *c* of TB *epiyac*, TA *opyāc*, but none so far is convincing. An idea which I abandoned consisted in suggesting that, since *epiyac* is frequently used together with the verb *yam-* 'to do', with the meaning 'to remember', a palatalization of *epiyat yam* (< *epiyate yam*) to *epiyac yam* could hypothetically have occurred. There is, however, no explanation for the supposed loss of final -*e*, and, more importantly, no reason to consider an inner Tocharian palatalization of -*t* + *y*- > -*c* + *y*-. This proposal is thus not attractive in order to explain the palatal -*c* of *epiyac* 'memory' and other solutions must be put forward.

One could, for instance, imagine a derivation from a participle *i̯ānt- to which the Old Iranian *abi-i̯āta- could go back. It would thus be *i̯ānt- → *i̯āθi̯a- (cf. *hant-* → *haθi̯a-*), but this is no longer a productive derivation process in Iranian. Another possibility is an ancient gerundivum: *abi-i̯ā-tii̯a- 'what needs to be remembered', but this word too is not found anywhere.

Yet another possibility to explain the final -*c* in TB *epiyac*, TA *opyāc*, consists in viewing the word as an ancient allative form PT *epiya-c or *epiyate-c (> *epiyatc) 'towards the memory'. Memory, indeed, represents a moving process, cf. the possible etymon *HiaH- 'to go' of Old Iranian *abi-i̯āta-. Memory can be conceptualized as a motion (of the mind) towards the past. One could think that this notion of movement would have been expressed as an allative in Proto-Tocharian. If the allative was added to a bare form *abi-i̯ā- instead of *abi-i̯āta-, one can perhaps compare this bare form to other unsuffixed Old Steppe Iranian words carrying the same meaning as the suffixed form: for instance *etswe* 'mule' as if from *atᶴu̯a-tara- 'mule' and not *atᶴu̯a- 'horse' and *kertte* 'sword' as if from *karta-tara- 'sword' and not *karta- 'knife' (nonetheless, see a different scenario p. 34 and p. 197-198). These examples are discussed in section 2.6.4.f of the present chapter. A form *abi-i̯ā- would thus perhaps stand for *abi-i̯āta- in Old Steppe Iranian. A simplification *epiyate-c > *epiyatc is much more specu-

lative, as it has no parallels in our corpus. To support the proposal of an al-lative, Dr. Federico Dragoni (p.c.) informs me that Khotanese *byāta* 'memory' is strictly used in collocation with *yan-* 'to do', and that it is generally assumed that it is originally a frozen instrumental.

To conclude, the Proto-Tocharian word **epiyac*, TB *epiyac*, *epyac*, TA *opyāc* was borrowed from Old Steppe Iranian. The etymology is evident, and it has many cognates among Iranian languages. However, the final *-c* of the Tocharian forms remains difficult to explain.

2.2.3 TB *etswe* 'mule'

The Tocharian B word *etswe* was first identified and recognized as a loan-word from Iranian by Peyrot (2015: 222-23), and then discussed extensively in Peyrot (2018a).[3] There is no doubt that this word must be a borrowing, as the Tocharian reflex of the same Indo-European etymon (**h₁eḱu̯o-*) is found in TB *yakwe*, TA *yuk* 'horse'. It was already known that Proto-Iranian **ć*, **j́* were reflected as **tˢ*, **dᶻ* in Old Steppe Iranian, see e.g. TB *entse* or TB *tsain*. Since *etswe* must derive from Old Steppe Iranian **atˢu̯a-* < PIr. **aću̯a-* < PIE **h₁eḱu̯o-* 'horse' (cf. TB *yakwe*, TA *yuk* 'horse', cf. Peyrot 2015: 223), this borrowing shows that the Proto-Iranian cluster **-ću̯-* was reflected as **-tsu̯-* (see Peyrot 2018a: 271-72). In Proto-Khotano-Tumshuqese the cluster **-ću̯-* remained as such, and is regularly reflected as *-śś-* in Khotanese, as in Khot. *aśśa-* 'horse' (DKS: 11). It is notable that the Iranian word for 'horse' was borrowed by other languages as well, for instance Ugaritic *s*, *śs*, *św* m. 'horse' (cf. Del Olmo Lete & Sanmartín 2003: 772).

If, as Peyrot suggests (2018a: 271), the word was borrowed from Old Steppe Iranian together with a horse or a horse-like animal, it seems that the speakers of Old Steppe Iranian had a different kind of horse than the "inherited" horses that the speakers of Proto-Tocharian called **yək^we*. Or else, the word should have undergone the semantic specialization within Tocharian that caused it to be rendered in Uyghur as *katır* 'mule'.

I would like to offer another possible explanation for the meaning of Tocharian B *etswe* 'mule'. Given that the meaning of *kertte* 'sword' rather corresponds to that of a hypothetical **kertetere* 'knife-like; sword' (or OSIr. **karta-tara-*), it is possible that the meaning of *etswe* 'mule' was extracted

---

[3] If the word were attested in Tocharian A, which it is not, we would expect †*atsu*.

from that of *etswetere (or rather Old Steppe Iranian *atˢu̯atara-) which would have meant 'mule' (cf. Khwarezmian ʾsptyr, and New Persian astar, both meaning 'mule' < *atˢu̯a-tara-, and naturally also Sanskrit aśvatara- 'mule'). Whether the suffix was deleted in Old Steppe Iranian, or in Tocharian, is an open question for the time being.

In conclusion, Tocharian B etswe 'mule' is a borrowing from Old Steppe Iranian *atˢu̯a- 'horse', or perhaps from *atˢu̯a-tara- 'mule', the suffix of which is not reflected in the Tocharian B form.

2.2.4 TB ainake, TA enāk 'common (vulgar); evil, bad'

It has long been correctly claimed that Tocharian B ainake and Tocharian A enāk 'common (vulgar); evil, bad' are of Iranian origin. Hansen (1940: 146) saw a connection of this word to MP ‹ʾynykyh› ēnīkīh 'Gewalttat'. The Middle Persian word, which is also cited as a cognate by Schmidt (1985: 766[25]), is not in the main Middle Persian dictionaries. It is, however, discussed in Bartholomae (1906: 93-94), who proposed the reading ēnīkīh. I could find it this form in F10 (Vd. 2:50):
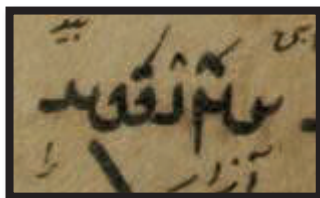
As it appears, the word should be transcribed as ‹ʾyndgyh› (because of the diacritic over the ‹y/d›). It is translated as آزار (āzār) 'great pain, great annoyance' in Persian below, and it notably translates Avestan axti- 'disease (or pain)' in the phrase nōit axtiš nōit mahrkō (Vd. 2:5) "neither disease/pain, nor death". A similar transcription is needed for other manuscripts, such as G10 (‹ʾyndkyh›, ‹ʾyrdkyh› or ‹ʾywdkyh›) Even if the meaning were 'evil' and the word ēnīgīh, it would not come regularly from Proto-Iranian *ai̯naka- vel sim. In fact, besides assuming a stem *ai̯ni- suffixed with -ka-, there would be no way to explain ēnīgīh. In any case, the problems surrounding this Pahlavi word are too complicated to let it be used in the discussion surrounding the Tocharian words.

Tremblay (2005: 425) proposes to connect Tocharian B *ainake* and Tocharian A *enāk* to a Middle Iranian proto-form *$a\underset{\smile}{i}naha$-ka-. However, a formation such as *$a\underset{\smile}{i}naha$-ka- cannot be Middle Iranian: what Middle Iranian language would have preserved such an archaic form? For instance, let us look at both the preservation of -*aha*- (even though probably contracted to *$\bar{a}$ in the Iranian source word) and the word-final -*ka* (> -*ke*), neither of which are preserved in any known Middle Iranian language. Also, Tremblay claims that the phonetics of the Tocharian forms are clearly Proto-Tocharian, that is, that borrowing occurred at the Proto-Tocharian period (Tremblay 2005: 425). This is chronologically impossible, as there are no parallels for Middle Iranian loanwords in Proto-Tocharian.

Isebaert (1980: 115) proposes to derive *ainake* from OIr. *$a\underset{\smile}{i}nāka$- 'misdadig, gemeen, lasterlijk'. Although this would work phonetically, it is only a transposition of the Tocharian form back into Old Iranian, with no evident basis in Iranian. Schmidt (1985: 763) proposed to connect the Tocharian words to Avestan *aēnah*- m. 'Übeltäter', similarly to Isebaert, and considered rightly that the preservation of the $a\underset{\smile}{i}$- diphthong indicated a borrowing of Old Iranian age. For the meaning 'evil', a derivation from the neuter stem is possibly preferable.

The preform posited by Isebaert is a plausible etymon. Indeed, since the Iranian form is too archaic to be linked to any known or unknown Middle Iranian language, it seems sound to ascribe the origin of the Tocharian words, which go back to Proto-Tocharian *$eynake$, to an Old Steppe Iranian form *$a\underset{\smile}{i}nāka$- < Pre-OSIr. *$a\underset{\smile}{i}naha$-ka-, cf. CSogd. ʾ*yn*ʾ*qwc* adj. 'abusive' (cf. Sims-Williams 2020: 45). A contraction of *aha to *$\bar{a}$ would explain the remarkable *$\bar{a}$ in the Old Steppe Iranian form, for which Isebaert had offered no explanation.

This word is particularly interesting because it is, to my knowledge, the only secure ancient *-*ah* stem (like *perne*, *entse*) among Old Steppe Iranian loanwords in Tocharian that is suffixed with -*(a)ka*.

Iranian languages have two strategies with regard to -*ka* suffixation of *h* stems: *ah- + -*aka*- > -*aka*-, like in Avestan and most Iranian languages (e.g. Avestan *vaβžaka* 'wasp' ← *$vaβžah$- < *$\underset{\smile}{u}eb^{h}$-*so-s*), and *$ah$-*a*-*ka*-, as is reflected in Sogdian (e.g. *sāk* 'number' < *$sahaka$-, SD: 453), and a few other languages, such as Yidgha-Munǰī (e.g. *sīγ* 'hare' < *$sahaka$-, cf. Morgenstierne 1938: 57). These two strategies are morphological in nature, rather than phonetic, and thus relevant in order to determine isoglosses within

the Iranian phylogenetical tree. The Iranian source language of this Tocharian form was closer to Sogdian, Yidgha-Munǰī, and a few others, in this respect, than to the rest of the Iranian languages.

Since *aịnaha-ka- would have been rendered in Proto-Tocharian as *eyne'eke > TB †aineke, it appears that the change *aha > *ā (then borrowed as PT *a) happened in the source language of the borrowing, not in Proto-Tocharian. In other words, there is no contraction of e.e to a in Tocharian, while a.a in Iranian would certainly yield ā, as in the examples cited above from other Iranian languages. For a more detailed discussion on this sound change, see section 2.6.2.c of this chapter.

Another interesting point concerning these words is that the archaic form of Tocharian B *ainake* is *eynāke* (cf. Peyrot 2008: 58). Together with *meyyā* (s.v. *maiyya*) and *krewpe* (and, to some extent, *newiya*), this shows the rendering of Old Steppe Iranian *a + yod or waw. Apparently, the Old Steppe Iranian diphthong sounded to Tocharians like *e (= OSIr. *a) + yod or waw rather than like the native Tocharian diphthong *ay.

In *ainake* we can also notice that it had initial stress, i.e. /áynake/, which is characteristic of Old Steppe Iranian loanwords, see section 2.6.2.g of this chapter. However, if we take Vedic *énas-* 'crime, mischief' into account, the initial stress in this word might be old.

One problem remains: the Tocharian A form *enāk* did not undergo vowel weakening, the rule being PT *ay _ *ā > TA e_a (see for example Pinault 1989: 45; Kim 2007: 1). If vowel weakening had applied, one would expect a Tocharian A form †*enak*. One can propose, very cautiously, that TA *enāk* was, in fact, borrowed from the supposed Sogdian form *ēnāk**, which is not attested as such, but can be deduced from its derived form, ʾynʾqwc 'abusive' (cf. Sims-Williams 2020: 45). This assumption would explain the absence of vowel weakening in the TA word. I have made a similar proposal for TA *paräṃ* 'glory, rank' below, in order to explain the lack of *p/m*-umlaut of the *a* of the first syllable, which did occur in *porat* 'axe' < *peret*, as well as in inherited words, such as TA *ñom* 'name' < *ñam*.

In conclusion, Tocharian B *ainake* 'common (vulgar); evil, bad', Archaic Tocharian B *eynāke*, derives from Proto-Tocharian *eynake* 'id.', itself a borrowing from Old Steppe Iranian *aịnāka-, with short *a being rendered as PT *e and long *ā as PT *a. Tocharian A *enāk*, though, might either derive from the same source-word, or from Sogdian.

2.2.5 TB *-aiwenta* 'units of'

It is difficult to translate Tocharian B *-aiwenta*. I translated it in the title as 'units of', but it is in fact an individualizing plural suffix. For the translation of this suffix, see Winter (1962).

In Tocharian B texts, it is found in various words, such as:

THT 368 a4
tänwaṃñana eśaiwentasa
"with lovely eyes"

IOL Toch 188 b1
ṣpä śle yasar misaiwentasa
"and with blood and pieces of flesh"

PK AS 16.1 a5
naś kwalyiṃne yokaiwenta tesare
"... they put the individual pieces of hair in the *kwalyiye**"

It is also found with the word for house/habitation: *ostuwaiwentane* (multiple texts), which can be translated as "in the individual groups of houses".

Since *misaiwenta* is built on *misa*, formally a plural, Winter (1962: 116) notes that it is likely that *ostwaiwenta* is built on the plural stem *ostwa* and not on the singular *ost*, which would have yielded **ostaiwenta. He thus translates it as "several groups of houses" and, accordingly, translates *eśaiwenta* as "[many] individual pairs of eyes". He rightly corrected Krause's analysis of the suffix as "-*īwe*-": *eśaiwenta* for instance clearly shows the morphological division to be *eś-aiwenta* and not *eśa-iwenta.

Krause (1954) explained the suffix (which, according to him, was used for a "Plurativ" number of Tocharian)[4] as deriving from PIE **Hoi̯u̯o-* 'one, alone': "Hat man das hieraus zu erschließende Element *īwe* vielleicht etymologisch mit idg. **oi̯u̯o-* „ein", „allein" zu verknüpfen?" (1954: 6[1]). The second element is the Tocharian B plural suffix *-nta*.

---

[4] Winter (1962: 117) argued against a "plurative" as a morphological category of inflection in Tocharian. In my view, he is right, as data is too scarce to indicate a real paradigmatic morphological use of the suffix *-aiwenta*, which only occurs in a limited number of texts and with a limited quantity of nouns.

No forms ending in *-aiwe are attested, and it seems that there was only a plural -aiwenta in Tocharian B. However, given that plurals in -nta were very productive, it is possible that the suffix -nta was added much later to a suffix *-aiwe which was, of itself, individualizing.

Since it is also added to the plural stem of ost 'house', ostuwa, it is possible that the addition of the suffix complex -aiwenta compensated for weakening of the plural meaning of those words. For instance, *ostaiwe may have meant 'a single house' → *ostwaiwe 'a group of single houses'. Then, because the plural meaning of *ostwaiwe was lost, or perhaps to insist on their quantity, the suffix -nta was added again. This type of constant reinforcement is very common in living languages (Fr. hui 'today' → aujourd'hui 'on the day of today' > 'today' → colloquial au jour d'aujourd'hui 'today' < 'the day of the day of today').

As to the etymology, this suffix could theoretically be inherited, as proposed by Krause (cf. also Pinault 1988: 202). However, of all Indo-European languages, only (Indo-)Iranian and Greek have forms that go back to *Hoi̯u̯o-. In Sanskrit it became a particle, eva. Furthermore, it is only in Iranian that *Hoi̯u̯o- has become a productive suffix, found in several languages such as Middle and New Persian, Balochi, some Eastern Iranian languages, etc.[5] It would seem very coincidental that this suffix also arose in Tocharian, which has not inherited *Hoi̯u̯o- as an independent word, without at least influence from some Iranian languages.

Pinault (1988: 201) has established the meaning of a new form iwār: 'at once', which he translates as "'aussitôt', 'd'un seul coup, tout à coup' ('auf einmal')", etymologically constituted of *iw- and of the adverbial suffix -ār (1988: 201f.). He wondered if *iw- could come from "*(e)i̯u̯o-" while *aiwe would derive from *Hoi̯u̯o- or from *-a-iwe, the latter hypothesis being disproved by Winter (1962, see above).

Pinault's etymology is difficult formally, as there is no other evidence for ablaut in *Hoi̯u̯o- in Indo-European. Pinault (p.c.) no longer believes this etymology: indeed, the suffix of distributive numerals -ār (in fact *-arə) can no longer be considered a cognate of the adverbial suffix -ār (< *-ōr), cf. Pinault (2008: 560-61).

---

[5]   On this suffix in Middle Persian, see Josephson (2011).

There is thus no other trace of PIE *$Hoi̯u̯o$- in Tocharian. Because of this, and because *$Hoi̯u̯o$- is rather rare in Indo-European, being only found in Indo-Iranian and Greek, it seems to me that -*aiwenta* can more straightforwardly be derived from Old Steppe Iranian *-*ai̯u̯a*- by means of borrowing. This *-*ai̯u̯a*- could have been an individualizing suffix, possibly also indicating indefiniteness, as in Persian.

The word *$ai̯u̯a$- has many cognates among Iranian languages, at least one in every language, and is the basic word for 'one' in Proto-Iranian. It became a suffix for instance, in Middle and New Persian (Josephson 2011: 25). This suffix possibly initially indicated individuality, but in attested texts, it is an indefinite suffix, thus not entirely similar in function to the one in Tocharian.[6] However, both the individualizing and the indefinite function can derive from a primary meaning 'one' or, secondarily, 'single'. It is, however, remarkable that Proto-Iranian *$ai̯u̯a$- seems to have become a suffix in Old Steppe Iranian too. Nonetheless, I do not think this grammaticalization process can be used as an argument for the classification of Old Steppe Iranian among Iranian languages, as it is possible that it occurred independently in various Iranian languages.

In conclusion, the Tocharian individualizing suffix *-*aiwenta*, composed of *-*aiwe* and of the plural suffix -*nta*, has been viewed as deriving from Proto-Indo-European *$Hoi̯u̯o$- 'single; alone'. However, this Indo-European word is rare, being only found in two branches: Indo-Iranian and Greek, and being otherwise completely absent from Tocharian. Because of this, I suggest that this suffix originates from Old Steppe Iranian *-*ai̯u̯a*-, with a probably similar if not identical meaning.

2.2.6 TB *kertte* 'sword'

The Tocharian B word *kertte* 'sword' was recognized as being of Iranian origin for the first time by Van Windekens (1963: 486). The Iranian etymon

---

[6] There also exists a Persian individualizing plural suffix, namely Contemporary Persian: -*ihå* (< Cl. Pers. -*ē-hā*), which has the same function as the Tocharian suffix -*aiwenta*, and which is also built on *-*ai̯u̯a*- (> -*ē* > -*i*) + plural suffix -*hā*, e.g. *åb-i(-)hå* 'individual bodies of water' (individual waters + plural).

is clearly *karta-, cf. Avestan karəta- 'knife', but there has been discussion about the exact source language of the borrowing.[7]

Winter (1971: 218) has claimed that the source was Bactrian, but Schwartz (1974: 409) has shown that this is not possible. Bactrian does not preserve (short) word-final vowels of Old Iranian in polysyllabic words, which proves problematic to explain the final -e of kertte. One could theoretically oppose that Pre-Bactrian loanwords may display a final vowel in Tocharian, as in the case of mālo and koṣko (see Del Tomba 2020: 126; Bernard & Chen 2022). However, no correspondence with final -e is found thus far, and for sure Tocharian loanwords from Bactrian do not show e for Bactrian a internally, so that the first syllable of kertte certainly cannot be explained from Bactrian. Last but not least, the Bactrian word for 'sword' is χαγγαρο (Sims-Williams 2007: 276),[8] and the meaning 'sword' of the Tocharian word is well assured from the attestations, while the bare Old Iranian noun *karta- generally means 'knife' in most Iranian languages.[9] So, to sum things up, neither the phonetics, nor the documentation of Bactrian, nor the semantics of kertte support Winter's hypothesis of a Bactrian borrowing.

Once Bactrian has been eliminated, it is clear that this word must be from another Iranian dialect. No other Middle Iranian language seems to fit with the form (cf. Sogdian ākartē 'sword', Khotanese kāḍara- 'id.') and the vocalism e_e indicates a likely Old Steppe Iranian origin. An Old Steppe Iranian *karta- would regularly have yielded Proto-Tocharian *kerte. The To-

---

[7] Some scholars have considered that kertte could be inherited from Indo-European (Isebaert 1980: 89), but this is improbable. Although an Indo-European form *kor-to- would indeed yield TB kert(t)e, no other Indo-European language, not even Indic, reflects such a form. Indeed, as Professor Lubotsky informed me (p.c.): the nomen agentis *kortó- 'cutter (?)' is never found outside of Iranian. In his forthcoming etymological dictionary of Proto-Indo-Iranian, Professor Lubotsky reconstructs *krti- 'knife' for Indo-Iranian, and considers *kartá- as a properly Iranian innovation.

[8] Compare Christian Sogdian xnγr (var. xγr) 'sword' (Sims-Williams 2020: 216).

[9] For example in THT 79 a2 (sto)rmeṃ kertteṃ oṅkor malkānte śle-yärke lāntaś weskeṃ "(stand)ing they put [their] swords [in the] sheaths [and] speak with deference to the king" (my translation, on the basis of CEToM); THT 404 b8 wes rano ñake kerteṃ yamamtär ścirona ṣñārä "We will now also make the swords of each of us sharp." (Peyrot 2013: 658).

charian A outcome should have been †*kart*, but apparently this lexeme was replaced in Tocharian A by *kāre* (see below).

One needs to explain the discrepancy in meaning between Proto-Iranian **karta-* 'knife' and Tocharian B *kertte* 'sword'. There are two solutions, and I will account for both in the present work. One, would be to suggest that the meaning derives from a suffixed form **karta-tara-* 'knife-like, sword' at the source of, for example, Khotanese *kāḍara-* 'sword' (on the derivation of *kāḍara-* from **karta-tara-*, see directly below). This **-tara-* suffix would have either been removed in Old Steppe Iranian, while the meaning remained, or Tocharian speakers would have borrowed the unsuffixed form with the meaning of the suffixed form (on this see section 2.6.4.f of the present chapter).

The other solution consists in looking at the Ossetic evidence: Ossetic has a word *kard* (‹a› in Ossetic stands for a long vowel, *ā*) from **karta-* with vowel lengthening. This word means 'knife; sabre; [epic] sword' (cf. Cheung 2002: 196), and it thus appears that Ossetic is the only Iranian language with a meaning 'sword' for the unsuffixed word **karta-*. Naturally, one could suggest that the hypothesis with **-tara-* deletion would apply to Ossetic as well, but this is unlikely, due to the meaning 'knife' that remains primary. Given the quantity of similarities between Old Steppe Iranian and Ossetic, it is not unnatural to assume a meaning 'knife; sword/sabre (in any case a cutting war weapon)' for the ancestor of Ossetic *kard* and Old Steppe Iranian **karta-*. Interestingly, this would show that Ossetic **ard > *ārd* is posterior to the separation of both languages. Further, the Old Steppe Iranian could also have meant 'knife', but the Tocharians, who were probably either under their rule or in any case in a type of martial relationship towards them, would have borrowed the meaning 'sword' exclusively.

In conclusion, Tocharian B *kertte* 'sword' derives straightforwardly from Old Steppe Iranian **karta-* or **karta-tara-*.

A note on TA *kāre* 'sword'

No etymogical cognate of Tocharian B *kertte* is so far attested in Tocharian A. It would have been, to all probability, †*kart*. Instead, the Tocharian A word for 'sword' is *kāre* (pl. *kāreñ*) 'sword' (see Carling 2009: 116), whose etymology is not known, to my knowledge. I suggest that it derives from

Khotanese, where the word for 'sword' is *kāḍara-* (DKS: 58). This, in turn, ultimately derives from PIr. *karta-* too.

Two different pathways are available to explain the Tocharian A word on the basis of the Khotanese. An enlarged form *kāḍaraka-* would have undergone a weakening of the medial syllable: *kartaraka-* > *kāḍaraa-* > *kāḍäraa-*, which would have yielded *kāḍärei/*kāḍärai in Old Khotanese.[10] This could have been borrowed as *kārāre in Tocharian A, then becoming *kāre* through syncope.

For *-ḍ-* → -r-, a parallel development can be found in TAB *kor* (besides TB *koṭ* and TAB *koṭiśvar* 'millionaire') 'ten million' from Middle Indic *koṭ* 'idem', among other examples.

Another solution is possible, albeit more speculative. Khotanese *kāḍara-* 'sword' has generally been explained as deriving from Proto-Iranian *karta-* 'knife, dagger' with an unexplained suffix *-ra-* (e.g. DKS: 58). Another etymology seems more reasonable to me: *kāḍara-* should derive from *karta-*, suffixed with *-tara-* 'sort of', 'a sort of dagger' > 'a sword', and haplology (*karta-tara-* > *kartara-*). Tocharian A *kāre* could theoretically also derive from the unsuffixed (and unattested) form *kāḍa-* < *karta-*, with the addition of a *-ka-*suffix: *kartaka-* > OKh. *kāḍaa-*, with OKh. nom.sg. *kāḍei/*kāḍai → (regularly) TA *kāre*. Unfortunately, these forms are completely unattested in Khotanese, and the first hypothesis is thus much more likely, but this second hypothesis is formally more straightforward.

In conclusion, I propose that Tocharian A *kāre*, pl. *kāreñ* is a Khotanese loanword, which ultimately either goes back to Khotanese *kāḍara* 'sword', suffixed with *-ka-*, or to an unattested Khotanese *kāḍa-*, also suffixed with *-ka-*. The weakening of the ending *-aka-* to *ei* or *ai* already in Old Khotanese led to the final *-e* of the Tocharian form, while the Khotanese *-ḍ-* was rendered as Tocharian A *-r-* as in Middle Indic loanwords.

2.2.7 TB *keṣe*, TA *kaṣ* 'arms' length (measurement unit)'

Tocharian B *keṣe* and Tocharian A *kaṣ* 'arms' length (measurement unit)' derive from Proto-Tocharian *keṣe*, which was borrowed from Old Iranian *kaša-* (cf. Av. *kaša-* 'armpit') 'arms' length' (cf. Isebaert 1980: 84-85;

---

[10] I thank Federico Dragoni for confirming these facts to me.

DTB: 213). The word must have been borrowed from from Old Steppe Iranian in view of the vowel correspondence OIr. *a : PT *e.

This word *kaša- could already have designated a measurement unit, as semantic convergence of body part-terms and measurement units is extremely common cross-linguistically. The semantic change from 'armpit' *vel sim.* to 'arms' length' could have occurred in Old Steppe Iranian itself, as a measurement unit is much more easily borrowable than a body-part term. However, given how essential words (such as TB *epe* 'or', s.v.) were borrowed from Old Steppe Iranian into Proto-Tocharian, it is not impossible either that the semantic shift occurred within Proto-Tocharian itself.

In conclusion, Tocharian B *keṣe* and Tocharian A *kaṣ* derive from Proto-Tocharian *keṣe* 'arms' length', itself a borrowing from Old Steppe Iranian *kaša- 'arm' probably with a meaning 'arms' length' (the length of two arms stretched together, Dutch *vadem*, English *fathom*) already present in Old Steppe Iranian.

2.2.8 TB *kraupe*, TA *krop* 'group, crowd'

The Tocharian B word *kraupe* (Archaic TB *krewpe*) and its Tocharian A equivalent *krop* 'group, crowd' are usually viewed as deriving from the verb TB *krawp-*, TA *krāwp-* 'gather, amass' (so Adams, DTB: 238) < PT *krawp- (cf. Peyrot 2013: 741). However, it appears that this verb does not have a convincing etymology.

The etymology of this verb as given in Adams is unsatisfactory: he claims cognacy of this word with Greek κρύπτω 'hide', but there is no evident semantic link between "hiding" and "gathering". Obviously, one can pile things up onto something in order to hide that thing, but the semantic stretch is too far in my view.

According to Adams (also Hilmarsson 1996: 179), *kraupe* is further cognate to Lithuanian *kráuju* 'pile up', *krūvà* 'pile', Old Church Slavonic *kryjǫ* 'cover, hide' *krovъ* 'roof'. However, it is doubtful whether any of these words can be connected to Tocharian B *kraupe* and Tocharian A *krop*, simply because we cannot account for the *-p-* in Tocharian. Given that the form is different, and the semantics are far from evident, it can be concluded that the cognates in the set proposed by Adams are most probably not related.

Tocharian B *krewpe*, *kraupe*, Tocharian A *krop* 'group, crowd' is more easily derived, both for the meaning and for the form, from Old Iranian

graṷpa- (or *graṷfa-*). Neither this etymology nor the Old Iranian word have hitherto been recognized. The basis for the reconstruction of this Old Iranian word is found in a few Iranian languages, namely, Middle Persian, New Persian, Balochi and Dawāni.

The facts are these: Middle and New Persian *grōh* 'group, company', borrowed into Armenian *groh* 'Nation, Volk, Truppe' (Hübschmann 1897: 132); Balochi *grōp* 'assembly, group'; Middle Persian *grōhag* seen in the Syriac borrowing ‹grwhqʾ› 'small round cakes' (cf. Ciancaglini 2008: 146), and New Persian *gurōha* 'globe, bowl, or any spherical figure; ball for a cannon; a kind of sweetmeat; an assembly, multitude of people' (cf. Steingass 1892: 1085); Dawāni *gorufa* 'ball of fiber or rope'.

Multiple etymologies were provided for these words, none of which is satisfactory.[11] Balochi *grōp* can go back to either *graṷpa-* or *graṷfa-* (cf. Korn 2005), but the forms with -h can only reflect an earlier *-f, and it is difficult to accept that the Dawāni form *gorufa* shows a sound change *-p- > -f-, although more information about this language would be welcome.

There is a difficulty with the form with -h- in Middle and New Persian: -f only becomes -h in New Persian, for example: Middle Persian *kōf* 'mountain' > New Persian *kōh* 'idem'. Here the -h is ancient, as shown by the Syriac borrowing ‹grwhqʾ› (see above). We thus have no other choice but to

---

[11] See Hasandoust (2014: 2385f.) for an enumeration of them. One can cite Horn's very cautious proposal of connecting this word to Germ. *Leute* etc., through perhaps a hypothetical *ṷi-raṷda-* (Horn 1898-1901: 186) which is impossible because of the *g-* in the Middle Persian form (*ṷi-* only becomes *gu-* in New Persian, and should have remained as *wi-* in Middle Persian). Furthermore *ṷi-raṷda-* is not fitting semantically: in Sanskrit *virodha-* means 'opposition, enmity, struggle'. Another attempt, by Nyberg (1931: 84), consisted in setting up a proto-form *gravaθva-* which, besides being completely *ad hoc*, does not explain the Balochi and Dawāni forms. Monchi-Zadeh (1990: 79) proposes a proto-form in *-θ-* (namely *graṷθa-*), which he derives from the Indo-European root *gel-* 'to group, to amass' (the root at the basis of Eng. *cloud*). This proto-form *graṷθa-* is obviously impossible, for the reasons evoked in this discussion. Even if one were to accept the implausible derivation from Indo-European *gel-*, Proto-Iranian *graṷpa-* or *graṷfa-* would still require an opaque suffixation in *-pa- or *-fa-.

assume borrowing from a variety that underwent the sound change *-f > -h* earlier than "Standard Middle Persian".

No matter whether it was *graṷpa-* or *graṷfa-*, my hypothesis is that an Old Iranian antecedent of the forms cited above was borrowed in Proto-Tocharian as *krewpe*, hence Tocharian B *kraupe* and Tocharian A *krop*. A weak point in my hypothesis is the limited attestation of the word in Iranian in combination with the lack of a cognate beyond Iranian, i.e. in Indo-Aryan or in Indo-European. We should thus assume that the word entered Iranian at some point, likely by means of borrowing, but we cannot be sure about the dating. Since the semantic and phonological correspondences with Tocharian are exact, with Old Iranian *aṷ* being reflected as Proto-Tocharian *ew* as in TB *newiya* ← Old Steppe Iranian *naṷiịā-* (and cf. further OSIr. *aị* → PT *ey*), I assume that the word has a longer history within Iranian and was borrowed into Tocharian from Old Steppe Iranian. [12]

The fact that the attestation in Iranian is limited may be explained by the circumstance that the lexicon of Middle Persian is so much better known than that of other varieties. The same applies to OSIr. *rataka-* 'army line', which is also only known from Middle Persian.

In Tocharian, a verb was derived from the noun *krewpe*, as for example, PT *rǝytwa-* was derived from PT *reytwe* (s.v. *raitwe*). The reconstruction of the verbs in TB and TA (cf. Peyrot 2008: 150; Peyrot 2013: 741), with two different presents, fits very well with the verbs being derived from the Proto-Tocharian noun *krewpe*. For instance, Tocharian B *kraupe* would not easily be derived from a stem *krawpa-*, etc., especially in view of the fact that the older vocalism of the noun is *-ew-*. This *-ew-* may have been subject to *a*-umlaut in stems like *krawpa-*, but the reverse development is not possible.

In conclusion, the Tocharian B *kraupe*, Tocharian A *krop* words meaning 'group, crowd' are straightforward borrowings from Old Steppe Iranian *graṷpa-* or *graṷfa-* (more likely *graṷfa-*) which had the same meaning. The verb TB *krawp-*, TA *krāwp-* is derived, within Tocharian, from the Proto-Tocharian noun *krewpe*.

---

[12] Since there are no cognates of this word outside of Iranian, we should obviously reject the option that the Tocharian and Iranian words are both inherited from Proto-Indo-European, instead of being related by borrowing.

2.2.9 TB *newiya* 'canal or channel'

TB *newiya* 'canal' is attested only twice, in the passage I cite below. It is a contract for the sale of a land (Pinault 1998: 364), and this contract discusses the boundaries of that land.

Tokyo National Museum, Shiryokan Archives n. 174 4-5[13]

> *omotruññaiṣṣe yateññe ckeṣṣe ārte sim kom kläskomeṃ orotsa newiya sim oṣṣalemeṃ armokiññe ckeṣṣe ārte sim orotsai newiyai täṅtsi*
> "of the south, the boundary is the arm of the river Yāte; on the west the boundary is the great canal; on the north, the boundary is the arm of the river Armoki up to the great canal inclusively." (Pinault 1998: 364-365)

The meaning 'canal' is probably based on the likely etymology of this word. Adams (DTB: 364) derives it from Old (Steppe) Iranian **nau̯ii̯ā-*, which is shortened from PIr. **nāu̯ii̯a/ā-* according to the sound law discussed section 2.6.2.i (p. 175f.). The original meaning of the Iranian formation is disputed, see for instance a review of the literature in Utas 1965-66; more recently see Filippone 2017, with a review of the more recent literature as well as a thorough analyses of relevant passages – she concludes that the precise semantics of the Avestan cognate *nāuuiia-* have not yet been securely determined.

As per Filippone 2017, one can see that there are two possible ways to understand Proto-Iranian **nāu̯ii̯a/ā-*. It can be analyzed either as meaning 'navigable (canal)' or 'channel'. Either meaning could be ascribed to TB *newiya*. One can hope that further attestations of this word in Tocharian might help us determine the meaning of the Old Iranian cognates with more exactness. It seems to me, in any case, quite straightforward to see TB *newiya* as the Old Steppe Iranian feminine gender substantivization of an adjective meaning **'*relative to *nau̯ii̯a-* rivers', the word for 'river' being a feminine noun, thus 'navigable (river)' (see for instance Utas 1965-66: 127f.) or 'channel' (cf. Filippone 2017 for references to previous literature). This 'navigable (river)' or this '(navigable?) channel', made into a noun, could have have designated canals, as they are (artificial) navigable rivers, made precisely for the purpose of navigation. As I wrote above, the meaning of

---

[13] Alternatively known as Otani 19.1.

this word and related words in Old Iranian is extremely difficult to ascertain, however, it is beyond doubt that they designate running waters, and it is also beyond doubt that the first syllable contains – originally in any case – a long *$\bar{a}$.

Although the derivation of TB *newiya* from Old Steppe Iranian seems straightforward, an alternative interpretation has been proposed by Widmer (2007). He also sees TB *newiya* as a borrowing from Iranian, but his explanation differs in crucial points. According to him, the meaning of Old Persian *nāviya-* is not connected to the notion of 'boat' but to that of 'flowing'. He supports his view with the Akkadian translation of the Old Persian passage, which has "the Tigris river was in flood". He also finds the meaning '(stark) fliessend, strömend' fitting for the Avestan cognates. Widmer (2007: 224) also mentions the Avestan word *nāuuiia-*, for which he proposes the possible meanings 'Fluss, Bach'.[14]

Widmer argues that the source of the Tocharian borrowing was a Middle Iranian form *$neu̯(i)i̯a$. In my opinion, this is impossible, since it presupposes an *ad hoc* metathesis of the original *$nāu̯i̯a$- to *$nāi̯u̯a$-, followed by a restoration of a suffix *-(i)i̯a*, which was no longer available in Middle Iranian, as far as I know. Also, I do not know any Middle Iranian language where *-$āi̯$- would become *$\bar{e}$ or *e, as usually OIr. *-$āi̯$- remains as $āy$.

Here, the data offered by Tocharian clearly contradicts Widmer's hypothesis, and the traditional meaning 'navigable' usually found in scientific literature is to be upheld. Furthermore, Akkadian and other versions of the text are not literal translations of the Old Persian, but simple different versions written in different languages.

The word *newiya* is of special interest since it seems to show that shortening of PIr. *$\bar{a}u̯$ to *$au̯$ occurred in Old Steppe Iranian (see 4.2.i, p. 175f.). Moreover, it gives a precious insight into the culture of the people speaking Old Steppe Iranian. Apparently, these people knew what canals were, and what navigation was. Did they practice it in the Balkhash Lake, which is far south, but one of the closest lakes nowadays in the region, or in another, smaller body of water? Or, perhaps, was there much more water in the re-

---

[14] Widmer (2007) does not cite Utas' seminal work of 1965-66; where the translations of the Old Persian passage are examined, and where the meaning of *nāviya-* is discussed.

gion?[15] Naturally, the Tocharian texts reflect a different sort of civilization from the original Old Steppe Iranian one, and the original meaning of *newiya* might have been different from the English definition of 'canal' – if we do not take it to have meant 'channel', as many scholars did for **nau̯ii̯a-*.

In conclusion, Tocharian B *newiya* 'canal or channel' is borrowed directly from Old Steppe Iranian **nau̯ii̯ā-*, a substantivized feminine adjective of a feminine noun 'river', meaning 'thing relative to navigation; canal', with shortening in the first syllable from PIr. **nāu̯ii̯ā-*.

2.2.10 TB *pāke*, TA *pāk* 'part, portion, share'

The Tocharian B noun *pāke* 'part, portion, share' (derived noun *pakeññe* 'member, partner'), plural *pakenta*, Tocharian A *pāk* 'idem', reconstructible as Proto-Tocharian *pake 'id.', has been considered as either of Iranian origin or inherited (DTB: 389). In both Tocharian A and B there is a set phrase *pāke yam*- 'to participate in' (cf. Meunier 2013: 149f.).

In order to determine whether it is more likely that TB *pāke* and TA *pāk* are inherited or borrowed, I will first present the generally admitted Indo-European etymology of those words, and then discuss their Iranian etymology.

I will present below the two main Indo-European reconstructions which have been proposed in order to derive Tocharian B *pāke*, Tocharian A *pāk* from Proto-Indo-European. The first one is exemplified in Pinault (2008), and the second one in Adams (DTB).

Pinault (2008: 30) considers both B *pāke* and A *pāk* 'part, portion, share' as straightforwardly derived from Proto-Indo-European *bʰag-os "de la racine **bʰag-* « (se) partager » (LIV², p. 65)". This is naturally only possible if one accepts the existence of a vowel *a in Proto-Indo-European. It is to be noted that the reconstruction with *a in the LIV²: 65 is based on the Tocharian forms. The reasoning here could be seen as circular: one reconstructs a

---

[15] The Elamite phrase *kurtaš numakaš* could have designated irrigation workers (very hypothetically, see Henkelman 2018: 233 with references to earlier literature). Possibly, if this were true, those equally hypothetical Tocharian *kerciye 'house servant' (s.v. *kercci*) could have been building canals for the Old Steppe Iranians. This, however, implies a type of society and empire that we can only hypothesize for the Old Steppe Iranians.

proto-form in order to explain two Tocharian words, which are themselves only explained through that reconstruction.

The second explanation can be found in Adams (DTB: 389), who suggests among others to derive it from *$b^heh_2góm$ (or *b^heh_2go-), while noting a problem with the -*nta* plural. Indeed, as Adams rightly notes, one would expect a plural **$p\bar{a}ki$ if the word was inherited from Proto-Indo-European *b^heh_2go-, and not *pakenta*. However, there is yet another problem with this reconstruction: *b^heh_2go- or *$b^heh_2góm$ would yield TB †*poke* and TA †*pak*. Adams (DTB: 389) does not completely exclude a borrowing from Iranian.

The verbal root *$b^he(h_2)g$- 'to divide (a portion)' is attested in Indo-Iranian (cf. LIV^2: 65; EDIV: 2; Beekes 2010: 1543), in the Greek verb φαγεῖν 'to eat, consume, swallow', and in Armenian *bekanem* 'to break' (see below). If inherited and cognate to these forms, Tocharian B *pāke*, Tocharian A *pāk* 'part, portion, share', which go back to Proto-Tocharian *$pake$ 'id.', can only come from PIE *bh_2go-.

For the reconstruction with a laryngeal *h_2 for this word, see Lubotsky (1981: 134f.; see also EDIV: 2). A reconstruction *bh_2go- would explain one Old Persian form, *bigna*, which could contain an -*i*- < *h_2. In *$baga$- < *b^heh_2go- the loss of the laryngeal would have resulted from generalisation from contexts where it would be deleted by Lubotsky's law. Nonetheless, this word *bigna* is only found in one personal name, *Bagābigna* < *$baga$- 'god' + *$\bar{a}$-bigna*- (cf. Kent 1953: 199). This word has been explained otherwise variously, notably as 'having the attacking power of God' (cf. Tavernier 2007: 14, with references to other scholars). In my view, this example, of unclear meaning, is not enough to warrant the reconstruction with a laryngeal.

If one connects, as is often done, this root *b^heg- 'to divide (a portion)' with *b^heg- 'to break' (LIV^2: 66), then the reconstruction *b^heg- without a laryngeal is warranted. In that case the Tocharian form – if inherited – could only derive from *$b^h\bar{o}go$-. This would more easily account for the Indo-Iranian forms. Only Greek φαγεῖν requires the reconstruction with *h_2 (cf. Lubotsky 1981: 134), but it is semantically distant from 'to divide (a portion)'. Even if we were to connect it to the other words, it could derive from *b^hng-, as suggested by Professor Lubotsky (p.c.). On the other hand, Armenian *bekanem* 'break', *bekor* 'fragment, piece', which fits the semantics of the Indo-Iranian much better than that of the Greek, does seem to be cognate of *b^heg- 'to divide' (cf. Olsen 1999: 524).

If we derive Proto-Tocharian *pake from Old Steppe Iranian *bāga- 'portion, share' by means of borrowing, the problems mentioned above are solved: no laryngeal *h₂ or a PIE *a is needed. Many Iranian cognates have exactly the right meaning, and the assumption of borrowing would also neatly account for the -nta plural pakenta, as it is a very common plural often used for loanwords, including Old Steppe Iranian ones (TB waipeccenta 'possessions'). Old Steppe Iranian *bāga- 'portion, share' in its turn is from the Iranian root *bag- / *baǰ- 'to bestow, divide, have a share' (EDIV: 1-2). The expected correspondences between Old Steppe Iranian and Proto-Tocharian pose no problem here, with OSIr. *ā corresponding to Proto-Tocharian *a and OSIr. *a → PT *e. As can be seen from Cheung (EDIV: 1-2) almost all Iranian languages, Ancient, Middle and Modern share very similar meanings to the Tocharian words. Suffice it to cite Bactrian βαγο 'share, part, portion, lot, fate' (Sims-Williams 2007: 200); Sogdian β'γ 'allotment, garden' (SD: 97), Christian Sogdian b'γ 'garden, vineyard' (Sims-Williams 2020: 50).

In conclusion, Proto-Tocharian *pake 'part, portion, share', yielding Tocharian B pāke and Tocharian A pāk 'id.', was borrowed from Old Steppe Iranian *bāga- 'id.'. The phonological correspondences observed are regular for Old Steppe Iranian, while attempts to derive *pake from Proto-Indo-European must cope with insuperable phonological problems.

2.2.11 TB peret, TA porat 'axe'

The Tocharian B word peret, corresponding to TA porat, both 'axe', is among the first Tocharian words correctly identified as being of Iranian origin (Lidén 1916: 16f.; see also Hansen 1940: 154f.). The vowel of the first syllable of the Tocharian A word shows p/m-umlaut of the Pre-TA *a to *o, a phenomenon which also affects inherited words (compare for example TA cmol 'birth' < Pre-TA *cmal, TA ñom 'name' < Pre-TA *ñam). Thus, the Tocharian words can be reconstructed as Proto-Tocharian *peret. The Iranian etymon to be compared is *paratu, and the rendering of Iranian *a with e in Tocharian shows that the Tocharian word was borrowed from Old Steppe Iranian. However, the Iranian data is somewhat complex, and I will discuss below how an Old Steppe Iranian form *paratu can be justified.

In Iranian, this word has a very limited distribution: it is not even found in Sogdian, which has tš /taš/, the inherited word for 'axe'. There are none-

theless three cognates of Old Steppe Iranian *paratu-, namely Old Kho-tanese *paḍa-* (acc. sg. *paḍu*) 'axe' < *paratu-, with transfer to *a*-stems and syncope (cf. Canevascini 1993: 269), Khwarezmian *pdyk* '(large) axe' < *partaka-* (Benzing 1983: 520), and Ossetic *færæt* 'axe'. Cheung (2002: 74) considers the Ossetic word etymologically unreliable since "it cannot go back to a clear Iranian proto-form". His hesitance is probably due to the fact that the Vedic cognate *paraśú-* 'hatchet, axe' does not correspond regularly so that an accurate Indo-Iranian reconstruction for this word is difficult.

Indeed, the comparison of Old Indic *paraśú-* 'hatchet, axe' yields an evident problem, in that we would expect PIr. *paraću-* (corresponding to the Old Indic form) to be reflected as Khot. †*parsa*. Bailey (1982: 55) explains this as dialectal variation, but this is impossible: Khotanese is clearly a language that reflects PIIr. *ć as *s* (or *ś*, cf. Sims-Williams 1998: 136; Peyrot 2018a: 271-272[6]), and not as *t* as would be required for the attested *paḍa-*.

In fact, the Iranian-internal facts lead to the reconstruction PIr. *paratu- and not *paraću-, which has otherwise been assumed by previous scholarship on the basis of external reconstruction (e.g. EWAia[2]: 87; Blažek & Schwarz 2017: 82). Besides the Indic and Nuristani comparanda (Ved. *paraśú-* 'hatchet, axe' but also Ashkun *pōs*, Kati *pɛċ* 'large axe'), Greek πέλεκυς 'axe, double axe, hatchet' is often adduced (e.g. DKS: 203; Beekes 2010: 1166f.; Carling 2020: 135). None of these "cognates" have an etymology in Indo-European, that is, they appear to be *u*-stem nouns but there is no root, or even a derived or underived noun, with which they can be connected.

In order to solve the problem of the discrepancy between the Iranian and Indic forms, Mayrhofer (EWAia[2]: 87) and others suggest a borrowing from Old Persian *paraθu- into the rest of Iranian. Although this could theoretically work for Ossetic, as it could indeed derive from *paraθu- as well as from *paratu- (cf. Cheung 2002: 20f.), it is an unlikely solution to this issue.

This borrowing hypothesis needs to be rejected for the following reasons: 1. this word is absent, to our knowledge, from the entire Western Iranian territory;[16] 2. even if the word *paraθu- existed in Old Persian, there existed

---

[16] Very speculatively, I would like to suggest that the Middle Persian name *prdk-y*, Parthian *prdk* (cf. Back 1978: 243), rendered in Greek as Φαρρεκ, could go back to *paratu-ka-* 'small axe' (perhaps originally a mocking name, referring to somebody with "a small axe"). However, if my etymology is correct, it should be

a phoneme θ in Khotanese and Khwarezmian. This phoneme only later yielded *h* in Khotanese, and never even completely ceased to exist in Khwarezmian. Therefore, the supposed Persic *θ is unlikely to have been rendered as *t* in Pre-Khotanese and in Khwarezmian.[17] 3. A word for 'axe' is attested throughout the Persic languages: Old Persian *tapara- 'axe', indirectly found in Elamite *dabara* (cf. Tavernier 2007: 322) and reflected in New Persian *tabar* 'axe'. It thus seems that not only assuming a borrowing from Old Persian *paraθu- is unlikely for several reasons, but even reconstructing such a preform is unwarranted.[18] Sims-Williams (2002: 239), noting the rather unexpected final *-t*, suggested that these words, or in any case, the Tocharian form TB *peret* TA *parat*, were borrowed from one undetermined Scythian language, which is of course entirely possible.

In addition, there would be no other Old Persian loanwords with the same geographical distribution, and loanwords from Old Persian were not very common in general, and less common than Middle Persian and New Persian loanwords among Iranian languages, for example.

In regard of these facts, the irregular correspondence between Old Iranian *paratu- and Indo-Aryan and Nuristani *paraću-, Greek πέλεκυς is a problem in and of itself. One could consider whether the word is a borrowing from the BMAC language, since those borrowings sometimes show irregular correspondences, although this correspondence is not listed among the irregular correspondences in Lubotsky (2001: 302f.).

The so reconstructed Proto-Iranian *paratu- 'axe' would, as far as we know, become Old Steppe Iranian *paratu, and the correspondences with Proto-Tocharian borrowing *peret(u) 'axe' are regular (on the final *-u, see below). Obviously, a theoretical Old Steppe Iranian *paratˁu-, to be ex-

---

a borrowing from a Scythian language, because of its initial *f-* (as seen in the Greek spelling), cf. Scythian Φαραδου 'PN'. Interestingly, a name with a similar meaning is also found in Sogdian: *tšˀkk* (cf. Lurje 2010: 391). For other, phonologically unconvincing suggestions concerning this name, see Back (1978: 244), with references.

[17] After *r*, Old Iranian *θ, does not yield Khot. *-h-*, but rather *-ṭh-*, even if the cluster *rθ* arose secondarily due to syncope. Examples are Khot. *paṭhu-* 'to burn' < *pari-θau̯ai̯a-*, or Khot. *baṭha-* 'cuirass' (cf. Avestan *varəθa-*).

[18] Yazghulami *parus* is completely isolated, even among Pamir languages, and thus likely to be a borrowing from an Indo-Aryan language.

pected if the Proto-Iranian form was rather *paraću-, would have yielded TB †perets, TA †porats (as in PIr. *aću̯a- 'horse' → PT *etswe 'mule' and many other examples). Thus, if the explanation of the Tocharian word as a borrowing from Old Steppe Iranian is accepted, it further confirms that the Proto-Iranian word was *paratu-.

The fact that the *-u of the Old Steppe Iranian form was not reflected is expected, since we also observe this phenomenon with tsain 'arrow' ← OSIr. *ȡ̌ainu-. Since the -wa plural (seen in tsainwa 'arrows') might reflect the original u-stem (see Peyrot 2018a: 270), it has been assumed that the form *tˢeynu and, by extension, *peretu, existed at some stage of the borrowing process (see also OSIr. *natsu- → Proto-Tocharian *nətswa- > TB mətstsa- and TA nätswā- 'to starve'). In this case, the -u was still there in Old Steppe Iranian, and Tocharian borrowed it, but it was later reduced within in Tocharian.

On the other hand, if we accept my etymology of TB mot 'alcoholic beverage' (s.v.), one could also imagine that the final *-u had disappeared after producing umlaut in the vowel (namely *madu- > *mod instead of *madu- > *modu-), already in Old Steppe Iranian. In this case, one should reconstruct *peret and *tsain for Proto-Tocharian, and analyse the Tocharian plural as a plural in -wa rather than *-u-a.

It is difficult to choose between these two options, especially since the inherited *-u from Proto-Indo-European is known to have been reduced in Tocharian. It should nevertheless be noted that the phoneme /u/ was unstable for a long time in Tocharian. In loanwords, notably, it has often been rendered as schwa: e.g., pañäkte /pə́ñəkte/ 'Buddha' < *pət 'Buddha' ← Skt. buddha + ñakte 'lord'; see also TB pat, TA pät 'stūpa' ← Skt. buddha) or rendered as /wə/ (cf. among others Fellner 2006). See further TB sakw /sə́kʷ/ 'luck' ← Skt. sukha- 'happiness', where the u of Sanskrit is rendered as a feature of the /kʷ/. For a more detailed discussion of short *u and *i in Old Steppe Iranian and Proto-Tocharian, see p. 166f.

A last point concerns the accent of the word, which is not visible in Tocharian. The syncope in pre-Khotanese (*paratu- > *parta-) and pre-Khwarezmian (*paratu- > *parta-, with syncope and secondary thematization, as in Khot.) suggests that the stress was on the first syllable, that is PIr. *páratu-. This agrees with the Greek stress in πέλεκυς, if it is indeed a cog-

nate – which is doubtful – (but disagrees with Old Indic *paraśú-*).[19] On this basis I suggest that the stress in Proto-Tocharian was also **péretu*. This also squares well with the information that Old Steppe Iranian most of the time had first-syllable stress (cf. p. 171-72).

Clearly the syncope that occurred in Khotanese and Khwarezmian had not taken place in the Old Steppe Iranian form from which Proto-Tocharian **peretᵘ* was borrowed. It is conceivable that a similar syncope occurred in Old Steppe Iranian after the borrowing into Tocharian. In this regard it may be interesting to compare Old Turkic *baltū* < *baltō* 'axe' (cf. Clauson 1972: 333; Wilkens 2021: 141), which seems to have been borrowed from an Iranian language that had undergone syncope but had not yet lost its final vowel (cf. Clauson 1972: 333, cf. also Wilkens 2021: 141). Khotanese *paḍa-*, for example, has undergone the syncope but has not preserved the word-final *-u* (cf. DKS: 203), so that it does not at first sight seem to qualify as a possible source for the Turkic word. Nonetheless the word could have been borrowed from the Khotanese accusative (acc. sg. *paḍu*), see Dragoni (2022: 174), but the correspondence *lt* : *ḍ* is a bit more complex. The Turkic cluster *lt* seems to imply either an Old Iranian dialect with **lt* < **rt* (possibly **rt* > **rd* > **lt*), or a change of **rt* to *lt* within Turkic. Naturally, both the syncope and the cluster *-lt-* seen in the Turkic word exclude borrowing from Tocharian.

A last point is that, if TB *mot* 'alcoholic beverage' (cf. section 2.3.8 of the present chapter) underwent Old Steppe Iranian *u*-umlaut (< **madu-*), **peretᵘ* did not, which corresponds to the Ossetic distribution, where we also find *u*-umlaut in *myd* < **madu-* but not in *fœrœt* < **paratu-*.

---

[19] The disagreement in stress, along with the different reconstructions depending on the family (Greek, Iranian, Indo-Aryan) lead to the conclusion that this word for 'axe' was borrowed independently in Greek, Iranian and Indo-Aryan. It is thus a true "Wanderwort" in the etymological sense of the term. It is also possible that the Greek word was borrowed from a different source, namely, from a Semitic language, cf. Watson (2013: 170), who derives it from Aramaic/Syriac *plq* m. 'axe' (*p-l-q* 'to split'). In that case, one could suggest that the Greek borrowing occurred in the second millennium BCE from a North West Semitic language, and that the *-u* of the Greek could be a trace of the ancient nominative case ending (I thank Dr. Benjamin Suchard for his help concerning this detail). Further research is needed on this subject in order to determine which hypothesis is the most plausible concerning the Greek etymology.

In conclusion, previous scholarship correctly interpreted Tocharian A *porat* and Tocharian B *peret* 'axe' as borrowings from an Iranian source. Given the regular sound correspondences between the A and B forms, it is likely that this word was borrowed in the Proto-Tocharian (or pre-PT) period from Old Steppe Iranian. Unlike what has been generally assumed in the past, the Iranian words for 'axe' go back to a common proto-form **paratu-* rather than **paraću-*, and they are not borrowings from Old Persian **paraθu-*, a form which does not need to be posited.

2.2.12 TB *perne*, TA *paräṃ* 'rank, glory'

The words Tocharian B *perne* and Tocharian A *paräṃ*, 'rank, glory' were first proposed to be loanwords from an Iranian language by Sieg, Siegling & Schulze (1931: 18) and Olaf Hansen (1940: 151-152). Both sources propose the borrowing to be from a Middle Iranian language, cf. *phārra-* in Khotanese and *prn* in Sogdian.

However, there is no reason to suggest that Sogdian *prn* /farn/ or Khotanese *phārra-* would be rendered as TB *perne*: /farn/ would yield TB †*pār(ä)ṃ*, and Khotanese *phārra-* would yield Tocharian A †*pār*, Tocharian B †*pār(r)o* (see Dragoni 2022 passim). The correspondence between the Tocharian B and Tocharian A forms is regular: the word-final vowel is lost in Tocharian A, and TB *e* corresponds to TA *a*. Thus, the Proto-Tocharian form can be reconstructed as **perne*, and it derives straightforwardly from Old Steppe Iranian **farna(h)* 'glory' by means of borrowing.

Nonetheless a problem remains with the Tocharian A form, since it has not undergone *p/m*-affection of the *a* as in *porat* 'axe', *opyac* 'memory'. There is no certainty as to why. Professor Georges-Jean Pinault's explanation is that the cluster /rn/ kept this change from happening (p.c.). I would like to propose another solution: it could be a Sogdian loanword. It is not unlikely that Sogdian *prn*, *frn* /farn/ would have been borrowed as Tocharian A *paräṃ*, but I have to admit that one could perhaps also expect †*pāräṃ*. I offer a similar explanation for TA *enāk* (see p. 29), which should have be-

come †*enak*, if directly inherited from Proto-Tocharian *eynake. For more detail, see the discussion on *puttiśparäṃ* below.[20]

Another point to note about TB *perne*, TA *paräṃ* is the meaning 'rank', which is a specific innovation that I could only find for two Iranian languages: Sogdian *farn*, for example in Buddhist Sogdian *ʾskʾprn* 'high in rank' (SD: 61) and Khotanese *phārra-* 'fortune, high position' (DKS: 261). A possible example of this usage in Khotanese can be seen in the personal name (or title?) *Hora-murta-phara* 'he who holds the position of a patron' (cf. Bailey 1982: 51), although we cannot be sure that this is Khotanese (as Dr. Federico Dragoni informed me). Bailey writes (1982: 51) that the etymon *farnah-* came to mean 'rank, position' in Buddhist Iranian languages specifically. On the meaning of Sogdian *farn* and Khotanese *phārra-* see Henning (1940: 62; cf. Asmussen 1965: 163[21]) and Bailey (1937: 914-915) respectively.[22] It is clear that the Old Steppe Iranian word borrowed by Tocharian cannot possibly have carried any Buddhist meaning, and most probably meant 'glory, dignity', the Buddhist uses and phraseology arising only later under the influence of Central Asian Buddhist culture (cf. Pinault 2002a: 246).[23]

Finally, I would like to note that Tocharian B *pernesa*, the perlative of *perne*, may have the grammaticalised meaning 'for the sake of'. This use has a remarkable parallel in Parthian *frhʾh* 'for, on account of' (DMMPP: 155), which I propose to derive from *farrah* (< **farnah-*) with an adverbial suffix -

---

[20] Since TA *enāk* is until now the only other example, and it reflects Sogdian *a* as TA *ā*, we have to remain prudent on this specific question of Sogdian loanwords in Tocharian A.

[21] On the meaning of *frn* in Sogdian, see further in detail the insightful study by Provasi (2003). In particular on the meaning 'rank, position' in Buddhist Sogdian, see Provasi (2003: 307ff.).

[22] Old Uyghur *kut* can be used in the same contexts as Tocharian B *perne*, Tocharian *paräṃ*. It originally meant 'favour of heaven' and, later 'good fortune' (Clauson 1972: 594). However, in Buddhist and Manichean contexts, *kut* could mean 'position', for example in *arhant kutı* 'position of an *arhat*' and *burhan kutı* 'position of a Buddha' (cf. Bailey 1937: 915, Wilkens 2021: 59). It is unclear to me whether this semantic shift occurred in Uyghur due to Iranian or Tocharian influence, or independently.

[23] However, Pinault attributed this word to a Middle Iranian source, in my view wrongly.

āh (on which see Henning apud Boyce 1954: 187). However, it remains to be seen if there is a connection between the Tocharian and Parthian uses. In Tocharian, for instance, there are more adverbs with -*sa*, and one could view this as a parallel innovation in both languages.

In conclusion, Tocharian B *perne* is from Old Steppe Iranian **farna(h)-* 'glory'. It acquired its meaning 'rank, position' in contact with Sogdian and Khotanese (and Old Uyghur) speakers, or, in any case, in a Buddhist context. The Tocharian A form *parāṃ* is somewhat irregular (one expects †*porāṃ*). It is possible, although historically difficult, that it was borrowed directly from Sogdian *farn*, and would thus not be a cognate of Tocharian B *perne*. This word represents a semantic isogloss with Sogdian and Khotanese, like for the word TB *śāte*, TA *śāt* 'rich' (for which see p. 74-75, p. 160-61 and p. 195). In both cases, this isogloss is not old, but due to common Central Asian Buddhist culture.

A note on TA *puttiśparāṃ* 'dignity of Buddha'

Tocharian A *puttiśparāṃ* means 'dignity of Buddha'. It corresponds to the Tocharian B phrase *pañäktäṃñe perne* (cf. Meunier 2015: 181-82). One can cite, as an illustration, the following passage:

A 18 b4-6
> *āṣānik ptāñkät käṣṣi nu äntāne mahāprabhā(se) ñoma wäl ṣeṣ waśirr oki prākräṃ pratimyo sne lotkluneyumänt puttiśparṣiṃ śäktālyi ṣñi sant(änaṃ) sāryāt*
> "Mais le vénérable, le Buddha, le maître, lorsqu'il était le roi nommé Mahāprabhāsa, avec sa détermination aussi dure que le diamant, sans retours dans sa séquence d'incarnations a semé le grain de la dignité de Buddha" (Meunier 2015: 182).

Discussing *puttiśparāṃ*, Pinault (2008: 94-95) considers it a compound of *puttiś** and *parāṃ* 'glory; rank'. Pinault views the element *puttiś**, which does not occur elsewhere, nor on its own, as derived from Sogdian *putīšt*, with "simplification prévisible du groupe consonantique final".

Another solution, based on Pinault's proposal, can be found to explain *puttiśparāṃ*: both elements, the first and the second, could have been borrowed together. Thus *puttiśparāṃ* could be viewed, not as a calque, but as a loan based on the Sogdian expression *putīšt farn* 'glory, dignity of the Bud-

dhas'. The sequence -*štf*- could have been simplified, either perhaps in spoken Sogdian or otherwise in Tocharian A, yielding TA *puttiśparäṃ*. This solution could support the hypothesis I mentioned above, that, although Tocharian B *perne* 'glory, rank' was certainly directly borrowed from Old Steppe Iranian, its Tocharian A equivalent, *paräṃ*, was perhaps borrowed from Sogdian.

A problem remains: why was *putīšt farn* borrowed with *ś* and not with *ṣ*? Since there are not many more examples of secure Sogdian loanwords in Tocharian A,[24] one can simply hypothesize that /š/ in Sogdian was phonetically palatalized due to the preceding *ī*, although this is naturally speculative.

From the TA form *puttiśparäṃ*, perhaps interpreted by speakers as *putt + -iś + paräṃ*, this construction was extended to other Buddhist concepts, such as *ārāntiśparäṃ* 'dignity of an *arhat*' corresponding to TB *ar(a)hanteññe perne* (cf. Winter 1971: 218; Wilkens 2021: 59).

In conclusion, Pinault's (2008: 95) proposal to derive *puttiś** in TA *puttiśparäṃ* 'dignité (statut) de Buddha' from Sogdian is very plausible. Moreover, I propose to derive the entire word *puttiśparäṃ* from a Sogdian collocation, directly borrowed in Tocharian A, with simplification of the sequence -*štf*- to -*śp*-, either in Sogdian or in Tocharian A.

2.2.13 TB *perpente* 'burden, load'

The Tocharian B word *perpente* 'burden, load' was almost always read as *perpette* (cf. DTB: 426). However, a reading *perpente* seems assured in most cases. I thank Athanaric Huard for having provided me with his new reading of fragment PK NS 216 where line a4 contains, among others, the words *oro[ce] perwendi* 'a great burden'. The spelling ‹nd› and the final -*i* are somewhat mysterious, and I suggest influence of the Sogdian form *prbnty* /parbandē/ (cf. Sims-Williams 2020: 135) on the spelling of the word in PK NS 216. In any case, in the cluster ‹nd› the reading with ‹n› rather than ‹t› is unmistakable, so this form decisively proves that the correct reading is *perpente*, not *perpette*. This cannot, however, be a Sogdian loanword, as one

---

[24] A likely example is Tocharian A and B *menāk* 'comparison, example' from Sogdian *mynʾk* (cf. Peyrot 2015a). If this is confirmed, it would be, to my knowledge, the only loanword into Tocharian B that needs to come from Sogdian.

would then expect †*parwandi*, †*pärwānt* or the like (pace K. T. Schmidt, apud Sims-Williams 1985: 63).

The etymology of this word is quite transparent: the source must be Old Steppe Iranian **pari-banda-* 'load, burden', itself possibly back-formed from **pari-bandaka-* 'load' → PT **perpente*. The etymology of the Iranian word is *band-* 'bind, attach' prefixed with **pari-* 'around', so 'the thing that is bound around (the ass, the horse)'. Other Iranian cognates of this word are Sogdian *prbnty* 'burden' < **pari-bandaka-* (cf. SD: 283) and Persian *parwanda* 'bundle, roll of tissues; coarse cloth where linens are put; attached document, (gathered) writing relative to a subject, dossier, file, folder', also from **pari-bandaka-*.

An oblique form *perpecce* has often been read in THT 401 b5 (e.g. in the edition, Sieg and Siegling 1953: 267). Even if it were a correct reading, this could be a writing mistake of the scribe. However, even though the ‹tt› here is indeed somewhat strange, it is clearly ‹tt› or ‹nt›, not *‹cc›, cf. Ogihara (2016: 265, 266; contra Hackstein, Habata & Bross 2021: 83).

It is noteworthy that the Old Steppe Iranian form has no **-ka-* suffix, while both Sogdian and Persian cognates show reflexes of **pari-bandaka-*. Some nouns start with the prefix **pari-* in Iranian languages (among which the well known Av. *pairi-daēza-* 'garden' < '(wall) built around', OP *paridaiza-*, which gave our *Paradise*), but they are usually deverbal. A short look at Cheung's list of words starting with **pari-* in Ossetic (2002: 184f.) will convince us of this: except for two items, they are all verbs. One can surmise that the Old Steppe Iranian word was originally **paribandaka-*, with a nominalizing **-aka-* suffix, like the preform of the Sogdian and Persian words, and the suffix is not represented in Tocharian as perhaps in the case of TB *etswe* 'mule' from Old Steppe Iranian **aťšu̯a-* 'horse' or **aťšu̯a-tara-* 'mule' and TB *kertte* 'sword' from Old Steppe Iranian **karta* or **karta-tara-* (cf. p. 185). Another solution would be that Tocharian speakers had simplified **perpenteke*, too long to their taste, to **perpente*. However, a form such as **perpenteke* would rather undergo syncope (see p. 171ff.) and become †*perpentke*, if *-ntk-* was an allowed cluster, thus invalidating this hypothesis.

To conclude, Tocharian B *perpente* 'burden, load' derives from Old Steppe Iranian **pari-banda-*, of identical meaning.

2.2.14 TB *mañiye* 'servant'

Tocharian B *mañiye* 'servant' has been recognized as an Iranian loanword since Sieg, Siegling & Schulze (1931: 11; cf. also Hansen 1940: 149). Isebaert (1980: 204) has correctly classified TB *mañiye* as an Old Iranian borrowing. It derives regularly from Old Steppe Iranian **mānı́i̯a-* '(servant) of the house' itself from Proto-Iranian **dmānii̯a-* 'of the house', see for example Old Persian *māna-* 'house' (MP, NP *mān* 'house, mansion'), Parthian *mān* 'house, mansion', Khotanese *damānā-* and Old Avestan *dəmāna-* 'house'.

TA *māññe* has been taken to be the equivalent of TB *mañiye* but rather corresponds to TB *-māññe* 'hall, pavillion', see below.

On the semantics of this word, see p. 195f. On the phonetics, see p. 171f. and 177f.

2.2.15 TB *-maññe*, TA *-māññe* '-hall, -pavillion'

Tocharian B *-maññe*, borrowed into Tocharian A as *-māññe*, is used in compounded forms such as TB *yärke-maññe* 'sanctuary, temple' < 'reverence hall', TA *talke-māññe* 'sacrificial hall' (for more detail see Pinault 2002: 320f., who established the meaning of this word). Pinault (op. cit.) suggests that this word was borrowed from Middle Iranian, which is unlikely, because of the vocalism of the final syllable: Middle Iranian **mānīg* or the like would not have yielded TB *-maññe*.

I would like to propose that this word rather derives from Old Steppe Iranian **mān-ı́i̯a-* 'of the house' (s.v. *mañiye*), cf. Old Persian *māna-* 'house'. I assume that originally the second syllable was stressed, but that it became unstressed in Tocharian because the word was found as the second member in compounds, in which regularly the last syllable of the first compound member is stressed. Because it was unstressed, the *-i-* in *-iye* was lost, i.e. **mañíye > *mañiye > *mañye*, and the cluster **-ñy-* then developed to *-ññ-*.[25] The meaning of this word was likely 'belonging to the (noble) house' > '(hall) belonging to the house'.

---

[25] Because of the assimilation in this cluster, we cannot see whether the nasal was originally palatal or not in Tocharian, but I assume that it was palatal because of *mañiye* 'servant' (see above).

In conclusion, Tocharian B -*mañ̃e* '-hall, -pavillion' derives straightforwardly from Old Steppe Iranian **māni̯a*-, with secondary loss of the stress due to its use as a second member of compounds. It was borrowed into Tocharian A as -*mā̃ñ̃e*.

2.2.16 TB *melte** 'pile', TA *malto* 'on top'

The basic meaning of Tocharian B *melte* is 'pile' and in some contexts it designates a 'pile of dung' more specifically (cf. DTB: 505). As suggested by Adams (DTB: 505), its closest cognate must be Tocharian A *malto* 'on top, in the first place'. Winter (1965: 205) has suggested, in my view convincingly, that TA *malto* is an adjective derived from an unattested *malt, in the same way as TA *parno* 'glorious' is derived from *parä̃* 'glory'.

Winter (1971: 220) separated *malto*, which he connected to Old English *molda* 'head, summit', from TB *melte*, which he connected to Avestan *kamərəδa*- 'head'. Tremblay, however, derives TA *malto* from an Old Iranian *marda- 'head' and does not cite *melte*.

A problem with Tremblay's explanation is that there is no Iranian word *marda- with the meaning 'head'. The Old Iranian word **mr̥da*- ~ *marda- rather meant 'neck' (e.g. Persian *mul* 'neck'), and **kamr̥da*- ~ *kamarda- means 'head'.[26] As I will argue in a separate article, *ka- means 'on, together with', and *ka-marda- initially meant 'on the neck', hence its meaning of 'head'.

From 'neck' to 'top', there is no easy pathway. However, the Sanskrit cognate of *marda-, namely *mūrdhán*- (borrowed in Tocharian as *murt** 'highness'), has the following meanings: 'forehead, head in general, skull, (fig.) the highest or first part of anything, top, point, summit, etc.' (M-W: 826). From a comparative perspective, we can easily start from a meaning 'top' → 'top of the body'. To the Indo-Aryans, the top of the body would be the head (as it is to us), but to the Iranians, it would be the neck, excluding the head which is not part of the body.

Viewing the head as different from the body is not uncommon at all. One has a body, but one is a head, in a sense. One can, for instance, cite Ronsard

---

[26] Instead of reading the hitherto unexplained Elamite PN *Muštimarda* as *Muštivarda as does Tavernier (2007: 486), I suggest to read it as *Mušti-marda* 'neck-fisted' (a compound meaning 'neck-sized fist, or wrist').

(1553: 39) "Voyci deux nuages tous plains / De Mores, qui n'ont point de mains ; / Ny de corps : & ont les visages / Semblables à des chats sauuages" ("Behold, two clouds full / Of Moors who have no hands / Nor do they have bodies, and their faces / Resemble wild cats"). From this poem it is clear that Ronsard did not understand the face, or the head, as being part of the body, since these Moors had no bodies, yet they had a face. It is possible that this vision of the head as different from the body is linked to the fact that the head is that from which we see, and the body is what we can see, thus excluding the head, but naturally, only a real anthropological study could solve this problem. In any case, Proto-Iranian *marda- could have designated the 'top', then the 'top of the body', that is, either the neck or the head.

In the present case, we see, once again, a sound change *rd > *ld at work in Old Steppe Iranian. This sound change is discussed in greater detail in section 2.6.2.k. The reconstruction of the Old Steppe Iranian form was thus *malda- 'top; top of the body'.

Here, we see that Old Steppe Iranian had preserved a meaning 'top' of *marda- which no other known Iranian language has kept. The Indo-European root behind Old Steppe Iranian *marda- was also inherited in Tocharian B mrāce* TA mrāc 'summit, top of the head' (cf. DTB: 514). TB murt* 'summit' (the second state of penetration in Buddhist thought) was borrowed from its Sanskrit cognate, mūrdhan- (cf. Thomas 1966: 266[15]). Another derivative of the same root was borrowed from Khotanese: TB kamartīke 'ruler', TA kākmärtik 'master' etc. (see below). Thus, strikingly, this root was both inherited from Proto-Indo-European and borrowed a number of times: from Old Steppe Iranian, Pre-Khotanese and Sanskrit.

In conclusion, Tocharian B melte* 'pile', Tocharian A malto 'on top' derives from Old Steppe Iranian *malda- 'top; top of the body' (or 'top of the body, top'), which shows a sound change *-rd- > *-ld- seen in other words, such as, for example, welke* 'stalk (?), petal (?)' < *weltke.

A note on TB kamartīke 'ruler', TA kākmart 'sovereignty'

The historical derivation of the Tocharian words B kamartīke (variant kamarttike) 'ruler', kamarttāññe 'sovereignty, rulership', Tocharian A kākmart 'sovereignty', kākmärtik 'master' is very important for both Iranian and Tocharian studies.

Both TB *kamartīke* 'ruler' and *kamarttāññe* 'sovereignty, rulership' have been known for a long time, and their meaning is not debated (see, for instance, DTB: 149). It has equally been noticed since long that these words are likely borrowings from an Iranian language.

Schwartz (1974: 411) proposed that these words and their Tocharian A equivalents are ultimately borrowings from Bactrian **kamarδə* 'head'. The Bactrian word that had to be inferred at the time has since been found in a document dating back to 710 A.D. under the form καμιρδο 'name or title of a god', see Sims-Williams (1997: 23; 2007: 220). Since Schwartz every scholar has considered the Tocharian B words discussed here to be Bactrian loanwords in Tocharian (among others, Isebaert 1980: 198; Sims-Williams 1997: 23; Pinault 2002a: 265f.; Tremblay 2005: 436; Del Tomba 2020: 130[180]; Pinault 2020: 327; Pinault 2020a: 392). I will discuss some of the points raised by various scholars, in order to further determine the etymology of these words. For this purpose, I will mainly cite Pinault 2002a, Adams (DTB) and Peyrot (2015a).

Pinault (2002a: 263) suggests that both the Tocharian A and B forms go back to an original noun **kamártV* "which would regularly develop to **kamártV* and ultimately, with loss of the final vowel, **kamart*." This would explain why the Tocharian B form is constituted of two *a*-vowels, as can be seen from *kamarttāññe* (/kamartáññe/), while the Bactrian form is καμιρδο, which seems to indicate a schwa as a second vowel, or in any case, a vowel that is not /a/.[27]

To sum up: if the word were from Bactrian, we would in principle expect †*kamärttāññe* for 'rulership' and not *kamarttāññe*. Thus, the base from which both *kamartīke* and *kamarttāññe* derive is clearly **kamarta* (cf. DTB: 149; Peyrot 2015a). Since this sequence -*art*- cannot correspond to Bactrian -ιρδ-, a derivation from Bactrian is difficult (DTB: 149; Peyrot 2015a).

Another considerable problem is that καμιρδο is a late form, while we would expect **καμιρλο* for earlier Bactrian (Old Iranian **rd* becomes -ρλ- in Bactrian, which, in late Bactrian, becomes -ρδ- "again"), cf. Sims-Williams

---

[27] In THT 128 b6, an archaic manuscript, one even finds the form *kamārtaññe*. Pinault (2002a: 263) regards the long vowel in the second syllable as a mistake. However, I do not understand how the scribe could have made the mistake of writing ‹ā› in lieu of ‹ä›.

(1997: 23) and Peyrot (2015a). Thus, in this case, neither the vocalism nor the consonantism fits with a Bactrian etymology.

These considerations are also valid for *kamartīke*, which thus phonologically is /kamartike/, with stress on the *i*, as indicated by the variants with ⟨ī⟩, and as shown by *kamarttāññe*, whose first two vowels are unaccented /a/.

The first element *kāk-* of Tocharian A *kākmart* 'sovereignty' has remained unclear for a long time. Pinault has proposed to explain it as due to contamination with the verb *kāk-* 'to call, invite' in its meaning 'to command' (first in a conference in homage to Bailey at Cambridge in 1999; then in 2002a: 265-66).[28] This seems very likely, as the first element **kā-* was not very common in borrowings, and not analyzable within Tocharian either.

Dragoni (2022: 77f.) suggested that this word was of Proto-Tumshuqese-Khotanese origin, namely, from an accusative singular **kamardu* "with early vocalization of PIr. **r̥ > *ar*." In the first two syllables, Khotanese *-a-* regularly corresponds to Tocharian *a*; the Khotanese final *-u* (of the accusative singular) corresponds regularly to Tocharian *-o*, with a Tocharian oblique *-a* (Dragoni 2022); and Khotanese *-l-* regularly derives from Proto-Iranian **-rd-*. This etymology thus fits perfectly formally, but it is also fitting semantically, since a number of Old Khotanese, Pre-Khotanese or Proto-Tumshuqese-Khotanese loanwords in Tocharian are titles or describe social positions (cf. Dragoni 2022).

The suffix *-īke*, however, cannot be Khotanese, because *-e* is not a normal Khotanese vowel ending in Tocharian. This suffix is nevertheless of Iranian origin (cf. Isebaert 1980). It is found in other loanwords from Khotanese, such as *aṣanīke* 'worthy, worthy one' ← *aṣāṃ* 'worthy' (DTB: 34). I believe that this suffix could be of Bactrian origin, compare Bactrian -ιγο. In view of the final vowel *-e*, as in *sapule* 'pot' ← Pre-Bactrian **sabōlə*, the borrowing would have to be dated in the Pre-Bactrian stage, when the suffix was **-īgə*. Perhaps the suffix was extracted from TB *spaktanīke*, TA *spaktānik* 'servant' ← Pre-Bactrian **spaxtanīgə*, compare Bactrian σπαχνιιο 'relative to service' (cf. Pinault 2020: 392).

In conclusion, Tocharian B *kamartīke*, Tocharian A *kākmart* and their cognates are not of Bactrian origin, unlike suggested in the literature, but

---

[28] Tremblay (2001: 24[37]) seems to mention Pinault's idea but does not provide a reference.

they rather derive from Proto-Tumshuqese-Khotanese, as proposed by Drag-
oni (2022: 77f.). Yet, the Tocharian B suffix -*īke* is indeed best derived from
Pre-Bactrian.

2.2.17 TB *maiyya*, arch. *meyyā*, etc. '(supernatural) power, might'

The Tocharian B word *maiyya* means 'power, might'. A Classical and Late
variant is *maiyyo*. The Archaic Tocharian B form of the word is sometimes
written as *meyyā* (cf. Peyrot 2008: 58). In fact, the only occurrences of the
diphthong ‹ey› in Archaic Tocharian B texts are found in the words *eynāke*
and *meyyā* (cf. Peyrot 2008: 58), apart of course from *e+y* found in various
paradigms where it is morphologically motivated.[29]

For the etymology of *maiyya*, Adams (DTB: 508) suggests that it was
originally an abstract meaning 'youthful strength', derived from *$*moh_1i\underset{.}{u}o$-,
which itself developed to TB *maiwe* 'young; youth'.

In my view, Adams' explanation is not convincing. Firstly, *maiyya* does
not refer to youthful strength, but to a great, powerful might, sometimes even
of supernatural nature. Secondly, a direct Indo-European transposition of this
word, as done by Adams, "[as if] PIE *$*moh_1iwyeh_a$-" does not yield the right
outcome. For instance, *$*moh_1iwyeh_a$- would not explain the final -*a* instead
of the *o*-stem to be expected from PIE *$*-eh_2$. Furthermore, it is morphologi-
cally problematic from the Indo-European point of view, because it would be
a root *$*moh_1$- with an unexplained suffix *$*-i$-, another suffix *$*-u$-, and yet
another suffix *$*-ieh_2$. Because of the synchronically exceedingly rare para-
digm of TB nom.sg. *maiyya* : obl.sg. *maiyya*, it is rather advisable to investi-
gate whether it can be a loanword.

Van Windekens (1976: 629) has proposed that this word was borrowed
from Sanskrit *māyā*-. However, this cannot be correct, as we would then
expect †*māy* (cf. DTB: 508). Furthermore, no Sanskrit loanword displays a
diphthong *ey* in Archaic Tocharian B as in the case of *maiyya* ~ arch. *meyyā*.

There is also an Avestan word *maiiā* (< *māiiā*-), which means 'craft, ca-
pacity, strength' (but also 'satisfaction, joy'). This meaning is extremely

---

[29] I was unable to find the Archaic form *meiyyo* noted by Adams between parenthe-
ses. The chronological distribution makes it clear that Archaic *meyyā* and Classi-
cal and Late *maiyya* are older than Classical and Late *maiyyo*, which is second-
ary.

close to the Tocharian meanings of 'power, might, extraordinary capacity'. One can understand the Old Persian name *humāyāsa- (Elamite ‹Ú-mi-ia-áš-šá›) as a bahuvrīhi meaning '(he who has) horses of amazing strength'. The noun *mā(i)i̯ā- is also reflected in the Persian name Humāy which comes from *hu-māi̯ā- 'having an amazing craft', cf. Remmer (2006: 139). Outside of Old Persian and Avestan, the word *māi̯ā- is also to be found in Sogdian myʾkčyk /mayākčīk/ 'happy', with regular shortening of *ā before yod (cf. Gershevitch 1961: 17).[30]

The shortening of *ā in front of yod is also found in Avestan, Sogdian and Khotanese and other Iranian languages, cf. p. 175f. It is possible to postulate it for Old Steppe Iranian as well. A reconstructed Old Steppe Iranian *mai̯(i̯)ā- would perfectly correspond to TB maiyya. To support the postulation of this sound law for Old Steppe Iranian, one needs to mention that OSIr. *ā was also shortened in front of waw (s.v. TB newiya). This would also explain the Archaic form TB meyyā, as the only other word with an Archaic TB sequence -ey- is eynāke 'common (vulgar); evil, bad' is also an Old Steppe Iranian loanword. The fact that the word has not been borrowed as **meya, and that Classical Tocharian B has -aiyy- in this word, and Archaic Tocharian B -eyy-, could suggest that the sound change *āi̯ā > *ai̯ā went through an intermediary phase *aii̯ā where the length was transferred from the vowel to the glide (perhaps in front of waw as well, with a possible intermediary form *nauu̯ii̯ā- simply noted newiya).

A parallel for the borrowing of OSIr. *ai̯ as PT *ey is eynāke. Instead of taking the latter to derive from PT *en-yənake (with a Tocharian en- prefix) as did Isebaert (1980: 115), it seems more plausible that the Old Steppe Iranian diphthong sounded to the Tocharian ear more similar to *e (= OSIr. *a) + yod than to the diphthong ay. This also seems coherent with the fact that Old Steppe Iranian *au̯ was noted in Tocharian B as e + w (s.v. newiya, TB kraupe, Archaic TB krewpe)

---

[30] Provasi (2013: 387[41]) suggests the meaning 'favorable' for Sogdian myʾkčyk. The semantics behind the Sogdian word remain to be explored: it is unclear to me how 'powerful' ('endowed with māyā-') took on the meaning 'happy'. Very speculatively, I wonder if there was a convergence with the meaning of frn, which can also, like māyā-, describe a power, a magic might (cf. Provasi 2003: 307), and which can also coincide with the notion of happiness.

Semantically, one has to assume a passage of the meaning 'craft, strength' to 'power, might', which does not seem unlikely to me. The magical meaning of *mā̆i̯ā- is not found in all the Iranian attestations (similar to Vedic māyā́- 'magic power') but the fact that in some contexts *maiyya* can have a magical meaning is interesting: it could be seen as either a preservation/generalization of the Old Iranian magical connotation of the word, or the influence of the meaning of the Sanskrit equivalent of this word.

There is yet another possibility for the source of the Tocharian B word: it could be borrowed from the BMAC language directly, especially if the Indo-Iranian words are themselves borrowed from the BMAC language, and not inherited (Lubotsky 2001: 314 writes that the word looks "fairly IE", and could be inherited, but could also be borrowed from the BMAC language). However, there is no positive indication that this could be a loanword from the BMAC directly, and there is so far no BMAC loanword in Tocharian ending in -a. This option is thus much less probable than the Old Steppe Iranian hypothesis.

In conclusion, Tocharian B *maiyya* '(supernatural) power, might' < PT *meyya* is a direct loanword from Old Steppe Iranian *mai̯(i̯)ā-*, possibly of identical meaning. The Archaic Tocharian B form *meyyā* confirms, together with Archaic TB *eynāke*, that the Old Steppe Iranian sequence *a + *i̯* was borrowed as PT *ey*.

2.2.18 TB *yentuke** 'India(n)'

The Tocharian B word *yentuke** occurs so far thrice in the corpus. Two occurrences are the following:

THT 1110 a4-5
tumeṃ karmapāyä weṣeñcatse tonak rekauna yentukäñe pele weṣ(le ||[31]
"thereupon the *karmavācaka* shall speak exactly these words in Indic" (CEToM).

---

[31] Since *reki*, *rekauna* is of alternant gender, we would perhaps expect **weṣ(lona)*, while the lacuna in the manuscript is too small for ‹lona›. The gerundive of obligation in its predicative use can be in the singular, while taking an accusative object (I thank Athanaric Huard for mentioning this point to me).

THT 1110 b1-b2

> *tumeṃ karmavācaketse nano toykk rekauna yeṃtukäñe pele (sāṅ)kāśc(ä) weṣä(le)*

> "Thereupon again the karmavācaka shall say exactly these words in Indic to (the community)" (CEToM).

As indicated on CEToM, *yentukäñe* stands for *yentukäññe, a derived adjective based on the noun *yentuke**. The next sentence is in Sanskrit, which implies doubtlessly that *yentukäñe* designates the language of the Indians. Pinault (1989: 21) writes "parallèlement à *kᵤśiñ-pele*, le syntagme de valeur adverbiale *yentukäñe-pele* signifie 'à la manière indienne', 'en indien', i.e. 'en sanskrit'."

As it is an adjective which refers to Sanskrit, it is clear that *yentuke** in this example refers to India. The following example is very fragmentary, but presents us nevertheless with a morphological question:

THT 424 b6

> *-ka yentukeṃne masa ///*

Taking *yentukeṃne* as a locative plural, and, accordingly, *yentukeṃ-* as the oblique plural, one could translate this as "he went to the Yentukes" (as a country) or "he went among the Indians" (DTB: 546). However, perhaps the word had an obl.sg. stem *-eṃ*, in which case *yentukeṃne* could also be the locative singular. If the word had an oblique singular in *-eṃ*, this would mean that the word indicated a human, and the passage would then have to be translated as "he went to the Indian" (if *yentukeṃne* is a locative singular) or as "he went among the Indians" (if *yentukeṃne* is a locative plural).

Schmidt (1985: 764-765) proposed that *yentuke** was of Old Iranian origin, deriving it from Old Iranian **hinduka-*. In my view, this is essentially correct, but his explanation of the initial *ye-* is not satisfactory, as he himself admits. The solution he suggested was that **hi-* was borrowed as **i-*, which became **yä-* at first and then **ye-* under the influence of the following **-u-*.

I believe a more satisfactory solution can now be proposed, because we have another example of an Old Steppe Iranian initial **i-* being rendered as *ye-* in Tocharian B, namely TB *yetse* (TA *yats*) s.v. In view of this second instance, it seems likely that Old Iranian **i-* became **i̯e-* or **i̯a-* (perhaps **i̯i-*) in Old Steppe Iranian. For Old Steppe Iranian, I posit **i̯anduka-* or **i̯enduka-* (perhaps **i̯induka-* ?). On the loss of *h-* of Old Steppe Iranian, see

section 2.6.2.d and on the Old Steppe Iranian development of initial *i- (and *u-), see section 2.6.2.f.

A very interesting consequence of this is that Old Steppe Iranian speakers must have known about India, and well enough to transmit their name of India to the speakers of Proto-Tocharian.

Nevertheless, it is alternatively possible that a people, an ethnic group, which lost contact with India, also lost usage of the name of the Hindu valley. The fact that Old Steppe Iranians kept it – or acquired it – does not, therefore, necessarily indicate that they used it to designate India and thus that they already had knowledge of the region of India. Indeed, for instance, the name of the Yidgha people itself derives from an older form *$i̯induka$- (on this, cf. Morgenstierne 1938: 16, 57), although they are, by all accounts, not Indian. It is thus possible that the Old Steppe Iranian name for the 'Indians' came to designate another people like the Yidghas. Because of its resemblance with the name of India, notably in Middle Iranian languages, the Tocharians could then have reinterpereted this word as meaning 'Indian'.

In conclusion, the Tocharian B word *yentuke* 'India(n)' is likely to be an Old Steppe Iranian borrowing. The consequences this may have for the history of Old Steppe Iranian and for Tocharian prehistory are still to be determined.

2.2.19 TB *yetse*, TA *yats* 'skin'

The Tocharian B word *yetse* 'skin' and its Tocharian A correspondent *yats*, reconstructible as Proto-Tocharian *$yetse$, have not yet received a satisfactory etymology.

Adams (DTB: 549) suggests that it derives from PIE *$h_1ēd$-so-, a lengthened grade derivative of the verbal root *h_1ed- 'to eat'. This is a mechanical reconstruction, which has to cope with the uncertainty whether PIE *ds would really become *ts in Proto-Tocharian. Also, the formation with a lengthened grade coupled with a -so- suffix seems designed for this etymology specifically. Adams further supposes that the reconstructed *$h_1ēd$-so- would have initially meant 'flesh' and then 'skin'. Adams does not provide any parallel for this semantic development, and I also could not find any. Most animals do not have an edible skin. On the whole, this reasoning is complicated, and requires a large number of intermediary steps.

A more straightforward etymology can be proposed. There is an Old Iranian word *iđa- 'animal skin': Avestan iza-* in izaēna- 'made of leather', Khotanese häysa- 'skin, hide', Balochi hīz 'leather churn', Yidgha īze, yiǰya 'goatskin-bag used for carrying sour milk in'. Since we see here the same correspondence between OSIr. *i- and PT *ye- as in TB yentuke* (s.v.), we can assume that this was a regular sound correspondence. This would allow us to derive PT *yetse from Old Steppe Iranian, since the correspondence of the final vowels and the Tocharian rendering ts of OSIr. *dz < PIr. *ĵ are securely attested.

Thus, Proto-Iranian *iĵa- 'skin' would have become Old Steppe Iranian *i̯eđa- or perhaps *i̯iđa- or *i̯ađa-,[32] regularly borrowed as PT *yetse, and subsequently becoming Tocharian B yetse, Tocharian A yats. This sound change was perhaps triggered by word-initial stress, but it remains unclear in the detail. As to the semantics, it seems that the Iranian meaning was rather 'animal skin, leather' and it could have perhaps shifted to the meaning 'skin' within Tocharian, or this could be a remnant of an earlier Proto-Iranian meaning for this word.

In conclusion, TB yetse, TA yats 'skin' is regularly borrowed from Old Steppe Iranian *i̯eđa- or *i̯ađa- (perhaps *i̯iđa-?), of similar meaning, perhaps 'hide, animal skin'.

2.2.20 TB retke, TA ratäk 'army'

Tocharian B retke, Tocharian A ratäk 'army' have been recognized as Iranian loanwords in Tocharian since the earliest times (Schulze 1932: 212; Hansen 1940: 155). An often cited cognate is Middle Persian radag 'row, (army) rank, line'. Although Winter (1971: 217) viewed Tocharian B retke, Tocharian A ratäk as a Bactrian loanwords, it is clear that this is impossible, as a Bactrian *ραδαγο or *ραδγο vel sim. would never yield the -e- vocalism that we see in TB retke, TA ratäk, which must rather go back to PT *retke.

Because of the vocalism, PT *retke is rather borrowed from Old Steppe Iranian, and more particularly from a form *rataka- 'line (of an army)', which underwent syncope of the middle syllable in Tocharian (see section 2.6.2.g). This adds to the Old Steppe Iranian military vocabulary found in

---

[32] For more detail, see section 2.6.2.f.

Tocharian, which appears to be quite representative of the type of contact Old Steppe Iranians and Proto-Tocharians had.

As to the semantic shift of *rataka-* 'line', one could believe that 'line (of an army)' took the meaning 'army', as is usually assumed. One can also think of a collocation such as *rataka- spadahya* 'line of army' or perhaps a compound *spada-rataka-* 'army line'. Perhaps, in the borrowing process or perhaps in Old Steppe Iranian itself, *rataka-* took the meaning of the whole collocation as a *pars pro toto*, like *babūr* 'steamboat; obs. train' in Tunisian, from *vapeur* in French *bateau à vapeur*, *train à vapeur*. Besides, one can also think of the French expression *bataille rangée*, literally 'ranked battle', designating a battle between two similarly structured armies. In this sort of context, it would not be so surprising that *rangée* 'army line' took the meaning of 'army'. The English expression *rank and file*, designating enlisted troops is a concrete example of how 'rank' (< 'line') can come to designate an army.

In conclusion, Tocharian B *retke* and Tocharian A *ratäk* derive from Old Steppe Iranian *rataka-*, which I tentatively take to have meant 'army'. The Old Steppe Iranian word may have been used synecdochally to designate the army, from a hypothetical earlier construction *rataka- spadahya* 'army line', which perhaps itself came to mean 'army' as a *pars pro toto*.

2.2.21 TB *raitwe* and TA *retwe* 'combination, mixture' and TB *eraitwe* 'in connection with'

The Tocharian words B *raitwe*, A *retwe* 'combination, mixture', have generally been analyzed as derivatives of TB *rǝytta-* 'to join, to connect, to combine' (< PT *rǝytwa-*) and TA *ritw-*, so Bailey (1937: 896[1]), Pinault (1988: 183) and Adams (DTB: 588). As Pinault (1988: 183) mentions, the root TB *ritt-*, TA *ritw-* 'to join, to combine' textually corresponds to Sanskrit *yuj-* 'to join, unite, bring together', and he translates the phrase *raitwe preke* as "moyen [et] moment (opportuns)".

The TA word *retwe* has been analyzed as analogically remade from Pre-TA *retu*, either "under the influence of Tocharian B or with the addition of the productive TA abstract suffix *-e*" (Pinault apud Weiss 2015: 191-192[21]). In the following, I will provide a new etymology of Tocharian B *raitwe*, in the framework of which it will be more economical to view TA *retwe* as borrowed from the TB form.

The meaning of the Tocharian A and B forms mentioned above is well established. Often in a medical context, they designate a combination or mixture of ingredients, for instance applied on a specific part of the body. We may also mention the Tocharian B adverb *eraitwe* + comitative 'in combination with, in connection to' as in the following examples.[33]

IOL Toch 244 b2
  tumpa eraitwe saṃtke ///
  "in combination with this, the medicine…"

PK AS 4A a3
  ṣamāññempa eraitwe cimpim palsko yātässi āstreṃ yāmtsi
  "In conjunction with monkhood may I be able to tame the spirit [and] to make it pure" (based on CEToM).[34]

PK NS 80.2 a2
  /// *(ok)ompa nai eraitwe yātalñe cwi orotsts(e)*
  "Indeed, in connection to the fruit his ability is great" (based on CEToM).

The origin of TB *rǝytt-*, TA *rätw-* 'to join, to connect, to combine' (< PT *rǝytwa-*) had long been deemed uncertain. Poucha (1955: 260) suggested it was a cognate of Latvian *riedu* (*rist*) 'to arrange' and of Greek ἀριθμός. More recently, Weiss published an article (2015), where he connects TB *rǝytt-*, TA *rätw-* to Latin *rīte* 'duly, correctly', *rītus* 'manner, way' (← *rītū* abl. 'according to the customary fashion') and Proto-Iranian *rai̯θ-* 'join, mix, combine'. According to Weiss (2015: 190) the Tocharian and Iranian roots go back to PIE *(H)rei̯th₂-* 'join, connect', with a present stem in *-u̯a-* found in Avestan next to a *-i̯a-* present.

Weiss (2015: 190) is certainly right that for semantic reasons these Latin words cannot easily be connected to PIE *h₂rei̯(H)-* 'to count' as has been proposed earlier. However, I cannot agree with some of his other suggestions. In the following I will discuss his proposal more in detail, since it is

---

[33] Since adverbs do not normally govern cases, one can suppose that *eraitwe* was in the process of becoming a postposition.

[34] I wonder if a less literal translation "with regard to monkhood" is not possible too.

difficult to accept all the points he raised concerning the etymology of these words.

As mentioned above, Pinault proposes, apud Weiss (2015: 191-92[21]), to see TA *retwe* as remade from Pre-TA **retu* under the influence of Tocharian B or with the addition of the abstract suffix *-e*. I do not know of any other example of a Pre-TA word having been remade under the influence of Tocharian B. In fact we can either have Pre-TA **retwey* (that is, **retu-* + abstract suffix **-ey*) or a loanword TB *raitwe* → TA *retwe*. Given the large number of TB loanwords in Tocharian A, and given the technical semantics of this word, it would not be surprising if it were simply borrowed from Tocharian B. This would also be more straightforward formally. That TB *-ai-* became TA *-e-* is normal and expected, as *-ai-* in Tocharian A is limited to Sanskrit loanwords (typically TA *śmoññe* 'place' ← TB *śmoññai* obl.sg.).

The existence of the root **rai̯θ-* in Iranian with the meaning 'to mix, mingle, pervade' is assured (see EDIV: s.v.). However, the present stem in *-u̯a-* found in Avestan is suspicious. Indeed, Iranian has very few present stems in *-u̯a-* (Kellens 1984: 162-63): **tar(-u̯a)-* 'to overcome', **har-* 'to guard, observe', **ǰai̯H* 'to live', all of which have **-u̯a-* already in PIE, and **rai̯θ-* 'to join, to mingle, to pervade'.[35] Synchronically, **ǰai̯H* is not a *-u̯a-*present (cf. Hoffmann & Forssman 2004: 184), which means we only have **tar(-u̯a)-* 'to overcome' and **har-* 'to guard, observe' and **rai̯θ-* as Iranian verbs taking a present suffix *-u̯a-*.[36]

Of all these examples, **rai̯θ-* stands apart, partly because it contains a diphthong, but mostly because it is the only one without a clear Indo-European origin. The Latin and Tocharian forms connected by Weiss aside, this verbal root is not even found in Indo-Aryan, and would thus appear to be of Iranian date. A root only found in Tocharian and Iranian is generally suspicious in itself, as a borrowing from Iranian into Tocharian would in many cases be the most likely explanation. The Latin forms *rīte* 'duly, correctly', *rītus* 'manner, way' seem difficult to connect to a verb that means 'to mingle, to join, to connect', but their connection is not impossible.

---

[35] The present stem of the verb **fi̯ahu* 'to hail' (only attested in Avestan) should be interpreted as a "thematic denominative formation" and not as a stem in *-u̯a-* (cf. EDIV: 86).

[36] In other words, there is no root **ǰai̯H-* extended with **-u̯a-* in Iranian, but the root needs to be reconstructed as **ǰiHu-* for Proto-Indo-Iranian.

It should be noted that there is also a verb *raēϑβa-* in Avestan, with a present in *-i̯a-* (e.g. *raēϑβaiiāt̰*), which Kellens supposes could be a compromise between *iriϑiia-* and *raēϑβa-* (Kellens 1984: 163). In any case, this shows that at some point **rai̯θu̯a-* was analyzed as a verbal root, and not as a *-u̯a-* present, as also happened for **tar(-u̯a)-* 'to overcome' (cf. Kellens 1984: 163).

Very importantly, there is a noun *raēϑβa-* 'mixture', which corresponds formally and semantically to Tocharian B *raitwe*, Tocharian A *retwe* 'mixture, combination'.[37] In my view, this match is important, as it shows that the word could have been borrowed by Tocharian, either together with the Iranian verb or not.

Since the other *u̯a*-presents in Iranian have Indic and Indo-European cognates, a number of suppositions can be made to explain why they were reanalyzed – or not – as roots enlarged with *-u̯a-*. However, for **rai̯θu̯a-*, we have no clear cognates except the Tocharian one to explain its prehistory.

There are a number of possibilities: 1) either Iranian and Tocharian are the only branches to have preserved this verb, and the great similarity of meaning in the verb and noun is due to coincidence or Iranian influence; 2) Iranian is the only branch to have had this verb originally and Tocharian has borrowed it from Iranian; 3) these two verbs are unconnected (so Cheung, EDIV: 310) and the identity of meaning of the derived noun is coincidental; 4) or Tocharian has borrowed the noun and not the verb. It is very unlikely that Tocharian had borrowed the verb and that the noun was made from the verb at a much later period and then became a perfect match with Avestan *raēϑβa-*.

Even if the Tocharian and Iranian forms were cognates, there would be no good Indo-European etymology for both verbs. For Iranian, there is not even a convincing – or known – Indo-Iranian etymology (see below).

An important point to underline is that **tar(-u̯a)-* 'to overcome' is only suffixed in *-ai̯a-* although the bare form **taru̯a-* is attested in compounds, cf. Kellens (1984: 163), while **rai̯θ-* receives two different suffixes, besides the form **rai̯θu̯a-*. This could show that the various treatments **rai̯θ-* received

---

[37] Secondary meanings such as 'moyen' discussed by Pinault (see above) may have evolved in a Buddhist context.

as a verbal stem are later than those of *tar(-u̯a)-. This conclusion could be backed by the fact that there is no Indic cognate of PIr. *rai̯θ-.

In fact, I believe that PIr. *rai̯θ- is a properly Iranian formation. I had first believed that, on the basis of the Proto-Indo-European root *lei̯H- 'to cling to, to adhere' (Ved. láyate, LIV²: 405), an abstract noun was made, probably in Proto-Iranian, but possibly in Proto-Indo-Iranian: *rai̯θu̯a- (with the *-θu̯a- abstract suffix) 'clinging, adhesion (of ingredients)' > 'a mixture, a combination'. This is the noun rai̯θu̯a- 'mixture', attested in Avestan raēϑβa- (Bartholomae, AiW 1482 'Mengung').

I then renounced this etymology since, as Professor Lubotsky pointed out to me, the existence of the Proto-Indo-Iranian root *rai̯H- is uncertain, and the Indic cognates mostly mean 'to dissolve, disappear'. Rather, he suggested that it derives from PIE *loi̯- 'to pollute' (cf. Milizia 2015), from which Proto-Tocharian *raimo (s.v. raimo*) probably derives as well, through borrowing from Old Steppe Iranian. The original meaning would thus be 'to mix' > 'to contaminate', as in Latin contaminare 'to defile, corrupt, deteriorate by mingling' ← contamen 'contact', Sanskrit lepa- 'spot, stain, impurity' ← lip- 'to smear, adhere to' etc.

From PIr. *rai̯- 'to mix' a noun *rai̯θu̯a- 'mixture' would then have been made, and the verb *rai̯θu̯a- 'to mix together, to mix' is clearly denominal. It has been thought that *riθ-i̯a- 'to die' and *rai̯θu̯a- 'to mix together, to mix' are etymologically connected; see for instance de Vaan (2003: 216) who assumes one single root riθ- 'to mingle; to die'. However, semantically there is little overlap between both meanings and the idea of their cognacy should be abandoned (as in EDIV: 309).

All this explains quite well why the root is completely absent from Indic. If my hypothesis is accepted, then Tocharian either borrowed the noun from Old Iranian, or the verb, or both. In Tocharian, it was in any case possible to derive verbs from borrowed nouns, such as TB spalk-, TA spāltkā- both 'make an effort' ← PT *speltke (s.v.) 'zeal' (cf. Peyrot 2013: 837, 837[991]) and *netsu- (from Old Steppe Iranian *natˢu-) 'dead' from which the verb *nətswa- was made (s.v. TB mətstsa-). Possibly this also happened with *tsəwa- (s.v. tsəwa). This set could have furthered the derivation of the verb ritwa- from raitwe, since it may have been unclear to the speakers whether the noun or the verb was primary.

There are also parallels for derivation of the noun from the verb: traiwo 'mixture' derived from trəywa- 'be mixed, get mixed, mingle, shake' could

have played a major role, given the semantic closeness to TB *ritt-* / TA *ritw-* 'to join, connect' and TB *raitwe* TA *retwe* 'combination, mixture'. Possibly there even existed a noun *traiwe* with a similar meaning (completely conjectural, see DTB: 341). There are other nouns derived from verbs in a similar manner, such as *traike* 'confusion, delusion' ← *trəyka-* 'go astray, be confused'. Nonetheless, there is no trace of a change in that direction for Old Steppe Iranian loanwords in Tocharian, so that it is much less likely, and much more likely for the noun to have been primary and the verb secondary.

Taking TB *raitwe* (probably borrowed in TA *retwe*) as a borrowing from Old Steppe Iranian **raiθu̯a-* is thus very tempting. From *raitwe* a verb *ritwa-* could have been made, as a back-formation, on the model of parallel nouns, like *traike* and *trəyka-*. It is also possible, but unlikely, that Tocharian borrowed both the verb (from Old Iranian **rai̯θu̯a-*) and the noun independently (also from Old Iranian **rai̯θu̯a-*).

The semantic variations of the verb TB *ritt-* / TA *ritw-* are possibly due to influence from the Indian meanings of the corresponding Sanskrit verb, an influence which seems to be pervading in the texts, as shown by Weiss (2015: 193). The Tocharian verbs also have a meaning 'fitting, suitable' (cf. Weiss 2015: 194). As Weiss (2015) reminds us, in that case they render Sanskrit *yuj-* 'to join' (cf. Sieg & Siegling 1949: 162), which also means 'to be fit or suitable'.

In conclusion, I propose to see TB *raitwe*, TA *retwe* 'combination, mixture', and the connected verbs as borrowings from Old Steppe Iranian. The Tocharian verb and noun have no good Indo-European etymology. While a convincing Indo-European etymology has not yet been found for the Iranian verbal and nominal forms, I propose to see the nominal form as an abstract derivation from a shorter Indo-European root found in Indic but lost in Iranian. According to this hypothesis, the verb is derived secondarily from the noun in Iranian. If this is correct, then there is no reason to assume that the Tocharian forms are inherited from Proto-Indo-European, and they must instead have been borrowed from Iranian.

2.2.22 TB *raimo**, TA *rem* 'dust'

A Tocharian B form *raimaine* has recently been found in THT 1622 (a group of fragments of uneven size probably belonging to the same folio) by Atha-

naric Huard.[38] It appears in THT 1622.c+b, line a7, in a description of the Tapana Hell. I reproduce the line below, with courtesy of Athanaric Huard:

THT 1622.c+b a7

> /// o pälkoṣ waraṃśne raimaine /// iñcuwasa : kärweñän tarśkaṃ ṣ/pa ///
> Huard translates it as: "… dans du sable brûlant et de la poussière … sur du fer … des pierres et des tessons …"

TB *raimaine* 'dans la poussière', if correctly identified, clearly corresponds to the Tocharian A word *rem* 'dust' (cf. Poucha 1955: 261). These words can hardly be separated from Khotanese *rrīman-* 'faeces, filth, dirt' and Middle Persian *rēm* 'dirt, filth, purulence' (CPD: 71). The etymology of these words is discussed in Milizia (2015). Milizia connects these Iranian cognates, in my view convincingly, to Ancient Greek λοιμός 'plague'. The original idea, according to Milizia (2015: 92, 100f.) is that of 'bodily rejection; dirt', initially stemming from the human body or corpse (s.v. *raitwe* for more detail on this matter). Since the Proto-Indo-European root has **l-* and not **r-* it is impossible to consider the Tocharian words as cognates by way of inheritance. The Tocharian words TB *raimo** and TA *rem* must thus be, originally, borrowed from another source.

A likely candidate is the Iranian root **rai̯-* referring to polluted substances. If we depart from the root **rai̯-*, it appears that an Iranian source is the most logical choice, since there is no known cognate of this root to be found in Indic. The proto-form would then be PT **reyma* 'dirty substance' ← Old Steppe Iranian **rai̯mā-* 'id.'. OSIr. **rai̯mā-*, in its turn, would be a feminine derivative of **rai̯ma-* masc. 'dirt' and probably served as a feminine individualizing noun (on a collective masculine base).

There remains, nonetheless, a problem concerning the inflectional type of this word. We know the oblique singular of this word, but we do not know its nominative singular yet. There are three possible inflectional classes this word could belong to. First, one could think that it belongs to the type of *aśiya*: nom.sg. -*a* and obl.sg. -*ai*, but this seems very unlikely because this class consists of female referents (Del Tomba 2020: 76). Second, it could belong to the *okso*-type, with a nom.sg. in -*o* and obl.sg. in -*ai*, and third it

---

[38] I wish to thank Athanaric Huard for sending me his edition of THT 1622, including his notes about various words, among which *raimaine*, and also for drawing my attention to this word, as he himself believed it is of Iranian origin.

could belong to the *arṣāklo*-type, also with a nom.sg. in -*o* and an obl.sg. -*ai*. In Tocharian B, words of the *wertsiya*-type have their nominative singular in -*a* and their oblique singular in -*ai*, which theoretically could correspond, but they also contain a palatal element at the end, which is lacking in *raima.

The Tocharian B word *maiyya*, which is also a borrowing from Old Steppe Iranian, has its oblique singular in -*a*, not in -*ai*. Its type is rare, so rare that *maiyya* itself was remade into *maiyyo* : *maiyya* in historical times, corresponding to a much more common paradigm. To my knowledge, the oblique singular of *newiya* 'canal' (s.v.) is not known, so that it could have been *newiyai or *newiya, but the former is also more likely. These three words, *raima, *newiya* and *maiyya* are the only relatively certain Old Steppe Iranian long *ā*-stems (*-*ā* < PIE *-*eh₂*) reflected in Tocharian B, together with *wertsiya* 'assembly', and perhaps *atiya* 'grass' as a further candidate.

Because *wertsiya* and *newiya* were recognized as Iranian loanwords, it is possible that *raima**, also ending in -*a*, was assigned to the *wertsiya*-type. However, *maiyya* contained a *yod*, and was thus more susceptible to belong to this class, also being an Iranian loanword. It is thus unlikely that a putative form *raima** was lined up with the *wertsiya*-type only because of its etymology.

A more likely scenario would be that an Archaic Tocharian B form *raima was replaced in Classical Tocharian B by *raimo**, exactly like Archaic Tocharian B *meyyā*, Classical Tocharian B *maiyya*, was later made into *maiyyo* (Peyrot 2008: 99f.). TB *maiyyo* belongs to the *kantwo*-type (nom.sg. -*o*, obl.sg. -*a*, cf. Del Tomba 2020: 108f.), which contains a few loanwords (perhaps TB *kāswo*, s.v.; TB *maiyyo*; TB *tāno* s.v.). Here one only has to suggest that *raimo** was brought into the *okso*-type (nom.sg. -*o*, obl.sg. -*ai*, cf. Del Tomba 2020: 140f.). However, only the oblique singular *raimai* is known for sure. Possibly, an uncommonly inflected word with nom.sg. -*a*, obl.sg. -*a* was made into a word with a much more common inflection (nom.sg. -*o*, obl. -*ai*).

In conclusion, the nom.sg. of Tocharian B *raimaine* 'in dust' may at first have been *raima, but may later have been changed to *raimo. It is probably a borrowing from Old Steppe Iranian *raimā- 'dust, dirt' (a feminine derivative of a collective masculine noun), cognate of Khotanese *rrīman*- 'faeces, filth, dirt', Middle Persian *rēm* 'dirt, filth, purulence'.

2.2.23 TB *weretem** 'debt; surety'

The Tocharian B word *weretem** is not directly attested, but the adjective *weretemaṣe* 'debtor through bankruptcy' (THT 1111 b1) regularly derives from it with the addition of the adjectival suffix *-äṣṣe*. Another possibility, suggested by Prof. Georges-Jean Pinault (p.c.) consists in deriving *weretemaṣe* from *werete** + *-maṣṣe* a late, colloquial suffix ("servant à dériver des adjectifs de relation, surtout avec valeur technique", cf. Pinault 2002: 328). Nonetheless, I believe the first derivation to be more meaningful, because of the ultimate etymology I propose, as I will mention below. Prof. Georges-Jean Pinault (p.c.) also suggests a connection to Tocharian B *peri* 'debt' (probably with the alternation $p \sim w$ which occurs in some examples), which is semantically very attractive. However, I am not sure how the vocalism could be explained: Tocharian tries to syncopate or dissimilate two or three *e*s in a row, so that we would rather expect †*weritem*, *vel sim.*

Adams (DTB: 663) translates this adjective as "'surety' (?) or, less probably, 'bankrupt person' (?)'". According to Schmidt (1986: 129; 2018: 104) the Chinese version of the text implies a meaning close or identical to 'debtor through bankruptcy': "*weretemaṣe* bedeutet nach seiner Entsprechung in der chinesischen Version etwa "Schuldner durch Bankrott(?)"" (see also Chung 2004: 85, 87).

Adams (DTB: 663-64) convincingly proposes to derive TB *weretem** from Old Iranian **wratam*, related to Avestan *uruuata-* 'promise, contract' and Sanskrit *vratám* 'rule; commitment to the observance of a rule' (cf. Pinault 2017: 363), which go back to Indo-Iranian **u̯rata-* neut. 'duty, promise, contract'.[39]

The meaning of **u̯rata-* is not easy to establish. Schwartz (2002: 61) translates Old Avestan *uruuātā* as 'commitment' and *uruuat-* as '(bound by)

---

[39] For the Indo-European etymology of this Indo-Iranian word, see Pinault (2017). Noting that **u̯rata-* cannot be derived from from **u̯erH-* 'to speak solemnly', since we would then expect ***urHata-*, he rather proposes to derive it from **u̯er-* 'to perceive, watch out for' > 'to ward, guard' > 'to observe, respect', and thus **u̯r-etó-* meaning 'that which is to be observed'. According to Katz (2003: 151), Proto-Indo-Iranian or Proto-Iranian **u̯ratám* was borrowed into Proto-Ugric as **wḗrta*, Khanti ("ostjakisch") *vort* 'Nachricht, Botschaft', *wort-χoy* 'bridal suitor' (someone who has promised to marry a specific girl?).

solemn declaration' (2002: 54). For Sanskrit *vratá-*, Pinault (2019: 118) has argued for a meaning 'commitment'. However, Pinault now has a different view (cf. 2016a: 60-61 and especially 2017: 358f.), and translates the Vedic and Avestan words discussed here with 'rule'.

The meaning of Tocharian B *weretem* is not completely certain (see above). However, based on the meaning 'rule, contract' deducible for PIr. *ṷrata-*, it probably meant 'rule, contract' too, and the derived adjective *weretemaṣ(ṣ)e* could then have meant 'bound by a contract or a rule; bound because of a promise', as is the case with someone in debt, a debtor.

The formal correspondences between Tocharian B *weretem**, which presupposes Proto-Tocharian **weretem*, and the posited Old Steppe Iranian form **ṷrata-* need to be commented. The vowel correspondences for the second and third syllable of the Tocharian word are regular. The final *-m* remarkably retains the final *-m* of the nominative singular of the Old Steppe Iranian neuter (nom. and acc.sg. **-am*). This fact is to be brought together in conjunction with the systematic rendering of Iranian masculine and neuter stems with Proto-Tocharian **-e*, corresponding both to Proto-Iranian **a* and **ah*-stems, since it shows that Tocharian borrowings in *-e* reflect the masculine nom.sg., **-ah* in Proto-Iranian, and the neuter *h*-stem ending, likewise reconstructed as Proto-Iranian **-ah*. It also shows that the Proto-Tocharian rendering **-a* of Iranian feminine **ā*-stems is based on the nominative too, since the **-m* of the accusative would have been preserved in Tocharian.

The first syllable *we-* of the Tocharian word remains to be explained. Perhaps Proto-Iranian **ṷr-* had undergone a change in Old Steppe Iranian. It is tempting to think of the remarkable rendering of Ir. **i-* as *ye-* in Tocharian B *yetse* 'skin' and *yentuke* 'Indian', but the conditioning here is not exactly parallel. This could explain the lack of syncope, as, for instance, the word was not /weretem/ (which would have been

In conclusion, Tocharian B *weretem** 'debt', 'surety' is a borrowing from Old Steppe Iranian **ṷrata-*. This word was neuter, and Tocharian preserves the neuter nom.acc.sg. ending *-m* (DTB: 664). The Tocharian reflex *wer-* for Ir. **ṷr-* still needs to be explained.

2.2.24 TB *waipecce* 'property, possessions'

The Tocharian B word *waipecce* means 'property, possessions'. It was mentioned as an Iranian loanword for the first time by Winter (1971: 218), after a

suggestion by Warren Cowgill. Its Iranian etymology is generally accepted, see for example Isebaert (1980: 86, 113). Isebaert, however, takes this word as deriving from the accusative "*hvāi-pačyam" (1980: 118) which is unnecessary and unwarranted as a more straightforward etymology exists, and as the accusative in *-am would likely be rendered by PT *-em, as in weretem (s.v.).

The identification of this word with Av. xʸaēpaiϑiia- 'own' is evident: compare also Old Persian uvaipašiya- 'eigen', uvāipašiya- 'own possession' (Isebaert 1980: 86), Parthian wxybyh (Morano 1982: 39).

As to the meaning, either the (Proto-)Tocharian form underwent a semantic change from 'own' > 'own thing' or it derives from a form similar to OP uvāipašiya- 'own', with an identical meaning, thus *hu̯āi̯-paϑi̯a. It appears that both proto-forms would give the same result, waipecce, in Classical Tocharian B. Old Steppe Iranian *(h)u̯ai̯-paϑi̯a- or *(h)u̯āi̯-paϑi̯a was thus borrowed in Proto-Tocharian as (*waypetʲye >) PT *waipecye 'property, possessions (collective)' > TB waipecce. On the likely loss of /h/ in Old Steppe Iranian, see section 2.6.2.d. The Tocharian A descendant of this Proto-Tocharian form should have been †wepac or †wepaci.

While we have Archaic TB meyyā and eynāke (< OSIr. *ma(i)i̯ā- and *ai̯nāka-) with Old Steppe Iranian *a+i̯ preserved as Archaic Tocharian B e + yod, we find no occurrence of †weypecce, not even in archaic manuscripts such as THT 128 a2.

A reason for this could be that the Old Steppe Iranian form was indeed *(h)u̯āi̯-paϑi̯a- with *(h)u̯āi̯- regularly rendered as Proto-Tocharian *a + yod. However, for tsain en tsaiññe there is no evidence of an archaic variant †tseyn and †tseyññe either, and the same explanation cannot be reasonably evoked for these words: it is certainly not warranted to set up *ǯāi̯nu- and *ǯāi̯n(a)i̯a- vel sim. The -ey- variants of meyyā and eynāke are rather rare and not systematically found in archaic texts, which means that it is also possible that *weypecce existed, but is not attested, in which case it might as well simply derive from *hu̯ai̯-paϑi̯a-.

2.2.25 TB śāte, TA śāt 'rich'

Tocharian B śāte and Tocharian A śāt go back to Proto-Tocharian *śate. The Proto-Tocharian word is a borrowing from Old Steppe Iranian *čyāta- 'happy'. Hansen (1940: 156) and Tremblay (2005: 448) both consider it a loan

from Sogdian, but the consonant *ś-* would then be unexpected, since we would expect Tocharian *ṣ-* for Sogdian *š-* (however, see *puttiśparäṃ*), and the final *-e* of the Tocharian B form would also be unexpected.

Rather, if taken from Old Steppe Iranian, this word has *ś-* for PIr. **čy-*, which can be explained either as a further palatalization of an Old Iranian **č'*, or as a direct borrowing from an Old Steppe Iranian **ś-* (itself from **č'*). Theoretically, Old Steppe Iranian could first have had a stage with a palatalized **ș́i̯-* (OAv. *śii̯āta-* 'happy') and then could have lost the *yod* (YAv. *śāta-* 'id.'). This means that, if Tocharian borrowed its *ś-* from Old Iranian, it would have done so at a later period than if it borrowed a palatalized **č-*, which is supposedly older in Iranian. Another possibility is that an Old Steppe Iranian **č̦i̯-*, with phonetic palatalization of the **č-*, was borrowed into Tocharian as such, and that the development to **ś-* occurred within Tocharian. This problem is discussed in more detail in section 2.6.2.b of this chapter.

A very important point to note about this word is its semantics: the meaning 'rich' is unexpected for this word, as the original meaning is 'content', and it often means 'happy' in Iranian languages; cf. Avestan *śā̦iti* 'happiness', Middle and New Persian *šād* 'happy', etc.

In my view, the semantic shift from 'happy' to 'rich' is non-trivial, and it only occurs in two other languages: Sogdian *šʾt* /šāt/ 'rich, happy' (SD: 370) and Khotanese *tsāta-* 'rich' (Bailey 1958: 148; DKS: 146). The sister language of Khotanese, Tumshuqese, interestingly also has the word *tsāta-* but with the meaning 'property' (Bailey 1958: 148). In fact, the shift apparently only occurred within Tocharian, and in a literary Buddhist context, as we can see the same shift in the Sogdian and Khotanese cognates of this word, see p. 195.

In conclusion, Proto-Tocharian **śate* 'happy', which gave Tocharian B *śāte* and Tocharian A *śāt* 'rich', was borrowed from Old Steppe Iranian. The Old Steppe Iranian form from which it was borrowed was either **śi̦āta-* or **č̦i̯āta-*, and it meant 'happy, content'. The meaning 'rich' arose in a Buddhist context.

2.2.26 TB *speltke*, TA *spaltäk* 'zeal, effort'

Tocharian B *speltke*, *spel(k)ke* (on the simplification of *-tk-* to *-kk-* and subsequently to *-k-*, see Peyrot 2020) and TA *spaltäk* mean 'zeal, effort'. From

these nouns a verb was derived: TB *spalk-* and TA *spāltkā-* 'make an effort' (cf. Peyrot 2013: 837, 837-838[992]). Because of the different stem patterns of the Tocharian A and B verbs, it seems more likely that the derivation is of post-Proto-Tocharian date, cf. Peyrot (2013: 837-838[992]), and occurred independently in the daughter languages.

Tocharian B *speltke* and Tocharian A *spaltäk* were recognized as Iranian loanwords since Winter (1971: 219), who proposed to see them as coming from the unattested Bactrian outcome of "iranisch" **sparδaka-*.

Adams (DTB: 789) is against this proposal of Winter's, and writes that "the phonological connection is not as straightforward as is usually assumed". A similar opinion is also expressed in Pinault (2020: 327). Indeed, a Middle Iranian origin of these words, as has often been assumed, is impossible because of the vowel *e* in both syllables, and because of the sequence *-lt-*, for which no Middle Iranian source seems to qualify. Bactrian, for instance, would have **-rl-* here.

Adams' alternative proposal consists in deriving the verb from PIE **spel-d-ske/o-* "as in Latin *pellō*". According to him, the verb is primary, while the nouns are deverbal. Adams' etymology is weakened by the fact that there is no other known *s*-mobile form of this root. It would also require an **o*-grade (**spol-d-sk^e/$_o$-*) which would be aberrant, since verbs in **-sk^e/$_o$-* regularly have zero-grade in the root.

Pinault (2020: 327f.) suggests a different etymology: he sees PT **speltke* as deriving from the verb PT **spaltka-* (TB *spalk-*, TA *spāltkā-*) 'make an effort'. This verb is usually considered as denominal (e.g. Peyrot 2013: 837[991]), but Pinault rather views the noun as deverbal (2020a: 327). He sees the *-tk-* sequence in this verb as deriving either from an earlier (Pre-PT) **-t-sk-* < PIE **-d^{h}h$_1$-sk̑é/ó-* (2020: 328f.). According to Pinault (op. cit.), the part **spel-* of the verb derives from Proto-Indo-European **spol-*, itself from **sep-* 'to handle (properly), hold' with an *-(e)l-* enlargement, and thus a cognate of Latin *sepeliō* 'to bury' and Vedic *saparyáti* 'to serve ritually, to serve, to honor (a god)'.

Although this is tempting semantically, as one can easily 'serve with zeal', and the notions of 'zeal' and 'service' are often interconnected, Pinault's etymology requires many steps morphologically. This is not impossible, of course, but I would like to propose a solution that, as I hope, is more straightforward.

If we assume that PIr. *rd* regularly becomes *ld* in Old Steppe Iranian, we cannot only connect *speltke* to Proto-Iranian *spr̥daka-*, but it also explains two other etyma, namely *melte* and *welke** (s.v.). Since there are several cases in which Proto-Iranian vocalic *r̥* seems to have become Old Steppe Iranian *ar* (see p. 182f.) and short *a* is rendered as Proto-Tocharian *e*, I posit an Old Steppe Iranian *spaldaka-* as the regular outcome of PIr. *spr̥daka-*. This Old Steppe Iranian word *spaldaka-* was borrowed as PT *spelteke*, regularly syncopated to *speltke* (on the syncope, see section 2.6.2.g).

We can assume that the meaning of OSIr. *spaldaka-* was 'zeal, effort' in view of, for instance, Avestan *spərəδa-* 'Eifer', coinciding exactly with the Tocharian meaning. This etymology is therefore straightforward as far as the semantics are concerned. Furthermore, there are parallels for a noun borrowed from Old Steppe Iranian from which a verb was derived in Tocharian (for example, s.v. *raitwe*). This was also the case with Khotanese borrowings, see for instance the verb TB *krasa-* TA *krāṣäyññ-* 'to annoy, vex' or 'to be annoyed' which Dragoni (2022: 117f.) analyzes, convincingly in my view, as denominal verbs derivating from TAB (or PT?) *krāso* 'torment'.

In conclusion, while Tocharian B *speltke*, Tocharian A *spaltäk* 'zeal, effort' might theoretically be inherited as suggested by Pinault (2020), it seems that an Iranian origin is also possible, and simpler. However, one has to abandon a Bactrian derivation, as per Winter (1971: 219) and view this word as an Old Steppe Iranian borrowing in Tocharian.

2.2.27 TB *tsaiññe* 'ornament'

It was Schmidt (1985: 763f.) who proposed for the first time that Tocharian B *tsaiññe* 'ornament' is borrowed from Old Iranian, i.e. Old Steppe Iranian in my terminology. It would thus be a cognate of Avestan *zaiia-* m. 'instrument; weapon', Khotanese *āysän-* 'to adorn, equip'.

In my view, it could be derived from *ǰ́ai̯-na-i̯a-* with syncope of the second syllable. It would thus be an adjective 'what belongs to equipment/instrument', possibly designating small ornaments of weapons, or, more straightforwardly 'equipment, adornment' > 'ornament'. Old Steppe Iranian *ǰ́ai̯-na-i̯a-* could perhaps indeed yield PT *tseyñye* which would, in turn, become (Classical) Tocharian B *tsaiññe*.

Another solution, suggested by Schmidt (op. cit.), consists in taking *ʒˊai̯a- as having been borrowed from Iranian as PT *tseye, to which the adjectival suffix -ññe was added in Tocharian, yielding regularly tsaiññe. This would thus originally have meant *'of an ornament, of ornaments' > 'ornament'. In both cases, the formal and semantic details remain not entirely clear, and more work is needed to explain further details of the Tocharian B word tsaiññe 'ornament'.

In conclusion, Tocharian B tsaiññe 'ornament' is probably borrowed from Old Steppe Iranian. No etymology from Proto-Indo-European has been put forward to this day, and the element tsai- has a very Old Steppe Iranian appearance. There is no immediate morphological comparandum for it in Iranian, although the root it would derive from is well attested among Iranian languages.

2.2.28 TB tsain 'arrow'

Tocharian B tsain 'arrow', plural tsainwa, was recognized as being of Iranian origin since Schmidt (1985: 763). In Tocharian texts, it corresponds to Sanskrit iṣu- 'arrow' (M-W: 168).[40]

Tocharian B tsain probably comes from Proto-Tocharian *tseyn which in its turn derives from Old Steppe Iranian *ʒˊai̯nu-, probably meaning 'arrow', cf. Avestan zaēna- m. 'weapon', YAv. zaiia- 'kind of weapon', Av. zaēnuš which means 'baldric'. The form *ʒˊai̯nu- underwent regular loss of the final -u, either in Old Steppe Iranian or, more likely, in Proto-Tocharian, as a trace of its final *-u is possibly found in the plural tsainwa and in the derived adjective tsainwaṣṣe see section 2.6.2.f (cf. also Peyrot 2018a: 270).

Tocharian B prere, Tocharian A pär and Tocharian B āstär (borrowed from Sanskrit astra-) also designate arrows. All these words represent different types of arrows: tsain and āstär corresponds to Sanskrit iṣu- (see above) and both designate an arrow. TB tsain can also have a more general meaning of 'weapon(ry)', while prere strictly means 'arrow'. It was previously believed, following Couvreur 1964 (so DTB: 64) that TB āstär designates Māra's weapon, as opposed to tsain, which would be the Buddha's. This is

---

[40] For instance, Tocharian B tsain-yamāṣeñca corresponds to Sanskrit iṣukāra 'arrow-maker' (cf. DTB: 530).

incorrect, both words simply designate arrows, whether used by an army, by the Buddha, or by Māra.[41]

To note, Buyaner (2021: 75f.) has suggested that this word derives from Early Middle Chinese *tsian^h* 箭 (Modern Mandarin *jiàn* 'arrow'), cf. Pulleyblank (1991: 148). He writes (loc. cit.) that the Iranian reconstruction ***ǯai̯n*- does not have a firm basis, but also insists on the fact that the Tocharian word strictly means 'arrow', and not 'weapon'. However, the metathesis he suggests (*ia* > *ai*) is arbitrary, and the Old Steppe Iranian etymology is therefore to be preferred. Besides, the Chinese word

In conclusion, Tocharian B *tsain* 'arrow' comes straightforwardly from Old Steppe Iranian ***ǯai̯nu*-. For a discussion of the semantics, see section 2.6.4.e.

2.3 Old Steppe Iranian borrowings: possible cases

The etymologies presented in this section do function well as such, or seem plausible, but they lack a crucial element, which makes them more hypothetical than the ones presented in the previous section. For instance, no trace of a reflex of the suffix -*iya* of *wertsiya* is found in Iranian; there is no trace in Iranian of the meaning 'or' or 'neither' for **apa* (s.v. *epe*); the meaning of TB *tsetke* is not known; the -*i*- in *ekṣinek* is unexplained (and the meaning unsure); etc. Not all the etymologies here are equally probable: for instance, while it is hardly a matter of doubt that TB *mot* 'alcoholic beverage' is of Iranian origin, and that Old Steppe Iranian is its most likely donor, it cannot be excluded that at some point an Indo-European etymology will be found for *epastye* 'skillful'.

2.3.1 TB *ākteke*, *āktike* adj. 'wonderful, astonishing; astonished', noun 'wonder'

The Tocharian B word *ākteke* and its variant *āktike* (also *āktīke*) occur quite often (28 times) in Tocharian texts and the meaning is relatively clear: as an indeclinable adjective it means 'wonderful, incredible', or, sometimes 'astonished, bewondered'. As a noun, it means 'wonder, astonishing thing,

---

[41] I thank Professor Pinault for pointing this fact out in his review of the present thesis.

surprise'. On the details of the semantics of the adjective and the noun, see Pinault (2015: 170-172). TB *ākteke/āktike* is indeclinable as a noun.[42] The form *ākteke* is older than *āktike*, as shown by Peyrot (2008: 171-172). The etymology of *ākteke* has been considered uncertain by Adams (DTB: 40).

To my knowledge, the most recent proposal put forward concerning the etymology of this word is Pinault (2015: 171-172). He writes that the initial accent (‹ā› = /á/) "proves that it is not a single word, but the univerbation of two words *ā* + *kteke*." This argument relies on the fact that Tocharian B stress is usually on the second syllable from the left, but on the last syllable of the first member of compounds. In his opinion, this *ā* is identical to the interjection TA *ā*, *hā*, TB *hā* 'O woe!'. As for "*kteke*", he suggests it is ultimately derived from a particle **kté*. He writes:

> "**kté*: < **käté* < CToch. **kwätæ*, an interrogative used as an exclamative and otherwise found in the interrogative adverb TB *mäkte* 'how' < **mæ-kwätæ* […]". (2015: 172).

Pinault further proposes that *kteke* represents a reduplicated form **kte-kte* dissimilated to *kteke*. Basically, the semantic development which he implies is **'oh! what' > *'what' > 'wonderful, astonishing' etc.

I could not find a parallel in any language for a development from '(oh) what' to 'wonder', and Pinault gives none either, although one could imagine that it could have passed through an intermediary **'surprise' (< **'a (thing about which you say) what!'). Nevertheless, some problems can be found in Pinault's etymology, notably that 'O woe' would hardly make sense in the compound that Pinault proposes.

If the first element of *ākteke* corresponds indeed to *hā* in Tocharian B, perhaps one could expect at least one form written as †*hākteke*, given the high number of attestations of this word. There is also a chronological problem: the second element must be of an early date, since **kté* is not found as an analyzable morpheme in attested Tocharian, so that the reduplication and the dissimilation proposed by Pinault must have taken place at least in Pre-

---

[42] It does not have a plural form. TB *ākteke* can naturally take secondary case endings, such as the perlative, but does not have a primary declension, and remains identical in the nominative and the oblique.

TB (if not in Proto-Tocharian or before), while the first element is admittedly "imitative of Skt. *hā*" (Pinault 2015: 171).

I propose another etymology, which accounts for the fixed initial stress, and for the structure and meaning of this word: *ākteke*, in my opinion, goes back to Old Steppe Iranian **ā-gata-ka-*, a -*ka*- derivate built on the past participle of OIr. **ā-gam-* 'to arrive', **ā-gata-*, cf. Khwarezmian *āɣadik* 'what has come' < **āgataka-*; Sogdian *ʾʾɣʾtk* /āɣatē/ 'having come; brought' < **āgataka-*. As to the stress, it should be on the first syllable, as is usually the case with most secure Old Steppe Iranian loanwords.

I propose the following semantic path to account for the Tocharian B meanings 'wonderful' and 'wonder': 'what has arrived, happened' > *'what has appeared, happened' = '(out of the ordinary) apparition, (unexpected) happening' > 'a wonder, a miracle'. The original meaning is found in words such as Khwarezmian *āɣadik* 'what has come' < **āgataka-*, see Benzing (1983: 37; cf. Khwarezmian *ʾɣd* f. '(festgesetzte) Zeit' < **āgatā*). For the meaning 'apparition', one can look at Parthian *āɣad* 'to appear'; cf. Sanskrit *āgata-* 'newcomer, guest' (M-W: 129). One can think of the French meanings of *apparition*, which can describe the coming of somebody or something, but has a supernatural, magical use, describing, for instance, the sight of a ghost, or that of a divinity or demon.

Another possible argument in favour of the present hypothesis is found in Pashto *aɣə́lay* 'handsome, pretty', cf. Morgenstierne (2003: 8) "**ā-gataka-, with obscure semantics". The semantics could be *'wonderful' (as in Tocharian) > 'handsome, pretty', with a quite usual decrease in semantic emphasis. Cheung (2005: 128) derives it rather from **ā-gu(H)-ta-ka-* 'desirable, agreeable' (Late Khotanese *hagav-* 'to long'), which is possible, of course, although this etymon does not display similar semantics in the rest of Iranian.

It seems to me that the hypothesis that TB *ākteke* 'wonder, wonderful, astonishing' derives from OIr. **ā-gata-ka-* 'wonder' < 'what has appeared' is strengthened by the meaning 'to appear' and 'what has appeared, appeared thing' taken by the words **ā-gata-* and **ā-gataka-* respectively in Parthian.

This word shows initial stress, which is remarkable for trisyllabic words in Tocharian B. As I argue in section 2.6.2.g of this chapter, this initial stress is regular for Old Steppe Iranian loanwords (with some exceptions, which I think can be explained). The syncope of **ā-gata-ka-* to PT **akteke* is also expected: in early Iranian or BMAC loanwords with three times the same

vowel, such as *a_a_a* or *e_e_e*, Proto-Tocharian reduced the syllable that was after or before the stress, cf. section 2.6.2.g.

2.3.2 TB *ekṣinek** 'dove'

Tocharian B *ekṣinek** 'dove' is a hapax, occurring in the form *ekṣinekaṃña-na* 'pertaining to a dove' (DTB: 81). It is viewed, since Schwentner (1956: 238), as an Iranian loanword in Tocharian, corresponding to Ossetic *æxsinæg*, Khotanese *aṣṣänaka-* 'dove', *āṣṣeiṇa-* 'dark blue'. However, the idea that it is a Middle Iranian loanword (DTB: 81) should be abandoned because of the vowel correspondence Iranian *a* : Tocharian *e*, which rather indicates an early borrowing.

The problem with this etymology is that the phonological shape of the Iranian etymon does not fit exactly. The proto-form is **axšaiṇa-* 'dark blue' with the **-aiṇa-* suffix of color names (on the etymology of this word, see Garnier 2019: 368). From this, **axšaiṇaka-* 'dark blue bird' was derived, like New Persian *kabōtar* 'dove' ← **kapauta-* also 'dark blue'. As we have seen elsewhere, the Old Steppe Iranian diphthong **-ai-* remains as such, including in the position in front of **-n-* (s.v. *ainake*).

Since we expect OSIr. **áxšaiṇaka-*, **-ai-* was probably unaccented in this word, while it was stressed in Old Steppe Iranian **áiṇāka-*, one could think that unaccented **-aiṇ-* had become **-īn-* in Old Steppe Iranian. However, until confirmation or disproval due to other examples is available, this proposition remains *ad hoc*.

I believe a better solution can be suggested for the phonetics of this word. Kim (2003: 51[20]) proposed the following sound law for Ossetic: Pre-Proto-Ossetic **-aiCịa-* > Proto-Ossetic *iyCa. This could perhaps also have taken place in Old Steppe Iranian, as this word could show. For a more detailed discussion, see section 2.6.2.h.

A last point is that the meaning of the hapax is not assured: its interpretation by Schwentner (loc. cit.) is based on its possible Iranian etymology. The context is *ekṣinekaṃñana misa* "meat of an *ekṣinek**". With Schwentner, we have to admit that it is difficult to see what else it could be.

In conclusion, TB *ekṣinek**, probably meaning 'dove', is likely of Iranian origin. If so, Old Steppe Iranian etymology is most likely because of the Tocharian *e* corresponding to Iranian *a*. However, it is difficult to explain the

vowel of the second syllable, which is -*i*-, against expected *-*ai*-, although possible solutions can be evoked (see also section 2.6.2.h).

2.3.3 TB *epastye* 'skillful', TA *opäśśi* 'id.'

The Tocharian B word *epastye* 'skillful, adept, capable, able' and its Tocharian A cognate *opäśśi* 'skillful' derive from Proto-Tocharian *epəstye. Various etymological attempts (listed in DTB: 94)[43] suggest a stem *-*sth₂*- 'to stand' for the second syllable and a prefix *h₁op- or *h₁opi- for the element *ep*-. It would thus be 'what stands by' or something similar.

This morphological analysis is difficult to accept if one considers the fact that no other Tocharian word contains an Indo-European prefix, except for the negative PT *en- and the locatives PT *yən- and *en-. This would be the only such example, which is uneconomical and unlikely.

Rather, I propose to see this word as a borrowing from Old Steppe Iranian. An Old Steppe Iranian adjective **abí-st(H)-iịa*- 'relative to excellency' could have easily been made on the basis of the noun **abí-st(H)a*- *'standing over, in front' > *'mastership, excellency', in a parallel way to *abi-staH-ta- 'standing (lit. stood) in front' > Middle Persian *awistād*, *ōstād*, New Persian *ustād* 'professor, master' (cf. Horn 1893: 20),[44] and thus referring originally to the position of the teacher. There was no RUKI-effect on *abi-staH- because it was extracted analogically from unprefixed forms, and thus *abi-staH- does not come *recto itinere* from Proto-Indo-Iranian.

The adjective would originally have been **abí-st(H)-iịa*-, with *abí- carrying fixed accent on the -*i*-, yielding Proto-Tocharian **epə́stiye*. This protoform **epə́stiye* became Tocharian B *epastye* (< Pre-B *epəstiye), cf. Peyrot (2021), and Tocharian A *opäśśi* (<*opästyi).

The meaning of this adjective would have been 'master-like; excellent', and once it was made into a noun, it would have meant 'excellency, mastership', which fits the Tocharian meaning quite well. The form fits perfectly as well. For these reasons, and the fact that it is more difficult to assume a single word going back to a prefixed form in Tocharian, than to view this (cul-

---

[43] To which one ought to add Pinault (2008: 451).

[44] The form **aụa-st(H)a-ta*- suggested by Horn (op. cit.) would not have yielded the variant *awistād* (*awestād* according to DMMPP: 76), of which Horn was not aware.

tural) word as a loan, I believe it is indeed another example of a borrowing from Old Steppe Iranian into Tocharian. An obvious drawback to my derivation is that the supposed formation is not so far attested anywhere in Iranian.

In conclusion, Tocharian B *epastye* and Tocharian A *opäśśi* both meaning 'skillful', may be seen as regularly deriving from an otherwise unattested Old Steppe Iranian adjective **abí-st(H)-iįa-* 'relative to mastership, to excellency', which derivation would fit both the semantics and the phonetics of the Tocharian forms.

2.3.4 TB *epe* 'or'

It is agreed upon that the Tocharian A and B word *epe* means 'or' (see for example TEB II: 85). However, the etymology of this word has remained unclear until now (see DTB: 95). The only Indo-European etymology provided for this word by Abaev (1965: 137), who connects it to Ossetic *ævi* 'or, either', and derives both from Proto-Indo-European **h₁ou̯ē*, is not possible since that preform would yield Tocharian B †*eye* (TA †*ay*), as mentioned by Adams (DTB : 95).

I believe that Tocharian B *epe* was borrowed from Old Iranian *apa 'away from', and more specifically from Old Steppe Iranian *apa which I suppose had developed a meaning 'neither, nor'. Tocharian A *epe* is a borrowing from Tocharian B. If I am correct, this etymology has important relevance to the interpretation of Iranian - Tocharian contacts. A part of my argumentation is based on my analysis on the difference of function between *epe* and *wat*, which has hitherto not been studied.

2.3.4.1. Examples of Tocharian A and B *epe*

It is important to specify the meaning of Tocharian A and B *epe* in order to establish its etymology. In order to do so, I will give below as many examples of its use and meaning in Tocharian texts as I could find. Indeed, conjunctions and disjunctive particles such as 'or' can have multiple nuances depending on their specific discursive or syntactic function. These nuances can be used in order to suggest a plausible etymology, since usually conjunctions evolve fast in use and meaning, but at various stages, their specific evolution is triggered by their previous function.

Excluded from these examples are incomplete sentences where the first or second element introduced by the disjunction is unknown, and, naturally,

very fragmentary manuscripts. In particular I will try to show that *epe* represents originally an exclusive conjunction and not an inclusive one. I will compare it to the use of *wat*, which is also generally translated by 'or', with a few selected examples.

The difference between exclusive and inclusive conjunctions is explained in more detail below (p. 80f.; cf. also Hurford 1974). An exclusive conjunction, or rather disjunction, expresses the impossibility of the coexistence of two logical elements: *X is a man **or** a woman* (one cannot be both). An inclusive conjunction, in most basic terms, says that either two things are possible, or only one. I discuss these two concepts in greater detail below.

In the examples below, provided with accepted translations when available, I will endeavour to analyze whether the meaning given by the Tocharian conjunction in bold has an exclusive or inclusive function. All these examples, when clear, are exclusive: for instance in A 6a2-3, where it is not possible that the mechanical girl is both the sister, and the wife, etc. – only one of these alternatives is possible. In THT 107, the alms can either be given to the speaker, or to someone else, but not both, etc.

Tocharian A

A 6 a2-3

*kuss aśśi sās yaṃtrācā(res mä)skatär ṣar ckācar **epe** śäṃ **epe** spaktānik* ***epe** nṣäkk oki lokit kakmus näṃ*

"Who may she be? Is she the sister, the daughter, **or** the wife, **or** the servant of the mechanic, or has she come to him as a guest, just like me?" (Peyrot 2013: 282).

"Qui est donc celle-ci ? Est-elle la sœur du mécanicien, ou bien sa fille, **ou bien** "son épouse, **ou bien** sa servante ? **Ou bien** est-elle venue (ici) en hôte exactement comme moi ?" (Pinault 2008: 258).

A 342 b2-3

mā te[45] nātäk caṃ brā(maṃ) e(pe) mā ne was entsaträ

"**ob** der Herr diesen Brahmanen **oder** uns behalten mag" (TEB II: 124).

---

[45] Cf. TEB II: 35 for the restoration as *(t)e* (CEToM has *ne*).

Tocharian B

Concerning the next example (and the previous one as well), one could think that the conjunction could also be inclusive, in that the lord could possibly keep either the Brahmin or the speakers. But in the context, it is unlikely, and the lord rather has a choice to make concerning who he should keep.

THT 79 a5

(e)pe saswe wess eṅträ epe brāhmaṇeṃ mā ra tsak wes co(mp) ///
"**Whether** the lord keeps us **or** the Brahmin, we are, at any rate, not able, either, to (endure *vel sim*.) that one ..." (Melanie Malzahn in CEToM)

"**ob** der Herr diesen Brahmanen **oder** uns behalten mag" (TEB II: 124).

The next example is unfortunately too obscure in order to make a secure judgment, but one could imagine that the *atka*-capacity and the *ṛddhi*-power are two incompatible powers that cannot be used at the same time by the same person. In any case, the *ṛddhi*-power is a strong supernatural power, whereas *atka*, whatever it is, designates here a capacity, rather than a strength.

IOL Toch 178 b7

/// (po) yente kärkāte (c)w(i) no tsakṣträ ñke asāṃ • atkane tu cämpamñe epe maiyya räddhiṣṣa […][46]
"… he stole the wind and his throne burns; in concentration (?) [is] that power of magical strength" (Adams DTB: 10)

but I think the following translation is better:

"the wind carried all of this (?) away, but now his throne burns. In this *atka*-capacity **or** (with) *ṛddhi*-power…"

As Dr. Athanaric Huard (p.c.) informs me, this text has a parallel in Buddhic literature, namely, Śrāvastī's miracle, where the Buddha defeats heretics through magical powers which they cannot outmatch.

---

[46] For the restoration of the beginning of the line, I follow Schmidt (1974: 400).

IOL Toch 305 b4-5

kwri no mā yāmtär vaisarp • warkatsa • praśśāṅkaṃ erṣäṃ • epe bha-gaṇḍālänta mäskenträ tärrek

"But if one does not do [it], [this] causes erysipelas, hydropsy and *praśśāṅkaṃ*, or [he has] haemorrhoids, is blind […]" (Peyrot 2013: 676).

THT 107 a8-9

*ñi aiścer ce pinwāt **epe** se ññissa śpālmeṃ tākaṃ cwi aiścer*

"[…] do you give those alms to me, **or** do you give it to somebody who is better than me?" (Peyrot 2013: 702) "[…] est-ce à moi que vous donnez cette aumône-ci ? **Ou bien**, celui qui d'aventure sera supérieur à moi, est-ce à lui que vous donnerez ?" (Pinault 2008: 157)

THT 107 b1

*bram-ñikte weñā-meś ṣerśkana : se ñisa śpālmeṃ rṣāke tākaṃ cwi aiścer **epe** tuwak ñi aiścer*

"Le dieu Brahman leur dit: « sœurettes, celui qui d'aventure sera un sage supérieur à moi, est-ce à lui que vous donnerez ? **Ou bien**, est-ce à moi que vous donnerez justement ceci ?" (Pinault 2008: 157).

THT 107 b8

*ṣerśkana se nomiyeṣṣe bhājaṃ rerinu star-me **epe mā***

"Sisters, is this jewel bowl left by you [to me] **or not**?" (Peyrot 2008: 123)

"Sœurettes, ce bol de joyau, (à moi) est-il laissé par vous **ou non** ?" (Pinault 2008: 158).

W 15 a5

*satkenta **epe** puṣne **epe** rohinikene satkenta waltsa(na)*

"… remèdes **aussi** en cas de *puṣpa*, **aussi** en cas de tuméfaction à la base de la langue; [sont] les remèdes à combiner …" (Filliozat 1948: 83).

In clearer language, it could be translated as

"… remèdes en cas de *puṣpa* **ou** en cas de tuméfaction à la base de la langue". Here Filliozat uses *aussi* in a quite obsolete meaning (similar to 'or'), because it can be repeated twice in a more beautiful manner than *ou*.

PK AS 3B b3

ṣpel gairikäṣṣe ṣpel te śār kätnālle **epe** *no kroścana tonak śār lupṣallona*
• toṃ prayokänta y(amaṣṣällona) ///

"... pellet, saffron pellet: this is to be strewn over [it], **otherwise** these
very same [ingredients] are to be smeared cold over [it]. These treatments
have (to be made as)" (CEToM).[47]

2.3.4.2. Discussion on the origin of 'or' in various languages

In many languages an exclusive conjunction 'either, or' is differentiated
from an inclusive conjunction 'or': for instance, in Latin exclusive *aut* is
opposed to inclusive *vel*. Among other Indo-European languages, there is for
instance Latvian exclusive *vai* and inclusive *jeb*, next to exclusive *vai … jeb*.

The difference between an inclusive conjunction and an exclusive one
can be seen from a set of examples given in Hurford (1974). "Ivan is an
American or a Russian"; "That painting is of a man or a woman" are exclu-
sive, because they naturally exclude each other, it is either an American or a
Russian, either a man or a woman. On the other hand, "Inmates may smoke
or drink" is inclusive, because inmates may do both.

In the cases of the examples given above, we can say that A 6 a2-3, and
the following examples are exclusive, whereas PK AS 3B b3 is inclusive. In
A 6 a2-3 the mechanical girl can necessarily be either a sister, a daughter, a
wife or a servant. Her being one of those things excludes her being any of
the others. In A 342 b2-3 and THT 79 a5 the lord can either keep the people
who speak, or the brahmins, but not both. In THT 107a 8-9 the alms can
only be given to one person: either the speaker, or someone who would be
superior to him; likewise the example THT 107 b1. In THT 107 b8 the jewel
bowl can be left either to the speaker or to someone else, and *epe mā* is
clearly exclusive. In W 15 a5, the remedies can work for either disease, but
epe retains an exclusive function, in the sense that it is only expected for a
diseased person to have either *puṣpa* or tongue tumefaction, in any case even
if both can be had at the same time. In PK AS 3B b3 the ingredients can

---

[47] Compare Filliozat 1948: 52 "boulette…, à… dans une boulette d'orpiment aussi,
donc, ces choses sont bien à appliquer froides, ces moyens..."

either be smeared cold or perhaps smeared after having been heated? The context seems to be exclusive, but it is unclear.

As mentioned above, many languages differentiate between inclusive "or" and exclusive "or". In many other languages, however, the inclusive "or" should rather be referred to as a neutral "or", as it can take an exclusive function in some cases.

Such is the case of English *or* (neutral, inclusive), opposed to the strictly exclusive *either or ~ either* (exclusive). If I say "here, you can drink either coffee or tea" it clearly indicates that only one of the two options is possible. If I say "here you can drink coffee or tea" it sounds more as if, in this place, both options are available. In English, there are also uses of "or" which can be exclusive (cf. Hurford 1974), such as in the examples cited above (e.g. "Ivan is an American or a Russian"; "That painting is of a man or a woman").

One can cite similar examples in other languages: French neutral, inclusive *ou* vs. exclusive *ou bien* (see Meillet 1921: 164), where *ou* can also be exclusive, and should thus be primarily defined as neutral. One can also think of Contemporary Persian neutral, inclusive *yå* vs. *o yå* lit. exclusive 'and or'. In Tunisian Arabic, the neutral conjunction *wala* 'or' is found alongside exclusive *ou bien*, borrowed from French.

I will describe a functional path taken by a number of disjunctive conjunctions in world languages. I do not pretend that this is the only possible path, or that it is universal, but it is observed in a number of languages, and I propose to reckon Tocharian A and B among them.

(0) A neutral disjunctive morpheme exists. (1) An exclusive conjunction is created, because emphasis on exclusion is felt as pragmatically necessary. It is generally created by (a) the addition of morphemes, (b) the repetition of the neutral morpheme in two places in the sentence, or by (c) grammaticalization processes, and (d) borrowing can occur in cases of strong cultural contact. As Meillet (1921: 169) writes "[i]l n'y a pas d'espèce de mot qui ne puisse livrer des conjonctions." Afterwards, (2) the exclusive conjunction is weakened due to repetition (cf. Meillet 1921: 164ff.; 169f.), becomes neutral, and a new exclusive conjunction is created, using one of the processes described above. It becomes weakened in its turn (3) etc.

(a) English *either* is for instance derived from Old English *ægðer*, itself from *a* 'always', *ge-*, a collective prefix, *hwæðer* 'whether'. It originally means 'one of both'. French *ou bien* is constituted of *ou* 'or' + *bien* 'well', as

is Dutch *ofwel*.[48] Contemporary Persian *yå* derives from Cl. Persian *yā* itself from Middle Persian *ayāb*. The etymology of the latter is disputed, but it is generally admitted that it derives from the addition of three morphemes, **aδa-* 'then' **-u̯ā* 'or' and a third one, which is sometimes derived from **pi-*, an emphatic particle (e.g. Bartholomae 1916: 38[9]; Back 1978: 199).[49] It is probable that this conjunction was formed to be exclusive, opposed to **-u̯ā*, which was neutral. Now that it is neutral, the exclusive conjunction *oyå* has been formed in Contemporary Persian from *-o* 'and' and *yå* 'or'.

(b) In most languages, it is always possible to express exclusivity by repeating the neutral or inclusive 'or' before each proposition. For instance in French "tu prendras ou du thé, ou du whiskey"; "ce livre est ou à toi, ou à moi" : there is no doubt here that one has to consume either tea or whiskey, and that this book is either yours or mine. In these examples *ou … ou* corresponds and could be replaced by a single use of *ou bien*. Some languages combine their exclusive conjunction with the neutral one in this type of construction, such as English *either … or*.

(c) Grammaticalization processes in order to create a new exclusive conjunction vary across languages. One of them consists in deriving an exclusive conjunction from a word meaning 'one of both' (as in English *either*, see above). Another grammaticalization process leading to the creation of an exclusive conjunctive particle consists in the use of a particle meaning 'away from' to express disjunction. This is the case of Latin *aut* 'or (exclusive)' which ultimately goes back to Proto-Indo-European **h₂eu̯-* 'away' (cf. de Vaan 2008: 64). This would also work if my etymology of Cl. Persian *yā* and Middle Persian *ayāb* as containing an element **apa* 'away from' were accepted.

(d) The borrowing of an exclusive conjunction 'or' is not very common, but has been noted to occur in cases of strong cultural contact. One can cite Latvian exclusive *vai*, borrowed from Livonian *või* or dialectal Estonian *vai* opposed to neutral *jeb*, which is inherited (cf. Karulis 1992: s.v.). One can

---

[48] I thank Dr. Benjamin Suchard for reminding me of the Dutch form. He also suggested it is a calque from French, which is very likely according to Dr. Peter-Alexander Kerkhof (p.c.).

[49] In my view, it is equally possible to suppose **aδa-* 'then' + **au̯a-* 'off, downwards' (or, indeed **-u̯ā*) + **apa* 'away'.

also cite Tunisian Arabic exclusive *ou bien* from French exclusive *ou bien* opposed to inherited and neutral *wala*.

(2) In the second step of this process, the exclusive and inclusive conjunction coexist. Generally the inclusive or neutral conjunction is felt as being "too weak" and loses little by little its importance in the language, before being replaced by the exclusive one, which becomes more neutral. (3) Another exclusive conjunction is created, it coexists with the previous exclusive one become neutral, and ultimately replaces it, and so on and so on.

This scenario is entirely schematic, and of course should be nuanced. It is possible for a language to have multiple conjunctions at various levels of semantic development. Besides internal hesitations within the language, dialectal data can add to the variety in the use of conjunctions. I will try to examine the Tocharian data in order to reach a conclusion as to the state of disjunctive particles in Tocharian A and B.

2.3.4.3. The case of Tocharian

In Tocharian A and B there are three ways to express exclusive or inclusive disjunction: 1. *epe*; 2. TB *wat* TA *pat* and 3. TB *epe wat* TA *pat nu* and possibly TA *epe pat*.[50] Because the three are part of the same system, and it is impossible to analyze a conjunction on its own, especially if one has etymological aims, I will also shortly discuss the function of TB *wat* TA *pat*. Afterwards I will discuss the function of TB *epe wat* and TB *wat no* in a few examples. This will lead to a more elaborate discussion on the disjunctive conjunctions of Tocharian A and B, their prehistory, and naturally the origin of the conjunction *epe* in Tocharian A and B.

TA *pat* and TB *wat*

Tocharian A

A 5 a4-5
täpreṃ ats pälkäṣ mäṃ(tne) tsekeṣi pekeṣi **pat** *arämpāt*[51]

---

[50] TA *epe pat* is only found in an extremely fragmentary context (in A 428) so that its meaning cannot be ascertained.
[51] I follow the restoration as presented in Peyrot (2013: 263).

"then it looks exactly as if it were a fashioned **or** painted figure." (Peyrot 2013: 263).

A 10 a1

*mā nu caṃ täm knā(nmu)neyäṣ wärcetswātsuneyā ṣñi **pat** ālu **pat** täm ṣurmaṣ pruccamñe ya(ts)i*

"mais, du fait de la déficience de sagesse, pour cette raison, elle ne peut procurer absolument aucun avantage, **ni** pour soi, **ni** pour autrui." (Pinault 2008: 262).

A 59 b3-4

*(tā)pärk ṣñi wärpāluneyaṃ **pat** āssi ālu eluneyaṃ **pat** nā(täk yärm)*

"À présent le seigneur [est] l'autorité, **soit** [qu'il opte] pour sa jouissance, **soit** dans le don [de la jeune femme] aux autres" (Meunier 2015: 56).

A 275 a5

*śāt ekro **pat** wrasom tāṣ*

"if this was a rich **or** a poor being" (CEToM).

Tocharian B

IOL Toch 306 b3-4

mā walkeññe päst yā(mormeṃ) […] *yāmormeṃ istak **wat** päst yāmormeṃ • istak **wat** prekesa yāmorme(ṃ)*

"Having not moved for a long time … (the tumour (?)/imagined foetus (?) suddenly moves with pain (?))[52] … having (moved) then suddenly having moved again, suddenly again for a time having moved, (through the accumulation of) *vā(ta)*)" (Carling 2000: 91).[53]

PK AS 3A a1-2

*läksañana misa lykaśke kekarśwa tsatsāpauwa ampoññaṃtse sāṃtke kaṃnte kältsau ṣalype **wat** malkwersa **wat** pissauṣṣe war **wat** śār kuṣalle*

---

[52] This part corresponds to a restoration done by Carling (2000) on the basis of textual parallels.

[53] Cf. Carling (2000: 86) "[…] [it] refers to the movements of the tumour: It does not move for a long time, and thereupon it pulsates like a limbless foetus."

"Fish meat finely chopped [and] crushed [is] a remedy against abscess. Oil pressed a hundred [times] **or** [oil] with milk **or** aneth water has to be poured over [it], in case of a gall [boil]" (CEToM).

PK DA M 507.37 and .36a.83-85

*saswa ce we[s](s)i (– Pu)ttewante snai paille yāmu ste ce **wat** wäntare kakāmau ste sū – (–) nau[ṣ]ameṃ papaikau ste*[54]

"Oh lord! The (things that) Puttewante has done lawlessly [to us], **or** the things that have been taken (by him), they have been written above" (adapted from Ching 2010: 215).

IOL Toch 307 a6-b1

*paiykalñesa vācavārg waiyke reki mantanta läññi-ñ (k)oynm(e)ṃ (:) kₐce-sa ṣañ-añmä karstoymar alyekepi kaloym **wat** mīyäṣṣälyñe*

"May through the writing of the Vācavarga never go out from my mouth untrue speech by which I might destroy myself **or** bring harm upon someone else." (CEToM).

PK AS 2B b5-6

*kewiye miśosa malkwersa **wat** nastukārm niryuhaṃ yamaṣle*

"cow urine **or** milk, a nasal injection [or] an enema has to be made." (CEToM).

PK AS 5A a5-6

*(wewe)ñor ārtar **wat** no : ṣe keklyauṣor eñ(c)imar ārth vyajantsa ṣañ yk(nesa)*

"May I seize … what (has been) said or what has been praised at the first lecture [lit. hearing] with meaning and literary form (according) to the proper manner. ..." (CEToM).

PK AS 6B b1-2

*nandeṃ ālts(i)ś pūdñäkte ṣ(amā)neṃ mäntrākk= ālyeṅkä(ṃ) : wert-siyaine orotsai **wat** w(e)ña ṣkas toṃ ślokanma*

"In order to restrain Nanda, the Buddha lord told these six stanzas in the same manner to other monks **or** in a large assembly" (CEToM).

---

[54] I follow the transcription given by Ching (2010: 215), from which I removed the punctuation which she added against the manuscript.

Reviewing the examples above, one can see that some are inclusive, while some are exclusive. For instance, the example found in IOL Toch 307 a6-b1 is inclusive: one might both destroy oneself and bring harm upon someone else by untrue speech. PK AS 5A a5-6 is also inclusive, since what has been said is not by nature excluded from what has been praised. A 5 a4-5; A59b3; A 275 a5; PK AS2B b5-6; PK AS 3A a1-2 are exclusive. The last example, PK AS 6B b1-2 is neither exclusive nor inclusive, the use of *wat* in it is conjunctive, that is, it is equivalent to an "and".

Indeed, the Buddha was repeating the same words to monks and to large assemblies. This conjunctive usage of *wat* in this last example (implying a relative neutrality of the particle) is reminiscent of the one in the example in IOL Toch 306b3-4, where it was even left out of Carling's translation. It is interesting to note that we do not have any examples of a conjunctive use of *epe*, which apparently had a much "stronger" disjunctive function than *wat*.

2.3.4.4. TA *pat nu* and TB *epe wat* and TB *wat no*

To complete the global analysis of disjunctive particles in Tocharian, I believe it is useful to consider the function of the following conjunctions briefly as well: TA *pat nu* and TB *epe wat* and TB *wat no*.

A 69 b5

> *näs wrasaśśi klopaṃ (p)āṣlune yatsi* **kupre pat nu** *mā yātal*
> "**Oder wenn** (ich) nicht imstande (bin), den Menschen im Leid Schutz zu gewähren, ..." (Thomas 1954: 741).

A 226 b6-7

> *ke* **pat nu** *krī tāṣ ñareyäntwaṣ ke* **pat nu** *saṃ krī ñäkcī suk näṣ kälpīmār : ke pat nu ākāl ñäkci napeṃṣi ā*
> "Who wants to (be freed?) from the hells, **or** who has this wish, "may I obtain divine happiness!", **or** who has the wish, "… human and divine…"" (Peyrot 2013: 268).

Here Peyrot does not render the first *pat nu* because it is not fitting in English. If one were to translate it completely literally it would be: "**or but** who has the wish from the hells, **or but** who has this wish "may I obtain divine happiness, **or but** who has the wish "… human and divine…"

It is not necessary to examine all examples: while every single use of Tocharian A and B *epe* is exclusive, the other conjunctions studied here can be either exclusive or inclusive. They are thus "neutral". According to the theoretical discussion above, it appears that the meaning 'or' in Tocharian B *wat* and Tocharian A *pat* should be older, since it is weakened, compared to TAB *epe* which has a stronger exclusive disjunctive use as a particle.[55] Tocharian A and B *epe* is thus "newer" in the system, which implies it was either borrowed later, or that it acquired its exclusive function later.

2.3.4.5. The etymology of TAB *epe* 'or'

According to ongoing research by Jens Wilkens, Hans Nugteren & Michaël Peyrot, the Old Uyghur word *ap* 'nor' (for which see Clauson 1972: 3) was borrowed from Pre-Tocharian B **epe*. I imagine the meaning of this Pre-Tocharian B **epe* could have been 'nor', as in Old Uyghur.

Indeed, the meaning 'or' can derive straightforwardly from the meaning 'nor', identically to Tunisian Arabic *wala* 'or', which derives from *wa-lā* 'and not' = 'nor'. However, Wilkens (2021: 55) translates it as 'oder … auch, entweder… auch'. In this case, we can assume a direct borrowing from a possible meaning 'or' of Tocharian A **ap* or Pre-TB **epe*. Nevertheless, in case Clauson's translation is to be preferred, we can use the Uyghur word as a demonstration of the functional trajectory of Proto-Tocharian **epe*. Since TAB *epe* is solely found with an exclusive function in all Tocharian A and B examples reviewed above, it is more likely that this exclusive disjunctive function is rather late, for instance of Pre-Tocharian B (and A) origin, so that the particle did not have the time to be weakened in function. Thus, it would be more fitting to see Old Uyghur *ap* as having at least one meaning 'neither… nor' taken from Tocharian.

The etymology of the Tocharian word I consider straightforward: it was borrowed from the Old Steppe Iranian particle **apa*, which originally meant 'away from', as in Avestan and Old Persian, but took on the meaning 'nor' in a conjunctive use.[56]

---

[55] Beekes (2011: 249) explains *wat* as deriving from PIE **-u̯e* 'or' + *-t* (what this *-t* is remains unclear to me). The etymology of *pat* remains obscure.
[56] For the possible presence of **apa* in the New Persian word for 'or', see fn. 49.

A particle meaning 'nor' is always, by definition, exclusive, and when it becomes positive in meaning ('or'), it tends to remain exclusive. The change from the negative to the positive meaning might be due because the negative part of 'nor' is also assumed by the verb, or by another particle (such as PT *ma).[57] Once this exclusive particle exists, there is also a strong tendency for it to become weakened, and thus become neutral. At that point, usually, a new exclusive disjunction should appear, as per the cycle described above.[58] Here, we are at a stage when *epe* is still strictly exclusive, whereas TB *wat*, TA *pat* have become weakened and neutral. This suggests that the passage from 'nor' to 'or' occurred shortly before the first attested texts, which could fit very well with the Old Uyghur data (if Clauson's translation is to be preferred).

A theory that could explain the absence of a Pre-TA word *ap is that in Tocharian A it never became a disjunction 'or', but still meant 'nor': TA *pat* having become neutral, the speakers of A borrowed in their turn the exclusive disjunctive particle *epe* from Tocharian B.

In conclusion, Tocharian A and B *epe* 'or' derives from Old Steppe Iranian *apa 'away from' and more particularly, I argue, from its posited meaning 'nor'. It first took on the meaning 'nor', when the word was borrowed from Tocharian by Old Uyghur speakers. Later, TAB *epe* became an exclusive particle 'or', which is already the case in all our attestations.

2.3.5 TB *ñyātse* ~ *ñātse*, TA *ñātse* 'distress, calamity; danger'[59]

There is a set of obviously related words: TB *ñyātse* ~ *ñātse* and TA *ñātse* 'distress, calamity; danger'. Of the two Tocharian B forms *ñātse* and *ñyātse*,

---

[57] One can, for instance, suggest that, for Tunisian, sentences like *mē temšīš ġadi wala hnē* "do not go there nor here" were reinterpreted as "do not go there or here", thus yielding the positive equivalent: *emšī ġadi wala hnē* "go there and here (if you want)", although etymologically this meant *"go there and not here".

[58] On the weakening of conjunctions, see Meillet (1921: 164ff).

[59] This etymology I have discovered on my own, and I have worked independently on it (as well as on *entse*), and presented it at the Tocharian International Conference in Vienna (October 25th-26th 2019). Afterwards, I have discovered that Isebaert has published the same etymology (Isebaert 2019). My etymological discussion differs a little from his in the morphological and semantic detail.

ñyātse (pl. *ñyatsenta*) is more archaic, while *ñātse* (pl. *ñatsenta*) is later (Peyrot 2008: 63-64). The Tocharian A form *ñātse* was borrowed from the later Tocharian B variant *ñātse*.[60] In one bilingual Tocharian B-Sanskrit text (THT 543 a5) it is translated as *īti* (spelled ‹iti›) ‘distress, calamity, plague’.[61]

The etymology of this word is generally considered uncertain (DTB: 291). Hilmarsson (1991: 137-139) proposes to connect it to Proto-Germanic *nīþa-, “a thematic neuter with the approximate sense ‘ill-will, hostility, hate’”. He cites cognates such as Goth. *neiþ* ‘ill-will, envy’, OIcel. *níð* n. ‘libel, insult’, OE. *níþ* ‘enmity, hate, combat’, etc., and projects those words back to Proto-Indo-European *níH-to- ~ *néiH-to- (Hilmarsson 1991: 138). Based on an observation that “[i]n Tocharian the suffix *-to- was generally supplanted by *-ti̯o- […]”, he also proposes that Proto-Tocharian *ñyātse (Common Tocharian *ñyātsæ in his notation) goes back to PIE *niH-ti̯o-.

Although Hilmarsson’s etymology could potentially work phonologically, apart from the fact that I expect PIE *niH- to yield PT *ña-, not **ñya-, I do not find it entirely satisfying semantically. The meanings of ‘hatred, enmity’ and ‘danger, distress’, although belonging to a similar semantic field, are not evidently connected. Furthermore, *ñyātse ~ ñātse* would be totally isolated in Tocharian A and B, having no verbal root or other related form from which they could derive.

After Isebaert (2019), I rather propose to connect Tocharian B *ñyātse* to Parthian *niyāz* ‘need, distress’ and Middle and New Persian *niyāz* ‘need’.[62] However, these forms cannot be the direct source: with its final -e, *ñyātse* should rather have been borrowed from the Old Iranian stage, and it presupposes an Old Steppe Iranian *ni-āđah- or *ni-āđa-, depending on the exact etymology: if it goes back to a form closer to Proto-Iranian *anđah, then it should ultimately go back to a proto-form with final *-ah, whereas if its proto-form is closer to Avestan *āzi, a reconstruction with *-ah is not expected.

The problem is that no such formation is attested in Avestan, nor can it in any trivial way be posited for Proto-Iranian as the preform of the Middle

---

[60] The expected Tocharian A cognate would be †*ñāts*.

[61] In the same line *snai ñātse* translates Sanskrit *anīti* ‘absence of calamity’.

[62] The Persian form was ultimately borrowed from another Iranian language, as the Persian reflex of PIr. *ni-ājah* would be †*niyād*.

Iranian forms just cited. Rather, Avestan has *āzi* 'greed, Greed-Demon' (AiW: 343) attested, a deverbal noun derived from the verb *āz-* 'to long for' (< Proto-Iranian **ā-Haǰ-*). Thus, we would need to assume a thematic derivative **āza-* or **ni-āza-* for Old Steppe Iranian, either from the root **āz-* directly, or on the basis of the **āzi-* attested by Avestan.

The word *āzi* and its cognates are generally believed to be cognates of Skt. *āǰí-* 'race, competition, fight' and to be derived from **h₂eǵ-* 'to drive, provoke a motion'. The development of the semantics of *āzi* could thus be conceived of as 'to drive' > 'to strive towards' > 'want' > 'greed', which is naturally a possibility.

As an alternative, I suggest that Tocharian B *ñyātse* and the Middle Iranian forms could go back to a zero-grade **aȶah* of the root **h₂m̥ǵʰ-* 'to tie up, to restrain'. The long **ā* could then have developed after the prefix **ni-*, as in Eastern Iranian languages **ni-* + *a-* yielded **ni-ā-* (de Vaan 2003: 34). However, this would mean that forms without the preverb **ni-* would have to be analysed as back-formations, since their initial long *ā-* could not have arisen by sound law: Sogdian ʾʾz /āz/ 'greed', Middle Persian and New Persian *āz* 'greed, lust, covetousness'.

As far as the form is concerned, I would also follow Isebaert's explanation (2019: 269f.), who assumes that an original **ni-anza-* 'nécessité pressante' (derived from **h₂emǵʰ-*) was influenced by **āza-* or **āzi-* 'désir, convoitise, empressement' (derived from **h₂eǵ-* 'conduire, pousser'), yielding a form **ni-āza-*. However, he also assumes semantic influence in the same direction, and as I will try to explain below, this is not necessary.

In my view, the semantics of **ni-ādza-* can be explained from **h₂emǵʰ-* as follows. The primary meaning of **h₂emǵʰ-*, namely 'to tie up, to restrain' is seen in, for example, YAv. *niiāzata* 'tied up',[63] Lat. *ango* 'to compress, tighten', Hittite *ḫamanki* 'binds' (see the LIV²: 264-65 for more examples, see Isebaert 2019 for a similar discussion). This root is also attested in Germanic, cf. Proto-Germanic **angwu-* adj. 'narrow': Gothic *aggwus* adj. 'id.', Dutch *eng* 'scary; narrow' (Kroonen 2013: 28-29). In many branches, derivatives of this root take the meaning 'distress' or, in adjectival derivations

---

[63] Ardvi Sura Anahita Yašt, verse 127: […] *maiδim niiāzata yaϑaca hukərəpta fštāna* [...] "and she tied her middle so that her breasts (are) well-shaped". The form *niiāzata* can also stand for **niiazata* with a short *a*, cf. LIV²: 265 with reference to Hoffmann.

'oppressing, scary, terrible', due to the semantic shift 'tie up' > 'tightened' (> 'narrow'), yielding the notion of a difficulty to breathe, of distress.[64] This can be seen in OHG *angust*, German *Angst* f. 'fear' < Proto-Germanic **angusti-*; French *angoisse* 'great anxiety, terrible worry' < Lat. *angustia* 'narrowing'. The same semantic development can be seen in Avestan *ązah-* 'constriction; distress, peril' < **anđ̌ah* < PIE **h₂emǵʰ-o-*. [65]

The German word *Not* can be translated in English by 'need' as well as by 'distress'. It appears that in Germanic languages, too, the notion of 'distress' precedes that of 'need', although they are both intertwined (Kroonen 2013: 385). The same semantic development must have taken place in Iranian languages. Thus, the meaning 'distress' of **ni-āđa-*, preserved by Tocharian and Parthian, must be older than the meaning 'need', found in Middle and New Persian, among others.

The notion of danger, occurring rarely for this word, could be secondary, possibly a Tocharian-internal development. The meaning 'danger' occurs often in Buddhist contexts, in phrases such as "the danger of the *kleśa*s (passions)", where it could perhaps have originally be used in a sense "the distress (or the calamity) relative to the *kleśa*s". All those meanings, although they reflect different semantic developments, are linked within Iranian to the root **Hanj́-* < **h₂emǵʰ-*.

A formal problem remains in that we have traces of **āzi-*, and none of ***āza-*, and we have no Old Iranian trace of a noun **ni-āza-*: we should ra-

---

[64] Pinault (2019: 394) writes "Ved. *aṁhatí-* fem. […] refers to 'distress', precisely to the fear caused by the feeling of 'narrowness'." This narrowness, in my opinion, is more precisely narrowness in the throat: difficulty to breathe. Being one of the worst feelings man can endure, difficulty to breathe was seen by most peoples as the worst of all pains. This is perhaps one of the reasons why paradise, in the Zoroastrian tradition, is qualified as *vīspō.xᵛāϑrəm* 'all good-breathing' (although this last word is subject to various scholarly interpretations). Another possibility, pointed out to me by Prof. Lubotsky, is that 'narrowness' for nomads meant a lack of pasture, which thus meant death.

[65] The notion of need arises from that of distress, in some languages, through the idea of want, of an urgent or vital need, a vital need, for example food for the starving, or medicine for the dying; compare French *être dans le besoin* 'to be in need', semantically almost identical to *être dans la détresse* 'to be in distress' (cf. also Pahl. *niyāzōmand*, NP *niyāzmand* 'needy', Pahl. *niyāzōmandīh* 'poverty', etc.).

ther expect *ni-āzi-. However, a thematic derivative could have been made in Old Steppe Iranian. We do not have a trace of *ni-āzi- neither, and it is thus possible that *ni-āza-, although it must have been secondary, was the prototype of Parthian niyāz 'need, distress' Middle and New Persian niyāz 'need', etc.

In conclusion, while an Indo-European origin of Tocharian B ñyātse cannot be entirely excluded, it appears that an Old Iranian borrowing is a much more plausible explanation. The problem is that no Old Iranian *niādᶜa- is attested. Therefore, it is advisable to follow Isebaert and assume influence of *āzi- 'greed' on an original *ni-andᶜah- 'distress'. However, the meaning of the Old Steppe Iranian *niādᶜa(h)- that I posit can be derived from PIr. *Hanj́- 'to be narrow, to be tight'.

A note on TB ñyās, ñās, TA ñās 'desire'

Related to the discussion above is the etymology of TB ñyās, ñās and TA ñās 'desire'. The form ñyās 'desire' in Tocharian B is more archaic, and ñās is later (see Peyrot 2008: 63-64). That later form ñās was borrowed by Tocharian A, cf. Peyrot (2010: 140). The word is borrowed from a Middle Iranian form niyāz 'need': either from Sogdian niyāz (SD: 249), Bactrian νιαζο, Parthian or Middle Persian niyāz (DMMPP: 252), or another language.

Here one has to suppose that, in any case, Tocharian B ñyās was simplified from an original Middle Iranian niyāz → Tocharian B *ñəyās > ñyās.

Since those Middle Iranian forms derive from Old Iranian *ni-ādza(h)- (from which TB ñyātse was also borrowed), the meaning 'desire' of ñyās can be seen as being closer to the original meaning *'pressing, vital need' from which the meaning 'need' arose in most Iranian languages.

Malzahn's explanation (2007) that ñās is derived from the verb ñäsk- 'to desire, to seek' and that ñyās is a secondary form (analogically remade from ñās) does not fit the chronological distribution as established by Peyrot (2008: 63-64; see also DTB: 291). However, one could surmise that the meaning 'desire' of the Tocharian form, instead of 'need' as in most of Iranian, could have been influenced by the meaning of the verb ñäsk- .

In conclusion, Tocharian B ñyās, ñās, borrowed into Tocharian A ñās, is itself ultimately a borrowing from a Middle Iranian language. Unfortunately,

the exact source language of this word is undetermined, as the number of languages from which this word could have come is quite large.

2.3.6. TB *twere** 'door'

Tocharian B *twere** 'door', a masculine noun, has been taken as an inherited formation (cf. DTB: 345; Beekes 2011: 35), but I wish to suggest it is alternatively conceivable that it has been borrowed from a hypothetical Old Steppe Iranian *$d\mathring{u}ara$-, also meaning 'door', see for instance Avestan *duuara*- 'gate, door', Wanjī *devur* 'door' < *$d\mathring{u}ara$- (cf. Lashkarbekov 2008: 69, who derives it from *$d\mathring{u}ar$-), Parthian *bar* 'door' (DMMPP: 108), etc.

The Indo-European *$d^h\mathring{u}oro$- from which TB *twere** should have been inherited would have to be a post-PIE form, because the Proto-Indo-European word was a root noun *$d^h\mathring{u}or$- (cf. Beekes 2011: 35). Reflexes of *$d^h\mathring{u}oro$- are found, but they must all be secondary. For instance, Vedic *dvā́ra*- 'doors', only found after the *Śatapatha Brāhmaṇa*, is a secondary thematization of nom.sg. *dvā́r*, acc.sg. *dvā́ram*. Lat. *forēs* pl. f. 'door, gate; the two leaves of a door'; *forus* 'garden surrounded by walls'; *forum* n. 'market place, public space', and OCS *dvorь* 'courtyard' designate a place next to doors, and must thus be secondary derivations. It is thus more likely that Tocharian B *twere** is a loanword from Old Steppe Iranian, which would have continued the Iranian form *$d\mathring{u}ara$-.

Another possible argument in favor of TB *twere** being a loan, is that it is masculine, while the noun *$d^h\mathring{u}or$- from which it should derive was feminine, as it is for instance in Iranian languages. On the other hand, a word such as *santse* f. 'daughter-in-law' was maintained as a feminine noun throughout its history (cf. Peyrot & Meng 2021). They both go back to a feminine *o*-stem (Peyrot & Meng 2021: 407), but it has to be admitted that the word for 'daughter-in-law' is more likely to remain feminine for semantic reasons.

In conclusion, Tocharian B *twere** 'door' could be inherited from post-PIE *$d^h\mathring{u}oro$- as is commonly believed, or it could have been borrowed from an Old Steppe Iranian *$d\mathring{u}ara$-, which has many Iranian cognates, with an identical or almost identical meaning. One argument for it being a loanword is that its preform is post-PIE, and another is that it did not preserve its original feminine gender.

2.3.7 TB *mətstsa*-, TA *nätswā*- 'to starve, to die of hunger'

TB *mətstsa*- and TA *nätswā*- 'to starve' derive from PT **nətswa*- (cf. Peyrot 2013: 790). As noted by Adams (1999: 459; DTB: 493) the initial *m*- in Tocharian B is due to an assimilation of the initial **n*- to the following **-w*-.

Adams (DTB: 493) derives Proto-Tocharian **nətswa*- from "Proto-Indo-European" **n-h₁dˢ-tw-ye/o*-, according to him "a derivative of **h₁ed*- 'eat'". Adams here assumes a sound change **-TT- > -TsT-* which, although it is attested for multiple Indo-European branches, is not known for Tocharian. Also, the supposed formation has no parallels. To my knowledge, apart from this very difficult proposal, the Tocharian A and B verbs did not receive an Indo-European etymology, nor were they explained as loanwords. An argument against the Indo-European inheritance of these verbs is found in the following fact. An Indo-European **n* followed by **u̯* becomes **m* in Proto-Tocharian already: TB *mekwa** (with addition of the plural -*a* suffix), TA *maku* 'nail' < PT **mekwe < *h₃nogʰᵘo*- (cf. Krause 1971: 10 – I do not understand why there was no *p/m*-umlaut in Tocharian A, as the expected form should be **moku). Here, we have to reconstruct Proto-Tocharian **nətswa*- with an *n*-, because of the Tocharian A form, so that it has to be a later sound change.[66]

I believe that TB *mətstsa*- and TA *nätswā*- are derived directly from an unattested Proto-Tocharian noun **netsu*-, the **-u*- of which is visible in the verb (PT **-tsw*-). This noun would have either meant 'corpse' or 'demoness of Death', and would have been borrowed either from Old Steppe Iranian **natˢu*- masc. 'corpse' or from **natˢu*- fem. 'corpse; demoness of Death', in its turn from PIr. **naću*- 'corpse', compare Avestan *nasu*- 'corpse; demoness of Death' (from PIE **nek̂u*-). From the Proto-Tocharian noun **netsu*-, a verb was made, **nətswa*-, in the same way that PT **reytwe* 'mixture' yielded a verb **rəytwa*- 'to mix'. Interestingly, TB *eṅkwe* 'man', TA *oṅk* 'id.' derive from the same Indo-European root as OSIr. **natˢu*-, through the semantic development 'mortal' > (*'human being' >) 'man' (cf. DTB: 83).

---

[66] It is unclear to me whether the sound change that effected **nətswa*- in Tocharian B is "the same" as in Proto-Indo-European, that never stopped being effective in Tocharian B, or whether it is an identical sound change that occurred a second time in the language.

As to the semantics, there are two different pathways: either *netsu- was borrowed with religious semantics, as 'the demoness of Death', the verb meaning 'to be overcome by the demoness of Death' > 'to die (as a process)' > 'to die of hunger, to starve', exactly like in Old, Middle and Early New English, where *to starve* indicates dying as a process, as opposed to *to die*, which meant 'to cease to live'.

One can compare also Norwegian (Nynorsk) *starva* 'frieren, dem Tode nahe sein' (cf. Holthausen 1963: 320). In English too, *to starve* could mean 'to freeze to death' (Middle English and English) and, transitively, 'to make someone die of hunger'. In this sense, in Tocharian, from a noun 'demoness of Death', a verb meaning 'dem Tode nahe sein' could be made.

The second pathway, which seems a little bit more difficult to me, would make the word go through *'corpse' > *'stiff/thin like a corpse' (perhaps an adjective?) → 'to become thin like a corpse' > 'to starve'. In this case, it is perhaps possible to view the English and Norwegian meanings of 'to freeze to death', and perhaps even 'to starve' as ancient archaic meanings (although they do not appear in Old English, as far as I could find), relative to the original meaning of the verb 'to become stiff' (Kroonen 2013: 477).

In conclusion, PT *nətswa- is likely the verbal derivative of an unattested Proto-Tocharian noun *nets^u-, borrowed from an Old Steppe Iranian noun *nat^{s}u- (borrowed as PT *nets^u), a cognate of Avestan *nasu-* 'dead matter, demoness of Death'. However, it is not entirely sure whether PT *nətswa- was derived from a noun meaning *'corpse' or *'demoness of Death'.

2.3.8 TB *mot* 'alcoholic beverage'

For a long time, it was believed that TB *mot* 'alcoholic beverage' was borrowed from Sogdian *mwδ* 'wine' (e.g. Pinault 2003a: 183). However, things are more complicated: the normal Sogdian word for wine is *mδw* /mᵊδu/ (SD: 210).

The form *mwδy* from the Ancient Letter IV, 1.5. is now recognized as meaning 'price' (see Dragoni 2021: 302, with reference to literature). A Christian Sogdian form *mwd[* 'wine' exists, but it is attested in a very late text, and would be too late for Tocharian (cf. Dragoni 2021: 302[31]). In any case, this form *mwd[* likely stands for /muδ/, with *u*-umlaut of the schwa, and thus phonetically it is too far off the Tocharian word.

To explain the Tocharian B form *mot*, one can think of three solutions: 1. it is a borrowing from an unattested dialect of Sogdian, which had *mwδ* (< *madu-*) at an earlier stage than late Christian Sogdian *mwd[* 'wine'. This is unlikely, as no other such example exists. 2. It is a borrowing from an unknown Middle Iranian language. This is very speculative. 3. It is a borrowing from Old Steppe Iranian *mod(u)-* < PIr. *madu-* (or *maδu-*) 'mead'.

What could permit us to posit such an umlaut for Old Steppe Iranian is the parallel of Ossetic: in Ossetic, PIr. *madu-* 'mead' became Iron *myd*, Digoron *mud* 'honey'. This *u*-umlaut is documented in Ossetic, with the condition that the *-u* or *-ū* should have been in word-final position and there should be a labial consonant in the word (cf. Cheung 2002: 124f.). According to Cheung (2002: 125) an intermediary stage of the umlaut was a realization of the *a* as [ɔ]. If Old Steppe Iranian had an *[ɔ] in this word, it would have been rendered as *o* by Tocharian speakers, who, as far as we know, had no phoneme /ɔ/. Otherwise, we can simply assume that the intermediary form was close enough to Tocharian *o*.

Even though this derivation remains speculative since no other case of Old Steppe Iranian *u*-umlaut is recorded, it is nevertheless clear that Tocharian B *mot* 'alcoholic beverage' must ultimately be related to Proto-Iranian *madu-* (or *maδu-*) 'mead'.

In conclusion, the exact origin of TB *mot* 'alcoholic beverage' is unclear for the moment: it could be of Old Steppe Iranian origin but since we lack clear parallels for the sound change it presupposes, this hypothesis remains speculative. In theory, it could alternatively come from some unknown or unidentified Middle Iranian language.

2.3.9 TB *wertsiya* 'assembly', TA *wartsi* 'assembly, crowd'

The Tocharian B word *wertsiya* 'assembly' (verse variant *wertsya*, cf. Pinault 2006a: 82) and its Tocharian A equivalent *wartsi* 'crowd' have been derived from Proto-Indo-European *$H\underset{\ }{u}ord^h\underset{\ }{i}eh_2$-* 'mass, multitude', from *$H\underset{\ }{u}erd^h$-* 'to grow' by Adams (DTB: 665; see also Del Tomba 2020: 168).

In my view, there are several problems with Adams's etymology. First of all, *$H\underset{\ }{u}erd^h$-* means 'to grow (for example, a plant), to make grow (animals, babies); to make strong' and 'to become strong'. These meanings can also be seen in various Uralic loanwords from Indo-Iranian, which have the meanings 'to rear animals', 'to raise children, to give birth'; 'to feed' (cf. Holo-

painen 2019: 311f.). Second, *$Huerd^h$*- does not have any assured cognates outside of Indo-Iranian, cf. LIV²: 228. To assume that, first, this root should have also existed in Tocharian, and second, that it would have evolved to mean 'mass' (which is far from evident) and then 'assembly; crowd' in Proto-Tocharian implies just too many steps for it to be likely.

Lastly, to explain the Proto-Tocharian form *wertsiya* one needs to have both *$Huerd^h i$*- (to explain the *ts < *-$d^h i$-) and, secondarily, *-iya*. Thus, for this derivation, a Proto-Indo-European *i would be needed twice.

In my view, a more straightforward etymology can be suggested. I propose that *wertsiya* is a borrowing from Old Steppe Iranian *$u̯ard^{ɟ}ii̯ā$-, a cognate of *$u̯r̥d^{ɟ}āna$- 'community', cf. OAv. *vərəzāna*-, YAv. *vərəzāna*- 'community'; OAv. *vərəzāniia*- m. 'member of the community', *varəzāna*- adj. 'belonging to the community' (cf. AiW: 1378f., 1424f.).

These words could derive from *$u̯r̥d^{ɟ}ā$-, which presumably meant 'habitation', but also 'village; city'. I suppose that the Avestan and Old Steppe Iranian words were originally substantivized adjectival formations meaning 'those of the village' or 'those of the city' > 'the community'. This could explain why we have here a feminine suffix *-$ii̯ā$-, denoting appurtenance. As in the Old Steppe Iranian source of the Tocharian B borrowing *newiya* 'canal', this feminine suffix is in origin the substantivization of a feminine adjective.

Unfortunately, the *-$ii̯ā$- formation that I set up to explain the Tocharian word is not yet attested in Iranian. Alternatively, one could consider that Proto-Tocharian borrowed *$u̯r̥d^{ɟ}$- 'village, city' and that the suffix *-iya* was added by the Tocharian speakers themselves, following the pattern outlined by Del Tomba (2020: 168).

If my derivation of Tocharian B *wertsiya* and Tocharian A *wartsi* from Old Steppe Iranian *$u̯ardzii̯ā$- is correct, the Old Steppe Iranian word would show another instance of *-$r̥$- > *-ar- (cf. p. 172f.).

A problem for which I have no definitive solution is why the *r has remained *r in front of *$d^{ɟ}$ instead of changing to *l. Possibly, the change to *l occurred only before *d, and not before *$d^{ɟ}$, see section 2.6.2.k.

In conclusion, an Iranian source is very likely for Tocharian B *wertsiya* 'assembly, community', Tocharian A *wartsi* 'crowd'. The closest Iranian cognates resemble the Tocharian words very much formally, and, based on their attested cognates, it is likely that their semantics would have been very close to the Tocharian meanings of these words.

2.3.10 TB *welke** 'stalk (?), petal (?)'

Tocharian B *welki*, the plural of an unattested noun *welke**, occurs twice. Once in W 11 b5, where it is the only legible word of the line, and once in W 32 b2, which contains a medical recipe. The context as given by Adams (DTB: 665) is:

W 32 b2

> *ke̬-pyapy(ai)ntse welki • eñcuwañe keṃtse • te ṣeme yarm*
> "the petals of a dandelion [?] and iron-rust, each the same measure" (Adams DTB: 665).

> "… de fleurs …, de terre ferrugineuse ; cela, une mesure […]" (Filliozat 1948: 86).

As noted by Adams (DTB: 665), *welke** (pl. *welki*) designates a part of a flower. He suggests that it can be compared to Vedic *valká-* 'bast, bast fibre' and Avestan *varka-* 'leaf', or Vedic *válśa-* 'shoot, twig'.

In my view, there is a much more straightforward etymology: Old Steppe Iranian **u̯aldaka-* (with PIr. **rd* > OSIr. **ld*, cf. p. 179f.) 'of the flower', that is, either '(petal) of the flower' or '(stalk) of the flower', but more probably petals, as it is used as an uncountable quantity in the Tocharian B text, and petals were more commonly used as ingredients, and less countable than bare stalks of flowers.

*u̯aldaka-, which was probably accented on the first syllable, should regularly have become **weltke* in Proto-Tocharian. The simplification **ltk* > **lk(k)* would have occurred in the historical period, after the archaic stage (cf. Peyrot 2020), see for example TB *spelke*, *spelkke* 'zeal' from *speltke*, cf. TA *spaltäk* 'id.'.

To conclude, *welke**, which designates a part of a flower, either a stalk or a petal, used in a recipe, may be borrowed from Old Steppe Iranian **u̯aldaka-*, derived from PIr. **u̯arda-/*u̯r̥da-* 'flower' (on the development of Proto-Iranian vocalic **r̥*, see section 2.6.2.l), with the addition of the suffix *-ka-*.

2.3.11 TB *tsetke* 'a hundred, *centaine* (?), century (?)'

Tocharian B *tsetke* is found in two fragments: IOL Toch 158 and THT 1928. Below I cite the contexts in which it occurs:

IOL Toch 158 a4

/// *kalymi • tsetkesa tucenän kaunän*

"… direction … • By *tsetke* … yellow suns …"

THT 1928 b6

/// *·m̥ – – – 89 tsetke prākre snai – ///*

"… *tsetke* solid without […]"

Unfortunately, the meaning of the word cannot be established from these fragmentary occurrences. However, the phonological structure of the word is strongly suggestive of Old Steppe Iranian origin: 1) the initial *ts-* need not necessarily, but could well be of Old Steppe Origin; 2) the vowels *e_e* likewise are not necessarily of Old Steppe Origin, but are extremely frequent in words from that source; 3) the final *-tke* is a strong indication of Old Steppe Iranian origin. This latter feature could perhaps have been taken over from a verb if the word was of Proto-Indo-European origin, i.e. *tsətk-* or the like, but in Old Steppe Iranian vocabulary *-tke* is highly frequent. Even though none of these features is absolutely probative, together they make borrowing from Old Steppe Iranian a serious option to consider.

If borrowed from Old Steppe Iranian, *tsetke* should go back to either *$\!\ast\!\widecheck{t\!}aTaka$-* (with *T* representing either **t* or **d*) or *$\ast t^{\!\!s}aTaka$-*. This easily leads to a possible etymon: *$\ast t^{\!\!s}ataka$-* 'centaine, around a hundred'. Old Steppe Iranian *$\ast t^{\!\!s}ataka$-* would be a **-ka-* derivative of *$\ast t^{\!\!s}ata$-* 'hundred' (< PIE *$\ast(d)\hat{k}mto$-*), see for instance Middle Persian *sadag* 'centaine (around a hundred); century'; New Persian *sada* 'century' (and, of course, MP *hazārag* 'millenium' ← *hazār* 'thousand', cf. CPD: 43). The syncope of the second vowel of *$\ast t^{\!\!s}ataka$-* is regular (cf. section 2.6.2.g).

Although I stress that the meaning of *tsetke* cannot be established independently at this point, it seems to me that "a hundred" is compatible with the attestation in IOL Toch 158. Since there is no English equivalent to French *centaine*, which is to a hundred what a *dozen* (← Fr. *douzaine*) is to 'twelve', I kept the French word '*centaine*' to translate *tsetke*. For this reason I also think it would be easier to translate IOL Toch 158 a4 to French:

"des soleils jaunes par centaines" "yellow suns by the hundred"

The perlative of *tsetkesa* also corresponds to Fr. *par* in *par centaines* (and also to English *by* in *by the hundred*). A full translation for THT 1928 b6 is not feasible. As Dr. Dragoni informed me (p.c.) this could be "a variant of a

frequent Buddhist stock phrase referring to the innumerable Buddha suns, which are often hundreds or thousands in number" (cf. Skjærvø 2004: I, 68-69). An issue with this explanation is that this expression of the "many Buddha suns" is typically Mahāyāna, which is a branch of Buddhism. The Khotanese were Mahāyāna, but the Tocharians were Theravāda, which is another branch of Buddhism, where this phrase does not occur. However, as Athanaric Huard told me, the type of texts to which this Tocharian text belongs (meditation texts) "multiplies mental objects" and mentions elements in extraordinary numbers, filling the universe with them, so that the meaning 'centaines'

As a conclusion, the meaning of Tocharian B *tsetke*, which appears in IOL Toch 158 a4 and THT 1928 b6, cannot be firmly established. As far as its phonological structure is concerned, the word could well be of Old Steppe Iranian origin. A possible source would then be Old Steppe Iranian **tˢataka-*, a **-ka-* derivative of **tˢata-* 'hundred', the reflex of which is found in all Iranian languages, for instance Avestan *satəm*, Middle and New Persian *sad* 'a hundred', and this could be supported by a possible analysis of one occurrence as a frequent Buddhist literary image.[67]

2.3.12 TB *tsere* 'a measure of liquid volume of half a *lwāke* (?)' or 'cup (?)'

The Tocharian B word *tsere* designates "a kind of vessel or […] even a measure of capacity equal to 0.5 *lwāke*" (cf. Ching 2011: 68[21]). It is worth, in size or volume, half a *lwāke* 'jar'.

As a jar could very well fill only two big cups, TB *tsere* might have designated a cup, and could come from Old Steppe Iranian **tˢara-* 'head' (Avestan *sara-*, Persian *sar*, etc.). An argument in favor of this etymology could be the practice of drinking from dead enemies' skulls as famously reported about the Scythians by Herodotus.

Perhaps, less cruelly, one can also think of a metaphorical designation. This is a known process, where the parts of the jug are identified with parts of the human body. For instance, one can think of Dutch *kop* 'cup; head', where the meaning 'cup' is original, cf. also Latin *testa* 'jug, cup' > French

---

[67] Since Middle Persian *sadag* < **θataka-* < **tˢata-ka-* meant 'century (a hundred years)', one can also imagine Old Steppe Iranian word had this meaning to.

tête 'head'.[68] With suffixes, one can also think of English *handle*, Persian *dasta* 'handle' < **dastaka-* built on **dasta-* 'hand'.[69]

Evidently, a 'head' may be called a 'cup' because of their resemblance. The Old Steppe Iranians would have processed in the opposite direction, and called a 'cup' a head. Unfortunately, I could not find an example among Iranian languages of **tˢara-* (or a derivative therefrom) taking on the meaning 'cup'.

An advantage of this etymology is that it works perfectly formally, and there are some parallels, although not exact ones. PT *tsere could very easily derive from Old Steppe Iranian **tˢara-*. Two disadvantages of this etymology lie in the fact that (1) the Tocharian meaning is not certain and (2), as far as I know, there are no examples of the specific shift from 'head' to 'cup' among Iranian languages.

2.4 Old Steppe Iranian borrowings: difficult cases

The words discussed in this section are of possible Old Steppe Iranian origin. However, they present serious difficulties, e.g. clear Iranian cognates are lacking, the form is both problematic and its meaning unclear, and so on. In the case of TA *kump* 'cauldron', for instance, where an Old Steppe Iranian etymology is possible, a Middle Indic one is possible as well. The meaning of TB *tseriteke* is not clear, but its *-i-* is also unexpected, so that an etymology will always be doubtful until at least its meaning is assured. The Old

---

[68] I would even like to suggest that the ʔ in Arabic *kāʔs* 'cup, glass', which is not etymological, is due to the influence of Ar. *rāʔs* 'head', where the ʔ is inherited (compare Hebrew *rōʔš* 'head').

[69] On this theme, one can evoke Omar Khayyām's moving quatrain:

Īn kūza čō man āšiq=i zār=ē būda=st
Dar band=i sar=i zulf=i nigār=ē būda=st
Īn dasta ki bar gardan=i ō mē=bīn-ī
Dast=ē=st ki bar gardan=i yār=ē būda=st

"This jug was a desperate lover, like me, / (Like me,) it was in chains, tied to the hair of an idol / This handle that thou seest on its neck / It was an arm around a lover's neck." As we all become earth and soil, this soil is used by potters to make jugs, which, once part of the human body, now mimic the body, as the *tsere* of the Tocharians perhaps reminded them of the human head. In them we drink wine, and tomorrow, we will become the soil from which new jugs are made.

Steppe Iranian origin of TB *ecce*, TA *aci* 'hither, from' I find tempting, but there remain many problems in the details of this etymology.

2.4.1 TB *ecce*, TA *aci* 'hither, from'

Tocharian B *ecce*, Tocharian A *aci* 'hither, from' both regularly go back to PT **ecye*. These words remain without clear etymology (cf. DTB: 83). The meaning of these words was discussed in detail by Winter (1984). He concluded (1984: 122) that they are synonymous with TB *śka* 'hither'.

Dunkel (2014: 262) has proposed an Indo-European etymology for PT **ecye*: an Indo-European particle **h₁oti* 'away from' → neuter adjective **oti̯om* > PT **ecce*. This etymology seems implausible to me because **ti̯* would have yielded PT **ts* rather than **cy* or **cc*, and also because **h₁oti* is only supported by two very doubtful Balto-Slavic forms (see Dunkel 2014: 262).[70]

I am tempted to connect TB *ecce*, TA *aci* to Proto-Iranian **hačā* 'from', cf. Old Persian and Avestan *hačā*, and to see in it a borrowing from Old Steppe Iranian. However, **ecye* cannot have been borrowed from **hačā* directly, as this would have yielded PT ***eca*. Therefore, in order for this etymology to work, I have to posit a derived adjective **hačā-i̯a-*. As no such adjectival formation is attested, and no close parallel for this formation can be found either, I classify this etymology as difficult.

If a derived adjective **hačā-i̯a-* existed, it could have become **(h)ačai̯a-* through shortening of **ā* in front of *yod* in Old Steppe Iranian according to the rule discussed in section 2.6.2.i. This hypothetical Old Steppe Iranian **(h)ačai̯a-* would probably have become Proto-Tocharian **ecye* according to regular syncope rules (cf. Winter 1994).

In the posited Old Steppe Iranian **(h)ačai̯a-*, the suffix **-(i)i̯a* would be adjectival, and then the neutral adjective would have become an adverb again, according to the well-known circle adverb → derived adjective → (instrumental function or form of the adjective) > adverb. A parallel can for instance be found in Latin: *super(i)* adv. 'above' → adj. (**superinos* >) *supernus* > adv. *superne*.

---

[70] A connection with PIE **h₁éti* is also impossible: this would have yielded PT **yəc*, like PIE **h₁ék̂u̯o-* > PT **yəkwe* > TA *yuk*, TB *yakwe* 'horse'.

A point which could be useful to mention in connection to this etymology is that usually two variants of *hačā 'from' are assumed to have existed in Old Persian: *háčā and *hačā́, the first yielding New Persian *az*, and the second yielding New Persian *zi*, both meaning 'from'. I wonder, however, if New Persian *zi* cannot be derived from *hačíi̯a-. Formally, this is not impossible, but it has not, to my knowledge, been proposed before. Since there are no other examples of *-á > -i, the alternative derivation from *hačíi̯a- may even have to be preferred. A form *hačii̯a- would be parallel to *hačā-i̯a-, but it would have to be derived from *hač- rather than *hačā. Since no such base *hač- is attested, my suggestion has to remain hypothetical for the time being.

In conclusion, I have proposed to consider Tocharian B *ecce*, Tocharian A *aci* 'hither, from', which both go back to Proto-Tocharian *ecye*, as ultimately borrowed from an Old Steppe Iranian adverb, itself derived from an adjective, *(h)ačai̯a- which could also have meant 'hither' and ultimately goes back to the preposition *hačā 'from' suffixed with the adjectival suffix *-i̯a-. However, this etymology remains difficult since *hačā̆i̯a- is attested nowhere and its derivational pattern has no parallels.

2.4.2 TB *eprete* 'resolute, steadfast'

The Tocharian B word *eprete* 'resolute, steadfast' is considered by Adams (following earlier literature, see DTB: 95 with references) as being constituted of the prefix *e(n)-* and *prete-* 'decision, resolution', "the unattested TchB counterpart of TchA *pratim* (the latter borrowed in B as *pratiṃ*)." (DTB: 95; similar view in Pinault 2008: 334).

Adams' interpretation (already in 1999: 90) has been accepted in the scientific literature (cf. LIV[2]: 493). This form *prete-* has been connected to a root *pret-* 'erkennen', with the two cognates Gothic *fraþjan* and Lithuanian *prantù*, both meaning 'to understand', adduced by the LIV[2]: 493.

This etymology is not without problems. As far as we know, there is no clear example of a borrowing from Tocharian A into Tocharian B, while the other direction is much more common. Because of its semantics, a word such as TB *pratiṃ* 'decision, resolve, conversion' belongs to a rather intellectual or literary stratum, making it even more unlikely to have been borrowed from Tocharian A.

Under a different entry, Adams himself (DTB: 442) suggests that TA *pratim* was borrowed from TB *pratiṃ*, and not the other way around, contradicting himself on this point. He proposes that TB *pratiṃ* is ultimately derived from "the Prakrit descendant of Sanskrit *pratijña-* 'agreement, promise, decision, assertion'". Ji (1998: 291) also provides a plausible etymology in Sanskrit *pratimā-* 'decision, likeness, symbol'. No matter the etymology of Tocharian A *pratim*, Tocharian B *pratiṃ*, these two cannot be formally compared to Tocharian B *eprete*, despite the semantic closeness of all those words.

Apart from this, there is no Tocharian B word †*prete* (*vel sim.*) from which *eprete* would have been made, nor is there a Tocharian A cognate of this word (†*prat* ?).

In my view, it is much more likely that TB *eprete* regularly derives from an Old Steppe Iranian word **abi-ratu-* 'on (= according to) the decision, the judgment'. On the meaning 'decision, judgment' of *ratu-*, see for instance Bartholomae with Old Avestan *ratu-* 'judicium, Richterspruch' (AiW: 1502).

If **abi-ratu-* existed and was borrowed, it would most likely have been rendered as PT **epəret*, with regular loss of the *-u* (s.v. TB *peret*, TA *porat* for instance); and reduction of **abi-* to **epə-* (s.v. *epastye* for another such example). This reconstructed form **epəret* would then have been reduced, possibly already in Proto-Tocharian, to **epret*. In Tocharian, a stem vowel *-e* would have been added secondarily to make adjectival inflection possible.

If the word was **ep(ä)rete* in Proto-Tocharian, when the syncope took place (see for instance section 2.6.2.g), the word would have been reduced to either **epärte or **epräte, because **é_e_e* or **e_ə́_e_e* was systematically reduced to **e_0_e* in Proto-Tocharian. This suggests that the adjectival *-e* was added after the Proto-Tocharian stage.

In conclusion, Tocharian B *eprete* 'resolute, steadfast', which does not have a convincing etymology so far, may derive from Old Steppe Iranian **abi-ratu-* 'according to the resolution, the judgment', which one can further understand as 'following one's resolution, steadfast'.

2.4.3 TA *kump* 'cauldron'

The Tocharian A word *kump* is often translated as 'pot' (e.g. Carling 2009: 150; Meunier 2015: 280). The context of most attestations is fragmentary.

However, in one well-preserved text, which I give below, the meaning is clearly 'cauldron' rather than 'pot'.

A 341 a7

*lyalypurṣi want worta-ṃ kumpaṃ ṣūci-śpāl pakt-äṃ ymār tmaṃ kapśañi :
ṣñi lyalypäṃntu(yo) ///*

"The wind of deeds threw her in the caldron. She quickly boiled her body [in it] as a needle-headed (preta). (By) her deeds … […]" (adapted from CEToM, with a new reading *ṣūci-śpāl* 'Needle-Headed', name of a preta, rather than *ṣūti śpāl*, previously translated as 'head first').

It is, I believe, unthinkable that an entire woman's body would fit in a pot. The translation 'pot' is based on the Sanskrit word *kumbha-* 'jar, pitcher, water pot, ewer, small water-jar' (M-W: 293). However, in this precise context, it should be a cauldron, in which an entire woman's body could fit. This is confirmed by the Chinese parallel (Chavannes 1911: 251), which has a word translated by Chavannes as *marmite* (cauldron, big cooking pot) and *chaudière* (boiler, furnace). In particular, we read in the Chinese text that: "une marmite à trois pieds apparut ; un feu de charbon la faisait bouillonner ; cette femme enleva ses vêtements, les mit de côté et entra dans la chaudière ; sa peau et sa chair furent entièrement cuites ; il ne resta plus que de petits morceaux d'os ; mais alors un vent frais vint à souffler ; elle put sortir de la marmite et revenir à la vie ; elle mit ses vêtements et dévora sa chair cuite. Quand elle l'eut dévorée, elle partit." (Chavannes 1911: 251).

Clearly, the object that is mentioned in both the Chinese and Tocharian versions is very different from the water-jar or pitcher that is denoted by Sanskrit *kumbha-*. It is bigger, and used for cooking, or at least for boiling water.

This text is an extract of the *Koṭikarṇa-Avadāna*, which is also part of the Sanskrit *Divyāvadāna*. Sieg (1952: 37) writes that the Chinese version is closer to the Tocharian one than both are to the Sanskrit text. In the Sanskrit version, it is only said that "whatever food she eats turns into her own flesh." (cf. Rotman 2008: 74).

Another indication on the meaning of TA *kump* is the distributive *kumpa-kump* (e.g. A 318 a1; YQ III.2 b4), translated as 'pot by pot' by Adams (2017: 1384). It seems to me that the correct meaning should rather be 'group by group' or 'crowd by crowd' (cf. Carling 2009: 151, "in crowds"), as pot by pot does not fit the relevant contexts, where it often occurs together

with *kropa-krop* 'group by group' (cf. Carling 2009: 174 "crowd by crowd"). However, the meaning 'crowd by crowd' or 'group by group' cannot be easily derived from 'cauldron-per-cauldron', since cauldrons rarely come in groups.

In my view the original meaning was rather 'cauldronful by cauldronful', which exists with a distributive use also in other languages (cf. Ugaritic *dd dd*, Del Olmo Lete & Sanmartín 2003: 265), or perhaps better, in order to give the meaning of various people together, 'quantity (of what is contained in the *kump*) by quantity'. Strangely enough, *kumbha-* in Sanskrit also designates a rather large quantity: a measure of grain, which is "equal to twenty Droṇas, a little more than three bushels and three gallons; some make it two Droṇas [...]" (M-W: 293). Even only two droṇas would be equal to more than twenty kilos. There is a real discrepancy between *kumbha-* as a receptacle, which is rather small, and as a measure of quantity, which is very large.

To bring some nuance to these facts, it should be noted that in some Prakrits *kumbhī-* fem. designates a 'large round pot', or even, in some modern languages, a pool or a bucket (cf. CDIAL: 170). In that sense, it is not impossible that Tocharian A *kump* 'cauldron' comes from a Middle Indic language.

In Iranian, reflexes of *xumba-* designate either a pot (Av. *xumba-*), a pool (Balochi *kumb*), or a jar (Persian *xum(b)*, cf. Horn (1898-1901: 59). Kumzari, a Persic language, has a word *xumba* (< *xumbaka-*) 'clay storage jar'. Interestingly Wakhi has a word *xəmbák*, which designates a very big bin or chest to store grain (cf. Steblin-Kamenskij 1999: 411). It thus seems that the discrepancy in sizes is seen among Iranian languages as well.

In these languages, the *xumba-* is either a jar to keep water or other liquids, as in Sanskrit, or it is a big container that contains grain (it is likely that Balochi *kumb* initially designated a grain-storage pool). The two meanings found in Sanskrit thus may be connected by comparison with Iranian languages. I cannot easily understand the two distant meanings: 'jar, pot' and 'big container of grain', but both meanings, that is, small jar and big container, both being storage tools, might have coexisted originally in Indo-Iranian.

A scenario that I can propose here is that Proto-Tocharian borrowed *kumpe* from Old Steppe Iranian *xumba-*, which had the meaning 'big receptacle', possibly used as a kitchenware, and which gave the meaning 'cauldron' to TA *kump*. Because the meaning 'cauldron' is absent in Indo-

Aryan, it is much more likely, in my view, that Tocharian A *kump* 'cauldron' ultimately goes back to an Old Steppe Iranian form.

It is striking that there is no Tocharian B †*kumpe* word or the like. To translate Sanskrit *kumbha-*, TB uses the word *lwāke* (possibly inherited, see Garnier & François 2020: 51-52), and the semantic equivalent of TA *kump* 'cauldron' seems to be TB *aise*, which Pinault (2008: 127) plausibly derives from the Proto-Indo-European root *$h_1ei̯$- meaning 'to be hot'.

However, it is also possible that TA *kump* was borrowed from a Middle Iranian language with the meaning 'cauldron'. This would make the absence of the word in Tocharian B less problematic, especially given that the Tocharian B equivalent might be inherited. Of course, we cannot exclude that there was a technical difference between *kumpe* and *aise*, which would have disappeared due to lexical generalization at a later stage.

To conclude, I propose that Tocharian A *kump* 'cauldron' derives from Old Steppe Iranian *xumba-*, which should also have meant 'big receptacle', and has many cognates among Iranian languages. In my view, the meaning 'big receptacle' can be reconstructed for Proto-Iranian, and it could easily have evolved to mean 'cauldron' through the meaning *'big receptacle (for cooking)'. However, a Middle Indic or Middle Iranian origin of this word cannot be entirely excluded.

2.4.4 TB *kercci* 'palace'

The Tocharian B word *kercci* (variants *kerci* and *kerc(c)ī*, oblique *kerciyeṃ**) means 'palace'. Its etymology is difficult. I will present the data, then discuss previous etymologies (2.4.4.1), detail a specific French etymology which I believe has had an impact on the previous etymologies (2.4.4.2), then explain my own etymology and problems relative to it (2.4.4.3).

TB *kercci* is a *plurale tantum*, whose nominative plural ending *-i*, written ‹i› or ‹ī›, comes from *-iyi*, cf. Peyrot (2021: 458). It occurs in many texts. In the past, two of its occurrences have been interpreted as the plural of the word *kertte* 'sword', but Del Tomba has convincingly shown that for those two occurrences the meaning 'palace' is preferable (s.v. *kertte*; cf. Del Tomba 2020: 258; Del Tomba 2020a). Tocharian B *kercci* 'palace' was borrowed into Old Uyghur as *karšı* '(royal) palace' and, from there, into Mongolian (cf. Clauson 1972: 664).

Tocharian A does not have an equivalent of TB *kercci* but uses a peri-phrastic figure *lāñci waṣt*, literally 'royal house', to designate a palace (cf. Meunier 2015: 84). Because of this, I believe the current meaning of *kercci* to be a secondary, post-Proto-Tocharian development. It is rather unnatural for a language that has a perfectly fitting word to discard it in favour of a more complicated periphrasis. In the following pages, I will argue that the word *kercci*, a *plurale tantum*, originally meant 'servants, people of the house', before meaning 'palace', and that it is likely an Iranian loanword. It is possible that the unattested Tocharian A cognate of Tocharian B *kercci* existed (perhaps under the form **karcañ* or **karccañ*)[71] took on another meaning, and no longer corresponded to 'palace', but this seems less likely and less economical than the option that Tocharian B *kercci* took on the meaning 'palace' secondarily.[72]

2.4.4.1 Previous etymologies

Previous etymologies of Tocharian B *kercci* 'palace' can be classified in two categories: inherited from Proto-Indo-European or borrowed from Iranian.

Of the first kind, Meillet (in Hoernle 1916: 379) and Lidén (1916: 21-2) proposed to connect TB *kercci* to Gothic *gards* 'house', Old Norse *garðr* 'fence, hedge, court', Old English *ġeard* 'enclosed space, garden, dwelling; land, region' (Modern English *yard*), OCS *gradŭ* 'enclosure, city', Sanskrit *gṛhá-* 'house, habitation, home', etc. Adams (DTB: 210-11) hesitates be-tween cognacy to this group of words and to another group of words: either "PIE **gʰordʰii̯o-*" or what he calls a "putative PIE **gʰortiyo-*" (DTB: 210), which would in its turn be related to Greek χόρτος 'enclosed place, feeding place', Latin *hortus* m. 'garden', Latin *cohors* 'yard, enclosure', and possibly to English *garden*.

These two reconstructions, namely **gʰordʰii̯o-* and **gʰorto-*, are connect-ed, although hesitantly, by Beekes (2010: 1645). The formal discrepancy

---

[71] According to Michaël Peyrot (p.c.), the plural (PT) **kercciyeñ* would have be-come (Pre-A) **karccyañ* > TA **karcañ* or **karccañ*, the geminate of which would be due to the **-cy-* cluster, similarly to *opäśśi* 'expert' which derives from **epəstəye* > TB *epastye* 'idem', s.v. TB *epastye*, TA *opäśśi*.

[72] The Tocharian A word *wimāṃ*, Tocharian B *wimāṃ**, from Buddhist Sanskrit *vimāna-*, designates a specific divine type of palace (cf. DTB: 656).

between all those forms indicates that there is a real problem in their recon-struction.[73] De Vaan (2008: 291) provides a solution: for him the Latin noun reflecting *-to- (and its Indo-European cognates) "might be regarded as a verbal adjective to a root *$g^h r$- 'to enclose'". He also reconstructs two PIE forms: *$g^h r$-to- and *$g^h or$-$d^h o$- which might have contaminated each other in the daughter languages or in PIE itself (2008: 291).[74]

On the borrowed side, Isebaert (1980: 88, cf. also p. 116) derives Tochar-ian B *kercci* from Old Iranian *$gardii̯a$- 'servant, courtesan', corresponding to Sanskrit *gr̥hya*- 'servant', etymologically '(the one) of the house'. In the plural, this would mean 'entourage, (royal) court'.[75]

Tremblay (2005: 426-27) apparently did not understand Isebaert's ety-mology, since he writes that Joki and Isebaert agree that TB *kercci* derives from Old Iranian *gr̥da*- 'house'. He himself proposes two solutions: the first is a borrowing from Old Iranian *gr̥da*- > Proto-Tocharian *karta* (his notation) → adjective *karciya*-, made within Proto-Tocharian. That adjec-tive "eventually ousted its related substantive". Tremblay's second solution is that PT *kerciye*- "was borrowed from a *vr̥ddhi*ed collective *$gardii̯a$- (instead of the expected *$i̯ardii̯a$-)." (2005: 427). Despite all this, Tremblay (2005: 427) believes that Adams' etymological proposal from either *$g^h ord^h ii̯o$- or *$g^h ortii̯o$- is also possible.

As regards semantics, the etymological proposals mentioned here can be divided in two groups:

---

[73] Cf. LIV[2]: 197, where the verbal root is reconstructed as *$g^h erd^h$- 'umschließen, umgürten'.

[74] The forms going back to *$g^h ord^h o$- could ultimately reflect a univerbated PIE compound *$g^h r̥$-$d^h h_1$-ó 'enclosed'. There is a parallel for this formation in Hittite *u̯arpa dai-[i]* 'to enclose' (cf. Kloekhorst 2008: 966).

[75] Joki (1973: 269) suggests that Tocharian B *kercci* derives from the Old Scythian ancestor of the Ossetic word Iron *kært* 'Hof, Bauernhof', adding that the ultimate etymology of the Ossetic word is uncertain. Normally *rd* and *rt* yield Ossetic *rd* (cf. Cheung 2002: 29) and only *θr* yields Ossetic *rt* (cf. Cheung 2002: 38). There are some cases of word-final devoicing in some Ossetic words, including the variant *art* of *ard* 'oath' (cf. Cheung 2002: 38). However, the initial *k*- in *kært* cannot derive from an ancient *g*- (see Cheung 2002: 22). In Joki's opinion (op. cit.), Old Persian *garda*- 'Hausgesinde' (< *gr̥da*- 'Haus') is not a cognate of this Ossetic word. In the absence of a clear origin of the Ossetic word this cor-responds to explaining *obscurum per obscurius*.

1. Adams (and others) propose an adjectival formation based on Indo-European reconstructed forms for 'yard', assuming a development 'yard' > *'lordly yard' > *'belonging to lordly yard' > *'belonging to lordly yards' > 'palace'. This is perhaps Joki's interpretation as well, as he suggested a link between the Tocharian word and the Ossetic one meaning 'garden; enclosure, fence'. The original meaning of the Tocharian would thus be 'enclosure', evolving into 'yard', etc.

2. The second type, which is that of Isebaert, consists of the following semantic pathway: 'of the house' > *'servant' > pl. *'servants' > *'entourage, court' > 'palace'.

My own explanation follows a similar pathway to Isebaert's. In my view, the first series of proposals, concerning an Indo-European etymology for the Tocharian word, rely on a specific semantic evolution, from 'yard' to 'royal court' which is based on a Romance (specifically French) parallel. I wish to show that this parallel is less self-evident than it initially appears, and I thus need to detail the French etymology that, in my opinion, implicitly motivated the etymology of TB *kercci* from *$g^hord^hii̯o$- and related.

2.4.4.2 French *Cour* 'yard; court' and its bearing on the etymology

The first type of proposals were presumably influenced by the semantic development seen in French *cour* 'yard; royal court'; secondarily 'palace (as a building)' (when it designates the royal court, *cour* is usually written as *Cour*),[76] Spanish *corte* 'idem', Italian *corte* 'idem', German *Hof* 'idem', Dutch *hof* 'yard, royal court'. Then, by metonymy, at least in French, *cour* can designate the palace itself.

Incidentally, *cour* derives from the accusative *cōrtem* of Vulgar Latin *cōrs*, *cōrtis*, itself from Classical Latin *cohors*, *cohortis* and is thus directly connected to the PIE form *g^hortos mentioned above. This connection has perhaps further prompted the etymological proposals discussed here.

---

[76] The Old French word *cort* was borrowed in English as *court.* The English judiciary meaning also found in *court* derives naturally from the fact that judiciary processes originally occurred in the royal court. The modern French orthography *Cour* (instead of *Court*) derives from an etymological confusion with Lat. *curia*, French *curie* 'Roman senate; the assembly thereof' (see Ménage 1694: 227-28, for references, arguments, and a discussion).

French *cour*, Old French *cort*, *curt*, initially designated an enclosed space, with a general meaning "terrain découvert, souvent entouré de murs ou de bâtiments, devant ou derrière l'habitation principale" (von Wartburg FEW II: 849). Later, in medieval times (only attested since the 12th century, von Wartburg FEW II: 850),[77] the lord's enclosure (stronghold) offered protection against raids to the 'lower people' (peasants) in the *basse-court* and to 'higher people' in the *haute-court* or *cour d'honneur* (in Spanish *corte noble*). In Modern French *basse-court* came to designate the animals associated with the lowest caste (poultry, rabbits, pigs). Because justice was held in the *haute-court*, the terms *cour de justice* and *haute cour de justice* are still used in Modern French.

By metonymy, this term came to designate the members of this *haute-cour* and the castle or palace where the lord (and later the king) resided. It is rather clear that the identical meanings found in Germanic languages for words that originally meant 'yard, enclosure' are due to French influence. Such an influence on Germanic semantics is often attested, and seems quite likely.[78] If indeed, Low and High German acquired this meaning for *hof* under French influence (and most other Germanic languages from German, cf. Hellquist 1922: 244 for Swedish),[79] as did Russian,[80] then the frequency of this semantic change is even reduced.

The semantic evolution of this word from 'yard, enclosure' to 'royal court' (and then to palace) took place in a very specific context: medieval

---

[77] If this semantic shift already occurred in Vulgar Latin, one cannot eliminate the possibility of the influence of Greek αὐλή 'open court; courtyard'; post-classical Greek 'house of a lord; royal entourage' (Bailly 2000: 309). This is however not certain, given the relatively late attestation of this meaning in French and the other Romance languages (12th century in French, even later in Italian and Spanish). A further difficulty of this etymology is that Lat. *aula* only designates the lordly court, so that the confusion with 'courtyard' cannot have happened, except among Hellenists. It should further be added that the meaning 'house of a lord' of the Greek word was first used to designate Persian palaces, which were very different from Greek ones, and contained large courtyards and gardens (I thank Romain Garnier for informing me of this fact).

[78] Cf. „Hof", in: Pfeifer & al. (online, accessed on the 15th of September 2020).

[79] I thank Professor Martin Kümmel for providing me with this information.

[80] As Professor Lubotsky (p.c.) informs me, French influence is generally accepted to explain the fact that Russian *dvor* 'yard; court' shares those two meanings.

Western Europe, in a world of strongholds, lords, and courtesans, with rather rigid social divisions reflected in 'low yard' and 'high yard'; so the semantic trait 'lordly, royal' could emerge. It is hard to believe such a specific evolution could have happened in the Tocharian world, given its cultural, geographical and historical distance with the Latin-speaking medieval world.

This casts much doubt on the inherited etymologies of TB *kercci*. Instead, like Isebaert (1980) and Tremblay (2005) I believe this word to be an Iranian loanword in Tocharian, as I will explain below.

2.4.4.3 TB *kercci* from OSIr. **gr̥di̯a-*

Having disposed of the inherited etymologies, I adhere to Isebaert's proposal, with some new considerations and arguments presented here. We start with Old Iranian **gr̥di̯a-* 'servant' lit. 'that (one) of the house',[81] attested for example in Elamite ‹kur-ti-e-bar-šá› ← **gr̥di̯a-br̥za-* 'the exalted servant' (Tavernier 2007: 191-92) and Elamite ‹kur-ti-ia-ma› < **gr̥di̯a-u̯ant-* 'having servants' (Tavernier 2007: 192); also Babylonian *gardapata*, *gardapatu*, Elamite ‹kur-da-bat-ti-iš› 'majordomus, steward, house-holder' ← Old Persian **gr̥da-pati-* 'house master' (Tavernier 2007: 424).

The pathway 'of the house' > 'servant' is relatively straightforward, often attested; in Old Iranian itself, we have **(d)māna* 'mansion, house' → **māni̯a* 'of the house' > 'servant', borrowed with this latter meaning in Tocharian (s.v. *mañiye*).

The Iranian noun from which the adjective **gr̥di̯a-* 'servant' derives is probably **gr̥da-* '(noble, rich) house', which had a variant **garda-* in Old Iranian. These variants are due to generalizations of the zero-grade and of the *a*-grade, respectively.[82] The form **gr̥da-* is seen, among others, in Avestan *gərəδa-* '(daevic) cave, burrow of obnoxious creatures' and in the Armenian loanword *gerdastan* 'household, body of servants and captives' (cf.

---

[81] This is exactly parallel to French *domestique* 'servant', also originally a substantivized adjective *domestique* 'of the house', and also to Tocharian *mañiye* (s.v.), from an Old Steppe Iranian word that originally meant 'of the house'.

[82] Either of two related nouns with different accentuation: **gr̥dá-* ~ **gárda-* (cf. Hoffmann 1992: 840, 854[8] for a discussion of such variations in Iranian). I thank Nicholas Sims-Williams for providing me with this information and the reference.

Olsen 1999: 333 and 333[290]). From *garda- derives for example Middle Persian *gāl* coll. 'the gang, the villeins labouring on the estates of the kings, the satraps, the magnates, etc.; in war their military crew' (Nyberg 1974: 80).

As one can see in these examples, *garda-/gr̥da- already had from Old Iranian times, alongside the basic meaning of '(important) house', the metonymic meaning 'workers and people of that house, household'. It is more likely that *gr̥di̯a- ~ *gardii̯a- 'servants, people of the house' is derived from the basic meaning.

The notion of 'court people' or rather, of 'people attached to the court' appears in the Elamite form *kurtaš* discussed in detail in Henkelman (2018). He calls *kurtaš* "dependent workers" (2018: 224). As can be seen from the texts he analyzes, these *kurtaš* were foreigners, coming from Bactriana, Lycia, or even Sogdiana. They could be men, women or boys. This implies that the *kurtaš* were, so to say, there to stay (Henkelman 2018: 235). A possible difference between the status of *gr̥di̯a- and that of *māni̯a- could have been that the former were permanently included into a class or a group, as possibly were the *kurtaš* in the Achaemenid Empire (cf. Henkelman 2018: 239).

As can be seen from the semantic range of the Middle Persian, Elamite, Babylonian, Armenian, the type of house and the type of servants described by *gr̥da- etc. is of a high standing. Not every house had butlers, householders, bodies of servants (and captives!), etc. Clearly, the type of house and households we are talking of are those of noble people and of kings. A royal meaning is even found in one of the meanings of Middle Persian *gāl*: 'the villeins labouring on the estates of the kings'. In the Middle Persian form and in many others, as can be seen above, there is a plural or collective meaning 'group of servants'.

Thus the semantic trait 'lordly, royal' is already attached to the Iranian word, and does not need to be acquired in Tocharian, which is a strong point for Isebaert's etymology. Logically, it would have been a singular noun in Old Iranian still, and only made into a *plurale tantum* within Tocharian. The shift from 'court, servants' to 'palace', which may have been connected with its becoming a *plurale tantum*, would thus have been made within Tocharian.

If this word is indeed borrowed from Iranian, it needs to have been borrowed from Old Steppe Iranian because of its vocalism, i.e. TB *e* for OIr. *a*.

An Old Steppe Iranian *gr̥di̯a- 'servant' would have been borrowed in Proto-Tocharian as *kercye or *kercce 'servant' → nom.pl. *ker(c)ciyi obl.pl. *kerciyeṃ 'body of servants, entourage',[83] which became our *plurale tantum* TB kercci 'palace', obl. kercciyeṃ* (this last form proves that the nominative was originally *kercciyi).

The geminate -cc- can be explained in three ways: (a) it is due to the preceding r (cf. kertte ← OSIr. *karta-); (b) it is due to the simplification of the cluster *ciyi > *cyi > cci in the plural, with influence of the spelling of the nominative on that of the oblique – however, *-iyi most probably became -i, not -yi (cf. Peyrot 2021); (c) it is a regular Tocharian reflex, or correspondent, of Old Steppe Iranian *-Ti̯V-, as one can perhaps see in TB waipecce ← Old Steppe Iranian *(h)u̯ai-pa ϑi̯a-.

We cannot ignore an important problem: I have suggested elsewhere (s.v. melte; speltke; welke) that in Old Steppe Iranian, *rd had become *ld, while *rt had become *rd. Thus we would expect †kelcci. If we reject this sound law, the -lt- in melte, speltke and welke needs to be explained differently only in order to explain kercci.

I propose three solutions to account for the r of kercci: first of all it is possible that *-rdⁱ- (that is, *r + palatalized *d) had not, in Old Steppe Iranian and at the time of borrowing, become *ld. This is of course *ad hoc* in the absence of any other example, but a phonetic explanation is available:

If we consider that OSIr. */d/ was realized as a dental, then *rd > *ld can be explained as an assimilation (/l/ being generally more dental than /r/ which is generally apical, cf. Ladefoged & Maddieson (1996: 182f., 215f.) for a detailed discussion).[84] In a cluster *-rdⁱ- the *d would be palatalized and thus no longer dental.[85] Perhaps it was already closer to *[dᶻ] than to *[d̪]. Following this, *-D(i)yV- (where D notes any dental sound) would be

---

[83] Compare TB epastye 'skillful' → nom. pl. m. epasti < *epə́stiyi, for which see Peyrot (2021).

[84] This explains why, in some Iranian languages, both *rd (< PIr. *rd) and *rz (< PIr. *rȷ́) became l.

[85] I also believe that Tocharian l was alveolar or dental, because it was in opposition with palatalized l noted ‹ly›, and in this type of opposition the two elements need to be maximally differentiated in the place of realization. However, in Polish and Russian, the non-palatal l is velarized, which is also a possibility for Tocharian.

rendered by Proto-Tocharian *-c(i)yV-, as seen in *waipecce* 'good, belonging' < *h_uai-paϑia-.[86]

A better solution would be that the PIr. sequence *-d_iV- was already palatalized in Old Steppe Iranian (on this see the section 2.6.2.a). This solution differs in that it would suggest that this palatalization would be older than the sound law *rd > *ld of Old Steppe Iranian. This also implies that *rǰ would have been borrowed into Tocharian as *rcc and only *rd in a non-palatal context had become *ld in Old Steppe Iranian.

An even better solution, in my view, consists in viewing the sound change of *rd to *ld as more specifically *[rð] to *ld, then explaining the palatalization of *rdi̯ becomes easier. For more detail on this solution, see section 2.6.2.k.

In conclusion, the inherited etymologies lack semantic ground for the incorporation of the 'lordly, royal' trait in the etymon, while Isebaert's etymology accounts for it. The meaning 'palace' is most probably a Tocharian B innovation from 'royal household', as Tocharian A uses a periphrasis instead. Thus, on a semantic basis, I favour an Old Steppe Iranian etymology for this word, although it requires some phonetic explanations in order to explain TB *-rcc-* from OSIr. *-*rdi̯*- (instead of, for instance, *-*lcc*-). In general it can be said that if this phonetic problem could be solved, this etymology would yield a precious amount of information concerning Tocharian society and culture, as well as Tocharian prehistoric architecture.

2.4.5 TA *wankā-* 'to chat, gossip'

Tocharian A *wankā-* 'to chat, gossip', deverbal noun TA *waṅke* 'pleasant talk' (cf. Peyrot 2012: 212, with references), has no clear etymology. Isebaert (1980: 90-91) has proposed to derive it from Iranian. Indeed, as he notes, there is an Iranian form *u̯ānk(a)-* (or *u̯āng(a)-*) 'sound, cry' (Bailey writes "*vank-* (or *vang-*) 'make sounds'" DKS: 373) which can be reconstructed.

Khwarezmian has *wʾnk* 'Ruf' (Benzing 1983: 635); Middle Persian has *wāng* 'voice, call, cry, noise' (DMMPP: 335), borrowed into Armenian *vang* 'Laut, Ton, Silbe' (cf. Hübschmann 1897: 243); New Persian *bāng* 'loud shout, outcry', Gilaki *väng* 'cries, lamentation' and Balochi *gwānk* 'cry'

---

[86] Palatalization in front of Old Iranian *-*ii̯a*- is also seen in Tocharian B *mañiye* ← OIr. *mānii̯a*.

(Korn 2005: 99); Zaza *vang*, *van*, Bakhtiari *bang* 'shout, shriek, hail', and the list can be much longer.[87]

This root in Iranian is strictly nominal, and I was unable to find any primary verbal form relating to it. Bailey (DKS, s.v.) saw it in a few Khotanese verbs, but better etymologies have been found: *vaj-* 'to dispute' has been tentatively derived from **au̯a-Hi̯au̯d-* (Emmerick apud Degener 1989: 101);[88] *pyūṃj-* 'to deny' is convincingly explained by Federico Dragoni (p.c.) as deriving from the root **u̯a(n)č-* 'to stagger, shake' ('to shake away' > 'to deny'). As to Balochi *gwanj-* 'to cry', it is convincingly explained by Korn (2005: 137; already in essence in Gershevitch 1971: 284) as a back-formation based on the past stem of *gwānk ǰan-* 'to cry' (lit. 'to strike a cry'), *gwānk ǰat*, reinterpreted as a simplex *gwanǰat*.

It has been proposed that this **u̯ānk(a)-* derives from **u̯ač-* 'to speak', with a "nasal increment", or a nasal infix (cf. Gershevitch 1971: 279f.). Alternatively, it has been suggested that it derives from Old Iranian **u̯āna-ka-*, a cognate of the Vedic *vāṇá-* 'voice, music' (cf. Hasandoust 2014: 398f., with references). The first hypothesis is a bit difficult because of the *-k-* consonant, the second one because we would expect **u̯ānaka-* as a proto-form (yielding, for instance, Balochi **gwānak*, etc.), and also because a retroflex *ṇ* in Sanskrit does not regularly correspond to **n* in Iranian.

In both cases we could imagine the influence of the mimophone "*bang*" (English *bang*, French *pan*). The first part of the word could be connected to Vedic *vāṇá-* while the element *-ng* could be due to the influence of the mimophone. A slight problem is that the mimophone usually starts with a *b-* or *p-*, because these sounds imitate an initial explosion (that of the original sound). Perhaps an even better solution would be to see it as a cross-form between the root **u̯ak-/*u̯ač-* 'to speak' and this same mimophone *bang*. This hypothesis also helps explain why this root is strictly nominal in Iranian.

---

[87] In Bakhtiari, there occurred a shortening of the ancient **ā* before *-ng*, cf. Vahman & Asatrian (1987: 71).

[88] One often finds the spelling *vaṃj-* but, according to Emmerick's etymology, it might be better interpreted as *vaj-*, with unetymological *anusvāra*, common in Late Khotanese, and the spelling *vaj-* is more precise (I thank Federico Dragoni for informing me of this).

This discussion serves to explain the appearance of this purely Iranian nominal root, which is unusual from an Indo-European point of view (there is nothing in Indo-European that could yield a *-ng- or *-nk- sequence in Proto-Iranian).[89] If Tocharian was also derived from the mimophone "*bang*", it would rather yield something such as †*paṅkā- vel sim*. Most probably the Tocharian word thus derives from Iranian, as already suggested by Isebaert (1980: 90-91). The semantic pathway is very straightforward 'noise' > *'rumor, unimportant speech' > 'chatter, chitchat, chat'. It is a quite commonly attested semantic change: for instance French *bruit* 'noise; gossip'. A parallel semantic change is probably behind the Sogdian form *wnxrš* 'rumour, news'.[90]

The question that now concerns Tocharian is: what is the exact source, and was the verb borrowed first, or the noun? One might think it is a Pre-Bactrian or Bactrian borrowing, *οαγγο → *waṅk. From this noun a verb *wankā-* would have been made. The Bactrian form itself is unattested, but it is very likely to have existed, as this root is (almost) pervasive throughout Iranian.

There is a possible problem, since the first vowel would be expected to have been long, and thus probably yield †*wāṅke*. The verb can derive from either a long or a short vowel in the first syllable, but no Bactrian word was recorded as being borrowed as a verb in Tocharian to this day. If it is not a borrowing from an unattested Tocharian B word, the Tocharian A form rather constitutes an abstract derived from the verb (cf. Peyrot 2012: 211f.).

Another possibility is that the Tocharian A words are of an older date, and derive from Old Steppe Iranian. In this case, it would be a Proto-Tocharian borrowing *wanke* 'gossip' ← Old Steppe Iranian *u̯ānka-, or *u̯ānaka- (or *u̯ānga-) 'noise; gossip' although this latter etymon would normally yield TA *wāṅk. A verb would then easily have been derived therefrom, according to the same model as the other Tocharian verbs deriving from Old Steppe Iranian nouns (for instance s.v. *raitwe*). Then, to ex-

---

[89] Schwartz (1969: 446) saw such a root reflected in Sogdian *wnγr* /wanxr/, comparing also OInd. *vaṅkú-* (as 'noisy') and mentioning Tocharian A *waṅke*. Prof. Schwartz now tells me that the OInd. word should be removed from consideration since it is semantically uncertain, and since the root is limited to Iranian, he endorses an Iranian origin for the Tocharian A word.

[90] I thank Prof. Nicholas Sims-Williams for informing me of this example.

plain the vocalism of the noun TA *waṅke*, it would have been an inner To-charian A creation, a secondary abstract in *-e*, as suggested by Peyrot (2012: 211f.). Although it is difficult to ascertain this, it is in my view a more likely solution. According to this hypothesis, Tocharian B lost this lexeme, and replaced it with *käskor** at least for the noun.[91] Here again, one can evoke Professor Schwartz's solution of a proto-form **u̯ank-* (with only secondary lengthening), to explain the vocalism of the Tocharian A form, since Old Steppe Iranian **u̯ank-* would regularly yield Tocharian A *waṅk-*. However, this would imply that the word was borrowed directly as a verb, for which we have no parallel.[92]

To sum up, a possible trajectory for the Tocharian A word *wankā-* 'to chat, gossip', without going through a hypothetical Tocharian B loanword, is as follows: Old Steppe Iranian **u̯ān(a)ka-* 'noise; gossip' would have been borrowed as a Proto-Tocharian noun **wanke* 'gossip', from which a verb PT **wanka-* (prs. *wonko-*) was made, and later a Tocharian A deverbal noun *waṅke* was derived from the verb.

In conclusion, this discussion involves multiple assumptions and hypotheses, since we cannot be entirely sure of the etymology of Tocharian A *wankā-* 'to chat, gossip'. However, I consider an Iranian origin of this verb very likely, as no other origin can easily be put forward and we have a suitable Iranian source-word.

---

[91] Alternatively, it is possible that TB **wanka-* belonged to a much lower sociolect and was not represented in the texts, or perhaps even that it is simply not yet attested in our extant corpus.

[92] Professor Martin Schwartz has suggested another solution to me, which I cite verbatim: "I suggest that an Iranian root **u̯ank-* 'to make a sound', reflected in Sogdian /wanxr/ 'voice' < **u̯ank-ra-*, gave the noun **u̯ankV-* in Old Steppe Iranian, which is reflected in Tocharian A *waṅke*, etc. The early date of this event left the short vocalism of the first syllable (cf. the etymologically unrelated stem Avestan *xᵛanaṯ-* 'making a noise' in compounds) unaffected by the later development whereby **hu̯ānai̯a-* furnished the verb stem for 'to call' etc. (> Manichaean Middle Persian *xwān-*, Khotanese Saka *hvāñ-*, Khwarezmian *m|ʾfxwʾny-*, Bactrian χοαν-, χοην- and Ossetic *xon-*).

I suspect this affection occurred via phrasal collocation of **hu̯ānai̯a-* with **u̯angV-*, giving **u̯āngV-* reflecting in Middle Persian *wʾng*, Khwarezmian *wʾnk*, Balochi *gwānk*, etc. It seems that **hu̯ana-* survived in Parthian *xwn-*, Sogdian *xwn-* and Bactrian χοαν-."

2.4.6 TA *nātäk* 'lord', fem. *nāśi* 'lady'

The Tocharian A word *nātäk* 'lord' (nom.pl. *nācki*), corresponding semantically to Tocharian B *saswe*, and its feminine counterpart *nāśi* 'lady' have previously been seen as cognates of Greek ἄναξ, -κτος, 'lord, master' (cf. Mycenaean *wa-na-ka*, Beotian Ϝάναξ, etc) and ἄνασσα, -ης 'lady, queen' (Winter 1970: 53; Adams 2017a:1376).[93] This connection is rightly rejected by Del Tomba (2020: 79), who notes that the initial *$\underline{u}$- seen in the Greek forms should not have been lost in Tocharian, and that Greek has *-kt-* while Tocharian shows *-tk-*. The etymology of the Greek form is difficult (cf. Beekes 2010: 98).

Del Tomba concludes that "TchA *nātäk* cannot be derived from any internal source" and alternatively proposes it to be a borrowing from Skt. *nāthá-* m. 'protector, possessor, lord' (2020: 79). As Del Tomba (loc. cit.) notes, this noun is frequently attested in epithetic compounds such as *nāka-nātha*, *nāka-nāthaka-* 'sky-lord' (epithet of Indra),[94] *loka-nātha* 'saviour of the world (epithet of the Buddha)', etc.

Although very attractive from the semantic perspective, Del Tomba's Sanskrit etymology of TA *nātäk* 'lord' is difficult on the formal level. From Skt. *nāthaka-*, one would rather expect TA †*nātak* (type Sanskrit *kacchapa-* 'turtle' → TA *kāccap*, cf. Schwarz 1974: 406). Another problem, in my opinion, is the archaic nature of the morphology of Tocharian A *nātäk*: the plural *nācki* and the feminine *nāśi* 'lady'. The form *nātäk* seems to reflect a Proto-Tocharian syncope (PT **nateke* > **natke*, of the type of OIr. **rataka-* > PT **retke*, TB *retke*, TA *ratäk* 'army'), and it cannot be inherited (cf. Del Tomba 2020: 79f). Those two facts brought together strongly suggest a very early date for the borrowing of this word, namely, that it is a Proto-Tocharian borrowing, despite the fact that no Tocharian B cognate has yet been found for this word.

The vowel *ä* in *nātäk* could hardly be due to an early Tocharian A sound change. For older loanwords and inherited words, there existed a rule of vowel weakening: in the second syllable, Pre-TA **ā* was reduced to **a* if the

---

[93] To note, Winter (1970: 53) suggests that both the Greek and the Tocharian words are of (shared) non-Indo-European origin.

[94] I wonder if the translation 'supporter of the sky' (as in, the one who keeps the sky from falling) would not be more appropriate.

first syllable contained a vowel *a or *ā or a diphthong (cf. Kim 2007, with references). However, this rule is not found in Sanskrit loanwords (e.g. *āsāṃ* 'throne') nor in late Iranian loanwords. The word *kātäk* 'householder' is sometimes found in the literature as a counterpart of TB *kattāke* 'id.' (e.g. Pinault 2008: 234), but the form *kātäk* is not found in any text with the meaning 'householder', and as the counterpart of TB *kattāke* one should rather set up *kātak** (cf. Carling 2009: 110; Del Tomba 2020: 80).[95] There is thus no clear example of a secondary *ä* < *a* in the (TA) second syllable in any trisyllabic Indic loanword.

No borrowings from Sanskrit in Tocharian that predate the separation of Tocharian A and B are known, that is, Sanskrit borrowings for which we have a regular correspondence between A and B, as we find for Old Iranian words. Although in this case there is no known Tocharian B correspondent, the Tocharian A word is still much more archaic in its derivation than any Sanskrit loanword in Tocharian. I cannot accept Del Tomba's assumption (2020: 80) that the word was borrowed into Proto-Tocharian from Middle Indic either. Not only would the vocalism not be solved by that assumption, but it would also suggest a rather complicated inflection and derivation for a word taken from a Middle Indic language, for which there are no parallels.

Because it would explain the problems mentioned above better, in particular the syncope or the vowel-weakening, I prefer to see this word as an Old Steppe Iranian loanword in Proto-Tocharian. The major obstacle to this etymology lies in the fact that there is no word *nāθaka- attested in Old Iranian. This is, however, only a superficial problem, as I hope to demonstrate.

First, the basic meaning of Ved. *nāthá-* is 'refuge, help', cf. for example *a-nāthá-* n. 'Schutzlosigkeit'. This word has no clear etymology: a root *h_3neh_2- 'to help', suffixed with *-tHo- thus yielding the meaning 'helping' cannot yield the correct meaning, since no root with such a meaning is attested, cf. Frisk (1960: 395f.; LIV²: 302; *pace* EWAia²: 33f. and Beekes 2010:1083f.).

It has been suggested by Kroonen (2013: 388) that Vedic *nāthá-* is connected to Proto-Germanic *nēþō f. 'mercy, safety'. He thus projects *nāthá-* 'refuge, help' back as PIE *$nēt-h_2-o$-, and further connects the Indic and

---

[95] The form *kātäk* found in Tocharian A texts can rather be interpreted as the 3sg preterite active of *kātkā-* 'to arise'.

Germanic forms to a Proto-Indo-European root *net- 'to protect, favor', cf. Proto-Germanic *neþan- s.v. 'to grant safety (?)'. In any case, Kroonen's reconstruction (*nēt-h₂-o-) is preferable to the earlier postulation of a root *h₃neh₂- 'to help', for the reason evoked above (simply that such a meaning is not found, or not primary in the Greek forms such as ὀνίνημι, primarily meaning 'to be of use').

In Indic, the meaning 'saviour', 'lord', 'protector' might have been back-formed, or semantically reinterpreted from the epithetic compounds mentioned above. This well-known process can be described as follows: loka-nātha- initially meant 'refuge of the world', but since it was attributed to an individual or a god, it was taken to mean 'protector of the world', and nātha- was thus reinterpreted as a word meaning 'protector'. Another possibility is that a neuter *nātha- 'protection' was derived in *nāthaka- 'protector' and the meaning 'lord' of nātha- m. is secondary, as I explain below.

Once reinterpreted as a noun 'protector, lord' or 'saviour', nātha- could easily be enlarged with -ka-, as shown in the examples above (for more examples, see Del Tomba 2020: 80). For the Iranian part, it would be easy to simply suggest that *nāθa- 'protection' was enlarged with the denominal suffix *-ka- and thus took on the meaning of 'protector, lord'. The semantic shift from protector to lord is not difficult, especially in ancient times when lords were, first and foremost, protectors, or supposed to be.

The Indo-Iranian noun *nāθa- is not attested in Old or Middle Iranian, but seems to subsist in at least one word: New Persian panāh 'protection, refuge' < *pad-nāh < *pati-nāθa- (Hübschmann 1895: 43).[96] This etymology of Hübschmann's has long been neglected, with many preferring to see this word as a continuation of the Persic cognate of Avestan nas- 'hingelangen zu' (so Nyberg 1974: 150). In my view, this second explanation fits neither the form nor the meaning of New Persian panāh: indeed, 'to go somewhere, to return somewhere' seems rather far removed from the notion of 'protection'. Rather, Hübschmann's etymology should be retained as I will explain below.

I would also like to adduce another Persian word which, in my opinion, goes back to OIr. *nāθa-. It is MP nihān, NP nihān ~ nahān noun, 'secret,

---

[96] Hübschmann (loc. cit.) was still unsure whether *patn- could yield pan-, but I believe this is now firmly established (cf. e.g. panhān < *pad-nahān, discussed here as well).

hiding' adj. 'secret, hidden', *panhān* < **pad-nahān* adj. 'secret, hidden'. This word is usually explained as deriving from **ni-δāna-* 'set down' (from **dā-* 'to set') (so Hasandoust 2014: 2814). This cannot be correct, as -δ- does not become -h- but -y- in Middle and New Persian. Furthermore, PIIr. **ni-dʰāna-* seemingly means 'treasure' (lit. 'what has been deposited'), as one can deduce from the comparison between Sanskrit *nidhāna-* 'treasure' (M-W: 548) and Middle Persian *niyān*, also 'treasure' (DMMPP: 252).

In my view, *nihān* derives from **nāθa-* 'protection, refuge' suffixed with *-ān*. In the present case, this would not be the verbal suffix *-ān* (deriving from the athematic present middle participle **-āna-*, cf. Bartholomae 1895-1904: 109f. and Cheung 2015: 273) but the "other" suffix *-ān* (synchronically different: while it originally goes back to the middle participle in **-āna-*, it is mostly nominal and often non-analyzable already in the Middle Persian period), of the type of Middle Persian *wiyābān* 'astray, wandering' (cf. Cheung 2015: 273), and Parthian *žafrān* 'depth' (from *žafr* 'deep'). With the addition of this suffix, the word took a progressive, continuous meaning, so that we can propose the following semantic shift: **'protected' > **'covered' > 'hidden, secret'.[97] As to the form, we must first posit reduction of the **ā* of the first syllable, which is common before **h* (see Pahlavi ‹šh› *šah* 'king', cf. also Lazard 1963: 182), but could perhaps also be explained as a dissimilation (**nāhān → nahān*).

The *-i-* vocalism of Middle Persian *nihān* is, I contend, due to the influence of the verb *nihuft-/nihumb-* 'to wear, to hide'. Both words have strongly influenced each other. As is known, *nihuft-* derives from **ni-gufta-*, but **-g-* does not regularly become -h-. In the present case, it has been proposed that the -h- of *nihuft-/nihumb-* is due to the contamination of *nihān* (cf. Hasandoust 2014: 2814). I thus also propose that the *-i-* vowel of *nihān* is due to contamination with *nihuft-/nihumb-* (and the *-a-* in NP *nahān* is second-

---

[97] It is not even necessary to posit a verbal form **nāθ-* 'to protect', since the suffix *-ān* could simply be added to adjectival or nominal forms, once most of the inherited words containing this suffix were no longer perceived as verbal or rather deverbal forms.

ary, cf. Pisowicz 1985: 15). There probably was also interaction between the meanings of both *nahān* and *nihuft-*.[98]

Now that I have given my arguments for the existence of an Iranian word *nāθa-* 'protection, refuge',[99] it is not difficult to propose an Old Iranian *nāθaka-* 'protector, lord', along with *nāθa-* 'protector', either back-formed from the same type of compound, as in Sanskrit, or rather from *nāθaka-* 'protector, lord' itself. The semantic shift from 'protection, refuge' to 'protector, lord' is facilitated by the very common phraseology of the type "you are (like) a refuge", French *tu es mon refuge*, or *tu es ma protection*, usually said in a romantic context, but not far removed from a literal use (that is, when said to the lord, who literally is meant to be the protection of his people). The Old Steppe Iranian word *nāθaka-* 'protector, lord' was borrowed in Proto-Tocharian as *natke*, which regularly yields Tocharian A *nātäk* (and should have yielded TB †*nātke*).

The derivation of TA *nāśi* 'lady' is somewhat more complicated. As Del Tomba (2020: 80) notes, palatalization of the cluster -*tk*- yields -*ck*- and not -*ś*-. This is obvious from the palatalized nom. plural of *nātäk*, *nācki*. "If PTch *-*tk*- always palatalised as -*ck*-, then TchA *nāśi* cannot derive from TchA *nātäk* directly." (id.). Del Tomba tries to further explain this palatalization in two ways: firstly, through the addition of the palatalizing feminine suffix *-ʸa*: "*natakʸœ > *nataśʸa* (palatalisation) > *natəkśʸa* > Pre-TchA *nātśi* > *nāśi* (assimilation and simplification)", and second, "one may think that PTch *-y-* palatalised the cluster *-tk-* differently, yielding Pre-TchA *-śś-*:

---

[98] One could also think that *nāhān* directly became *nihān* under the influence of *nihuft-*. In that case the -*a*- vowel of *nahān* is simply due to assimilation of the second -*ā*-, as occurred in Iranian Persian in other words.

[99] Federico Dragoni has suggested a few words to me whose semantics are attractively close to that of the concerned root. These are Khot. *ānatu* used together with the verb *yan-* ('to do') in the collocation *ānatu yan-* 'to take care of, protect', and with *häm-* in the meaning 'to be taken care of, to be protected' (cf. also DKS: 18); Christian Sogdian ʾ*ntqy*ʾ 'modesty', and some other cognates. Although the forms look strikingly similar, with a semantic derivation that is far from absurd ('protecting' > *'hiding' > 'modest'), I do not know why they present a -*t*- and do not reflect a *-θ-*. Perhaps one can think of a PIE *nēteh₂-* 'protection' → *neth₂-o-* 'relative to protection, protecting, refuge' to explain PIIr. *nāθa-* and *ā-nat-* (< PIE *h₁ed-net-*) 'conferring protection' to explain PIr. *ānata-*.

*natkya > *naśśi > TchA *nāśi* […]" (Del Tomba 2020: 80). In my view, it is clear that *nāśi* is related to *nātäk* with the addition of the *-ya- suffix of the *aśiya*-type, although the matter is intricate.

The first proposal is difficult to accept because the syncope of *nateke* happened very early: this type of syncope (see section 2.6.2.g) occurred either in Old Steppe Iranian or in Proto-Tocharian already. In that sense, I am unsure whether we can even reconstruct a *nateke* stage. Most of all, I do not see how there could have been both *k* and *ś* in the sequence *-təkšʸ-*. In any case, there is no parallel for a simplification or palatalization of such a sequence to *-ś-*. The second proposal seems less complicated. Of course, the palatalization *natkya > *naśśi is unprecedented and *ad hoc*, but it is the best solution at hand, it seems.

In conclusion, I have tried to show the existence of Proto-Iranian *nāθa- 'protection, refuge', a cognate of Vedic *nāthá-* 'idem'. I have also argued that TA *nātäk* 'lord' ultimately derives from Old Steppe Iranian *nāθaka- or *nātaka- (depending on whether there existed a phoneme /θ/ in Old Steppe Iranian, see section 2.6.2.a) meaning 'lord' or 'protector'. As to the exact form of TA *nāśi* 'lady', there is no entirely convincing explanation, although Del Tomba (2020: 80) has advanced two hypotheses, one of which seems more convincing. In any case, the problems of the derivation of TA *nāśi* 'lady' exclude a borrowing from Sanskrit.

2.4.7 TB *tˢəw-* 'attach oneself to, stick to', TA *tˢäw-* 'to put together'

Before discussing the etymology of Tocharian B *tˢəw-* 'add' and Tocharian A *tˢäw-* 'to put together', I wish to present and discuss the data at hand. Peyrot (2013: 846) sets up the following verbs:

TB *tˢəwa-* 'attach oneself to, stick to' TA *tˢäwā-* 'fit (intransitive, active); obey (middle)'
TB *tˢəw-* causative 'add (transitive)' TA *tˢäw-* causative 'put together'.

Peyrot reconstructs for both languages a Proto-Tocharian form *tˢəwa-*. Adams (DTB: 808), giving the Tocharian B form as *tsu-*, translates it as 'cohere, adhere; embrace, contain' (active); 'adhere, stick, cling, attach oneself' (middle) and causative 'make cohere, add to'. Adams further writes that the etymology of this verb is uncertain. He mentions a connection proposed by

Seržant (2007) of this verb with Greek δύναμαι 'to be able' and Gothic *taujan* 'to make'.

Greek δύναμαι means 'to have the capacity of doing something potentially' (like 'I can swim'). Thus, the meaning of the Greek verb is far off that of the Tocharian one. In the Addenda and Corrigenda to the LIV² (available online)[100] Kümmel posits a PIE root *$deh_1\mu$- 'zusammenfügen, ordnen' with only the Tocharian verbal forms.

Apart from the semantic weakness of these connections, we would in Tocharian expect *śew- from *$deh_1\mu$- (> *tsēw > tsyew- > *śew-), or perhaps *śaw-. This is not a major problem, since the initial could have been generalized from a different ablaut grade, for instance from the zero grade. The lack of cognates is more problematic for the Tocharian etymology.

In my view, an Iranian etymology is also possible. If we start from the causative meaning 'to add to, to mix', and see the meaning 'adhere, stick to' as a back-formed decausative, we have a good Iranian candidate: Proto-Iranian *$ja\mu$- 'to pour a libation', seen in Avestan *zaoϑra*- 'libation', Buddhist Sogdian *zwtꞌk* and Christian Sogdian *zwty* 'beer, liquor' (according to Gauthiot 1913: 102 it could mean 'wine',), Yazgholami *zaw-/zod* 'to tread down, compress, squeeze', etc. (cf. EDIV: 471f.).

It is the meaning 'compress, squeeze', which I find quite strikingly close to the Tocharian meaning of 'put together, make adhere, fit'.[101] In my view, Buddhist Sogdian *zwtꞌk* 'wine' and Christian Sogdian *zwty* 'alcoholic liquor' (cf. Sims-Williams 2014: 102) originally comes from *zuta-ka*- 'what has been squeezed, treaded down', and refers to the process of squeezing fruit for the fabrication of liquor, or of treading down the grapes to make wine.[102]

From 'to squeeze fruit' to 'to make adhere, add', the semantic pathway is not evident. However, in a number of languages the root *$ja\mu$- gave a word

---

[100] https://www.academia.edu/402269/Addenda_und_Corrigenda_zu_LIV_ (accessed on the 20th of May 20, 2021).

[101] The word for 'pus' is also related to this verb 'to pour' in Ormuri *zū*, Waziri *zawa* 'pus, matter' (cf. Morgenstierne 1932) and in Khotanese *ysūa*- (gen. *ysūna*, from *ysun*- 'to pour, strain', cf. DKS: 353. Pus needs to be squeezed and compressed to be extracted.

[102] Henning (1946: 720) wonders if there is a connection of the Sogdian word to Greek ζῦτος, ζῦθος, an Egyptian type of beer (Lat. *zythum*), which, he writes, has an unclear etymology. He proposes that both the Greek and the Sogdian words have a common origin, "possibly in some Scythian language".

for 'broth': Khotanese *ysūma-* 'broth', Waziri *zēmna* 'soup' (DKS: 353). Interestingly, Tocharian B *sumo* 'libation, oblation' was probably borrowed from Khotanese *ysūma-* 'broth' (cf. Dragoni 2022: 217). Broth is constituted of hot water to which bones, meat or vegetables are added, or put together. If, again, we depart from the meaning 'to put together, to add' that the Tocharian A and B have in their causative, there is a much stronger semantic connection to be made between the Iranian forms and the Tocharian.

If this is correct, although the data at hand seems a bit unsure, and the semantic details still need to be made clear, it is possible that the Proto-Tocharian verb **tˢəwa-* 'to add, to put together (?)' derives from an Old Steppe Iranian verb **ǰu-* 'to add to a liquid, to compress together'. The semantic divergence from the original meaning of 'to pour (a libation)' can perhaps be explained through a preverb which was later removed, for instance **ni-ǰu-* 'to pour down (as a recipe)'. It is perhaps more likely, as no direct verbal borrowing from Old Steppe Iranian has been noted, that there was an Old Steppe Iranian noun **ǰu-ta-* 'mixture, broth' from which the verb would have been extracted within Tocharian, or perhaps, which would have influenced the Tocharian (and the Old Steppe Iranian meaning). It is also possible that the verb is inherited from Indo-European, according to the model I suggested above.

To conclude, the Proto-Tocharian verb **tˢəwa-* and its Tocharian A and B descendants TB *tˢəwa-* 'attach oneself to, stick to' TA *tˢäwā-* 'fit (intransitive, active); obey (middle)' could be inherited, deriving from a zero-grade **dh₁u-* (**deh₁-* 'to bind' with a present stem in *-u-*) becoming **duh₁-* with regular metathesis and resulting in Proto-Tocharian **tˢəwa-*. It could also be derived from an Old Steppe Iranian form **ǰu-*.

2.4.8 TB *tseriteke* 'young grass (?), green animal (?), turtle shell (?)'

As pointed out to me by Federico Dragoni, Tocharian B *tseriteke* has a very Old Steppe Iranian appearance, despite some phonetic problems (see p. 168). Its meaning is unclear. Below I cite one of the sentences in which it appears:

THT 324 a3

> *(me)nāktse ṣamāne : tseriteke menākäccepi ///*
> "a monk comparable with … to the comparable […]" (Ogihara 2009: 406).

Based on this context, the meaning of *tseriteke* cannot be established. However, if one takes this word to be of Old Steppe Iranian origin, it is tempting to see in it the rendering of **ɖaritaka-*, a derivative of **ɖarita-* 'green'. Ossetic *zældæ*, the descendant of Old Iranian **ɖarita-*, means 'young grass, grass' (cf. Cheung 2002: 253-54). If this meaning is very provisionally accepted for Tocharian B *tseriteke*, one can perhaps restore it elsewhere:

THT 3279 a6

/// – r kaccāp tseri(tekene) ///

The Tocharian B word *kaccāp* meaning 'turtle, tortoise', one could venture to think that a tortoise would naturally be in grass. Then perhaps one could read the first occurrence as meaning "a monk comparable to young grass", meaning "a young monk". Indeed, the metaphor of green grass, used to designate young, inexperienced people is quite common cross-linguistically; cf. French *un vert*, Dutch *groen als gras zijn* 'to be green as grass', *een groentje* 'a novice, uninitiated'. Of course, in the lack of a clear context or of bilingual evidence, this remains provisional. One could perhaps think as well of a turtle or a tortoise in a shell, as turtle and tortoise shells are also green. The translation is based completely on the assumed etymology and cannot be confirmed by textual evidence. Furthermore, it remains problematic that Old Steppe Iranian **i* would be reflected as Tocharian B *i* in this position (see section 2.6.2.f).

There is another occurrence of *tseriteke*, which I cite below:

PK AS 15D b4

(we)śeñaṃ su mā kca ṣäṃṣāte tserītekets wiyälyñe ramt ///

"Il ne tint pas compte des voix [des démons] … comme un effroi de *tseriteke*" (Athanaric Huard, p.c.)

As Athanaric Huard (p.c.) tells me, since we do not know if *tserītekets wiyälyñe* refers to the previous part of the sentence or not, this example is not very helpful. Nonetheless, from the context it seems unlikely that it re-

fers to 'grass', as 'grass' can hardly be afraid. It could possibly refer to a green or yellow animal.[103]

In conclusion, *tseriteke* could be a word of Old Steppe Iranian origin meaning '(young) grass', or 'turtle shell', or designating a yellow or green animal, depending on how we interpret the scarce evidence we have at hand. The meaning '(young) grass' would fit with the Ossetic meaning of *zældæ* 'young grass, grass'. Whatever its meaning, the appearance of TB *tseriteke* makes it very likely that it is an Old Steppe Iranian word. If the meaning 'grass' is wrong, another meaning such as 'green' or 'yellow (thing)' could also be possible for this word.

2.5 Old Steppe Iranian borrowings: rejected cases

In this section, I will discuss words which have been proposed to be or could perhaps considered to be of Old Steppe Iranian origin, but should in my view be derived from another source or cannot at present be plausibly attributed to Old Steppe Iranian.

2.5.1 TB *atiya**, TA *āti* 'grass; straw'

Tocharian B *atiya** can mean either 'grass' or 'straw' (in THT 194 b1 a fire of *atiya** is mentioned, which can hardly imply green grass). In a bilingual text, THT 530 b3, *atiya** corresponds to Sanskrit *tṛṇa-*. The word *tṛṇa-* itself has a wide range of meanings, such as 'grass; herb; any graminous plant; straw' (M-W: 453) and this wide range of meanings could very well be reflected in Tocharian texts, with calqued meanings as is not unusual. The precise meaning of Tocharian A *āti* is not entirely clear either, and it is usually taken to be the equivalent of TB *atiya** and translated accordingly.

Hilmarsson (1996: 51, with references) viewed these two words as cognates of Latin *ador* 'spelt, emmer wheat' and as deriving from PIE *h_2ed- 'to be dry', enlarged within Tocharian with the suffix *-(i)ya*. However, the type of grass referred to by these Tocharian words is not necessarily dry, and an Iranian derivation of this word might perhaps be considered.

---

[103] Athanaric Huard (p.c.) informs me that this word could correspond to Tocharian A *tsuri*, found in a A437 b6 together with other animal names. However, the vocalism of the two forms is difficult to reconcile.

Tocharian *atiya could in theory derive from an Old Steppe Iranian form *ādíi̯ā-. However, evidence within Iranian for such a form is extremely slim, as it is based on Scythian αδιγόρ 'locust'. This word has been analysed as consisting of *ādī- 'grass' and *xʷara- 'eater', a locust being a "grass-eater" (cf. DTB: 9). However, this comparison remains extemely weak in view of the uncertain analysis of this "Scythian" word and the apparently very limited distribution of it in Iranian.

One might alternatively think that *ādī- is related to PIr. *ādu- 'cereal', on which a lot has been written (see for instance Rossi 2010), but little consensus has been reached. PIr. *ādu- is based mainly on Sogdian ʾʾδwkh 'grain, cereal'.[104] It would take too long to enter on a discussion of these words here (cf. most recently Kölligan 2020: 227f.), but it is unclear how a stem *ādī- could have been formed to *ādu-. One could perhaps imagine that an Old Steppe Iranian *ādu- was borrowed into Proto-Tocharian as *at, which was later suffixed with *-iya. One can wonder why there was no assibilation, although epastye could provide a parallel. This etymological option, too, remains too hypothetical to be acceptable.

In conclusion, the etymology of Tocharian B atiya*, Tocharian A āti remains difficult. A derivation from an Old Steppe Iranian form *ādíi̯ā- would work formally, but such a word cannot be safely set up on the sole basis of Sogdian ʾʾδwkh 'grain, cereal' and Scythian αδι* in αδιγόρ 'locust'.

2.5.2 TB, TA āp 'water; river; flood'

Tocharian A and B āp mean 'water' but also 'river' and 'flood', thus referring to water in movement rather than stagnant water. Theoretically, these Tocharian A and B words could either be inherited from Proto-Indo-European *h₂ep- or borrowed from Old Steppe Iranian *āp- 'water' (cf. Avestan ap-, nom.sg. āfš 'water', Sogdian ʾʾp, Persian āb, etc.). However, it would be remarkable if such an essential word would be a borrowing. At the same time, if it is a borrowing, the most likely source would be Old Steppe Iranian, as no other language was the source of the borrowing of such fundamental vocabulary in Tocharian.

---

[104] The Scythian form αδι-γόρ has, to my knowledge, never been brought in the discussion of these Iranian words.

The plural of the Tocharian B form could be telling. In IOL Toch 23 a5, we could read either *āpäṃ* or *āpäṃ(ta)*, since the fragment is torn off afterwards:[105]

IOL Toch 23 a5

y(ku)wermeṃ orotstsana āpäṃ ///

"Having gone, the great rivers…" (based on CEToM).

If the correct reading is *āpäṃ*, this would be a good argument for inheritance rather than borrowing, as this type of plural is rare, unproductive and archaic, but if it is *āpäṃta*, then it could be either inherited or borrowed, as *-nta* plurals were common and productive at the Proto-Tocharian period (examples of the *-nta* plural found in Old Steppe Iranian loanwords are TB *waipeccenta* 'possessions', *pakenta* 'portions, shares').

There are three arguments against a reading *āpäṃta* in IOL Toch 23 a5: first, we would expect the stress to go to the right, and we would thus expect *apaṃta**; second, the spelling with ‹ṃ› is unexpected for this ending, which is usually spelled ‹nta›; third, the word is feminine, and feminine nouns normally do not form *-nta* plurals. In sum, a reading *āpäṃ* is more likely, and this in turn points to inheritance.

The semantics are not in favor of an Iranian borrowing either: in Iranian, **āp-* means 'water', and does not necessarily refer to rivers. Further, it is such an essential element of the vocabulary that it seems difficult to accept the borrowing hypothesis. If it meant 'river' in Iranian, a loanword hypothesis would be more plausible, as examples are more common, for instance English *river* ← Old French *riviere*.

In conclusion, although Tocharian A and B *āp* 'water; river; flood' could theoretically derive from an Old Steppe Iranian form, it could formally very well derive from Proto-Indo-European **h₂ep-*, from which the Iranian word for 'water' also derives. In the absence of evidence suggesting a borrowed origin, and with a plural that is apparently *āpäṃ*, inheritance is much more likely, and this option should be preferred.

---

[105] There is perhaps another example, IOL Toch 74 b1, which Adams (DTB: 47) and others have read as *(śt)w(āra) a(päṃ)* "four rivers". However, Peyrot (CEToM) reads it as *(śt)w(āra) k · t · − ///*, and this reading is accepted by Hartmann (2013: 448).

2.5.3 TB *ām* 'silence'

It has often been proposed that Tocharian B *ām* 'silence', as an adverb 'quietly, still', derives from Iranian. Tremblay (2005: 438), for instance, suggests that it is a direct borrowing from Sogdian *ʾrmyh*. However, this Sogdian word means 'deserted, isolated' (SD: 57) or 'solitary' (DKS: 244), which semantically does not fit. In fact, Tremblay probably took this word to mean 'still' in the phrase *ʾrmyh nyδ* "to sit *ʾrmyh*", but this should rather be translated as "s'asseoir à l'écart" (cf. Benveniste 1946: 66 "elle s'assit à l'écart").[106]

Isebaert (1980: 47) non-committally ("[g]eheel vrijblijvend") suggests to derive the Tocharian B word from another Iranian etymon: a Middle Iranian **āh(a)m* 'sitting' < **āh-(a)ma-*, from the root **āh-* 'to sit' (Cheung EDIV: 153-154), semantically parallel to Latin *sēdāre* 'to calm down, to restrain (cf. Eng. *to sedate*)'. Naturally, 'to sit' and 'silence' can be connected to one another thanks to the formulation "to sit in silence" which often occurs in Tocharian as well. However, in my view, this derivation is difficult as no Middle Iranian **āh(a)m* is attested; indeed, there is no other trace at all of an Iranian **āh-(a)ma-*.

On the other hand, **arma-* 'quiet' as extracted from Avestan *ārmaiti-* 'name of an Amesha Spenta' and Vedic *arámati-* 'piety, devotion' would theoretically fit semantically. However, it does not fit phonetically: Proto-(Indo-)Iranian **aráma-* is to be reconstructed to account for **arámati-*, which is to be metrically restored for Avestan *ārmaiti-* 'Proper, Devoted thought' (cf. Hoffmann & Forssman 2004: 36). The reconstruction of a form **arma-* coupled with the meaning 'quiet' is also anything but assured.[107]

Nevertheless, an Iranian adverb **armai̯* must have existed. It is found in the Avestan phrase *armaē šad-* 'sitting still, ruhig stehend', in the Avestan adverb *airime* 'still' (cf. Narten 1968: 247), perhaps in Sogdian *ʾrmyh* 'à l'écart' and in Bactrian αρμαυ- / αρμαστο 'to be present, be (temporarily) resident, stay, remain', which Sims-Williams (2007: 194) explains as deriving from **arma-* 'still' and **āh/āsta-* 'to sit, remain'. This derivation is fur-

---

[106] Note Gauthiot's outdated translation (1912: 485) "elle s'assit soumise".

[107] Pinault (2016: 123f.) proposes that there existed a PIIr. adjective **Hára-* 'fitting, proper, right' and a stem **Haráma-*, enlarged in *-ti-* (Avestan *ārmaiti-*, Vedic *arámati-* 'Devotion').

ther supported by the Avestan collocation *airime.aŋhaδō* 'sitting still' (Yašt 13:73, cf. Narten 1968: 247).

There is no trace of an Old Iranian noun **arma-* 'stillness' or 'silence', or 'quiet', but if such a noun were borrowed in Tocharian B, it should have been ***erme*. Perhaps one could suggest an *ad hoc* simplification of the cluster **rm* to *m* (in Iranian), but then it would have become ***āme* in Tocharian (if the vowel was lengthened in Old Steppe Iranian) or ***eme* (if it was not lengthened). From a Middle Iranian language one could perhaps indeed expect *ām*, again with *ad hoc* simplification of the cluster, but such a noun seems to be lacking from all Middle Iranian languages. Moreover, no meaning close to 'silence' is found in the Middle Iranian possible cognate Sogdian *ʾrmyh* 'à l'écart'. For completeness' sake I want to add that the hypothesis that TB *ām* 'silence' would be a cognate of Sogdian *ʾwrʾm* 'peace, quietude' (Sims-Williams 2020: 36), Persian *ārām* 'peaceful, quiet', Persian *ārāmiš* 'quietude', etc. (not found in the scientific literature) is also to be excluded, because these words are too different from the Tocharian form. There is no way to expect a reduction of **ārām* to **ām*, whether in Old or in Middle Iranian.

In conclusion, there is no clear Iranian source for Tocharian B *ām* 'silence', adv. 'quietly'(*pace* Tremblay 2005: 438): none that would fit for the meaning and none that would fit for the form. An Iranian source is theoretically not entirely excluded, but no suitable source form has yet been identified. Therefore, it is advisable to follow Adams (DTB: 47) for the moment, and to consider this word to be of unknown etymology.

2.5.4 TB *kāswo* 'skin disease, in particular leprosy and alphos'

Adams (DTB: 165) translates Tocharian B *kāswo* as 'eruption, inflammation of the skin', but since it corresponds to Buddhist Sanskrit *kuṣṭha-* 'leprosy' (see e.g. M-W: 297-98; Emmerick 1970: 79), 'leprosy' seems to be a more fitting translation (cf. Isebaert's translation as *melaatsheid*, that is, leprosy, 1980: 196). For a discussion of the meaning of Tocharian B *kāswo* see below.

This word does not have an evident Indo-European etymology. Isebaert (1980: 197), following an idea by Van Windekens (1977), proposes to derive *kāswo* from PIE **gʷos-u̯ā/ō(n)*, a derivative from the root **(s)gʷes-* 'to go out, to burn out' (cf. LIV²: 541-2).

Van Windekens' main argument (1977: 141) is based on the parallel with Vedic *jásu-* 'Erschöpfung, Verschmachten' and especially *jásvan-* 'elend', which "offre une structure morphologique qui se superpose nettement à celle de tokh. B *kāswo* : ce substantif féminin s'explique excellemment à partir d'i.-e. *(z)gᵘōs-u̯ō(n) […]". Naturally, *jásvan-* needs to derive rather from an *e*-grade *gᵘes-u̯ō(n) because of the *j-*, and the supposed ō-grade in the root has no morphological justification. Further, the semantic shift proposed by Van Windekens relies on an intermediary meaning *'disease' > *'leprosy', which relies on Dutch *melaats* 'leper' ← French *malade*.

From a methodological point of view, I think that loanwords should preferably not be used as parallels to support language-internal semantic changes. Nevertheless, 'sick' > 'leper' is an attested semantic change, in cultures where leprosy was "the sickness *par excellence*" (or "the grave sickness *par excellence*"). An example is found in Middle Cornish *claf* 'sick' > 'sick, leprous' and, in the plural, *cleyvon* 'lepers'.[108] In this way, I slightly disagree with Adams (DTB: 165) who writes about Van Windekens' proposal that "the semantics are anything but compelling", but he is not entirely mistaken, since the Sanskrit word does not mean 'sick'. It is principally on morphophonological grounds that Van Windekens' proposal needs to be discarded, the decisive argument being that his Indo-European derivation does not work.

Winter (1962: 113[10]) proposed to link TB *kāswo* to TA *kāsu* 'good thing', saying it seemed "inadvisable to separate [them]". Obviously, the semantic connection is difficult to see. For any semantic relationship to exist, there would perhaps need to be a certain cultural setting, in which diseases are particularly avoided subjects, and in which leprosy is one of the very worst diseases. One could then perhaps think that a word meaning 'good' was euphemistically taken to designate a specific (in fact terribly bad) disease.

There are no indications that there was such a cultural setting in the Tocharian environment, nor are there any clues for another way to explain this shift. Another argument against this etymology would be that, for this word meaning 'good', the original meaning would have been completely abandoned in Tocharian B, whereas its secondary meaning 'leprosy' would have

---

[108] Cf. Williams (2011). Probably through *claves bras* 'big disease' which designates leprosy in Modern Cornish (cf. George 2020). I thank Pierre Faure for bringing this latter fact to my attention.

never arisen (or not be attested) in Tocharian A. This makes the argument, already weak, even less probable. Kim (2019: 186[29]) also rejects Winter's suggestion.

Hilmarsson (1996: 107) connected this word to Proto-Germanic *haswa- 'grey, white' (on which Kroonen 2013: 213-14). This will be discussed more in detail below, in connection with Del Tomba's renewed etymology.

Del Tomba (2020: 123ff.) has proposed a semantically attractive connection of Tocharian B kāswo to the Proto-Indo-European root *ḱes-/*ḱseu̯- 'to comb, scratch', based on the idea that the skin of lepers often shows lesions which are similar to scratches and scabs, and look like they can be scratched off easily. He nevertheless notes that the vocalism of the Tocharian form is problematic.

Assuming that kāswo designates not leprosy in general, but "white (tuberculoid) leprosy" (cf. also Schmidt 1986; 68-70; 2018: 74), Del Tomba further makes a series of assumptions. Before addressing these, I need to mention a number of facts. White leprosy, or rather alphos or vitiligo, is a disease with which patients lose the pigmentation of (parts) of their skin. It differs in fact from tuberculoid leprosy, which is a less dangerous and contagious form of leprosy, with which patients usually have pinkish patches appearing on their skin. It is unclear to me whether white leprosy in the texts cited by Del Tomba (2020: 123 and 124) refers to alphos or to tuberculoid leprosy.

It is clear that when kāswo corresponds to Sanskrit kuṣṭha-, it definitely designates leprosy (either tuberculoid, borderline or lepromatous). This meaning of kuṣṭha- as referring to the most dangerous, limb-decaying forms of leprosy is confirmed by early Indian medical texts (Rastogi & Rastogi 1984).

As Del Tomba (2020: 124) notes, it can also be established that when Tocharian B kāswo translates Sanskrit kilāsa-, it designates rather alphos, or skin depigmentation (see Del Tomba 2020: 124 with references). Thus, in general, it seems safer to assume that the primary meaning of kāswo was 'skin disease', but a skin disease of a rather grave nature, not a simple rash.

Because of the correspondence of kāswo with kilāsa-, Del Tomba (2020: 124) rehabilitates Hilmarsson's etymology. He proposes to etymologize kāswo as from PIE *ḱh₂s-u̯o- 'having whiteness' → *ḱh₂s-u̯e-h₂ 'mass of

whiteness' > PTch *kaswå > TchB kāswo 'white leprosy; skin disease'.[109] In my opinion, this etymology is not impossible but there is another argument to be taken in account. Since TB kāswo does not seem to refer to one disease in particular, but rather to a series of diseases, all related to the notion of '(serious) skin disease', there are two main possibilities for its derivation: 1. either its initial meaning was much less specific or 2. it was borrowed and its exact meaning was unstable for some time.

If the first hypothesis is to be preferred, then I suggest a slightly different evolution. From *$\hat{k}h_1$-s-u̯o- 'being grey or pale' → *$\hat{k}h_1$-s-u̯e-h_2 'greyness or paleness' one could imagine a result TB kāswo 'skin disease which makes one's skin dark (as in the lesions of leprosy) or pale (as in alphos or lepromatous leprosy; or perhaps also pink as in tuberculoid leprosy)'. A parallel for another, perhaps better solution, is found in a semantic path to be observed in Champenois dialects. It is the word havé adj. (cf. Tarbé 1851: 74), variant havi 'desséché, contracté' (cf. French hâve 'pale and thin because of a disease or hunger'). Those forms are ultimately borrowings from the Germanic word *haswa- discussed above. The semantic shift is *'grey or pale' > *'sicklish' > 'weak, made thin' > 'dried up'. The skin of lepers appears to be particularly dry and ribbed. The meaning of TB kāswo would thus add one supplementary step from 'dried up', attested for Champenois, to 'leper'. This supplementary step I believe, is very plausible given the appearance of lepers. Whether the first hypothesis (which is more similar to Del Tomba's) or the second should be preferred depends on how one sees the disease or interprets the Tocharian word.[110]

It is also possible that the word kāswo is ultimately a loanword. In this case, the source is probably to be found in Iranian. Lidén (1916) and others after him (e.g. Oettinger 1983: 330; Tremblay 2005: 441) have suggested an Iranian source. The point of departure is found in Avestan kasuuiš designating 'someone having a specific disease'.

The etymology of Avestan kasuuiš is discussed in Kellens (1974: 367-68; cf. also Bartholomae 1906: 158). After rightly rejecting a number of etymologies he proposes to read it as a bahuvrīhi *kasu-viš- 'qui a un petit poison,

---

[109] There is no reason to reconstruct this root with *h_2; usually it is reconstructed as *$\hat{k}eh_1$- (cf. Lubotsky 1989: 57).
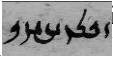
[110] I thank Philippe Hattat for providing me with the Champenois examples and references.

des pustules'. This proposal seems very unlikely to me: in what way are pustules small poisons? However, my main argument against this etymology is that *viš-* in Avestan and in Iranian in general designates liquid poison, a poison that springs out.[111] One could perhaps venture to say that **kasu-viš-* designates pus, and thus the pustule containing the pus by metonymy, but pus is not a poison. Naturally pus should not be drunk: but it is not poisonous (except that it might perhaps look like the yellowish poison of dragons?).

Rather than accepting Kellens' etymology, I believe Avestan *kasuuiš* to be an *-iš* adjective of a noun **kasū-*, cognate of Sanskrit *kacchū-* 'scabies', and probably meaning the same thing. This would fit very well the context of both attestations (cf. Bartholomae 1906: 158 for more details).[112] Lepers are already cited in Vendidad II:29 and Yašt 5:92 (*paēsō*). The list of people who are forbidden to approach the libation in Yašt 5:92 is based on physical appearance (cf. Bartholomae 1906: 158). It is possible that the translations by Pahlavi ‹kynyk› and Persian كينه كش *kīna-kaš* 'vengeful' (cf. fn. 112) are based on a confusion with NP *kana* گَنَه 'scab-causing mite; leech'.[113]

---

[111] As in the Hom Yasht XI:11, when the dragon Sravara ejects a yellowish poison; or as in the *Zand-i Wahman Yasn* IV:2 when the invaders of Iran, those of the race of Xešm, will "piss poison (*wīš*)".

[112] The Pahlavi translation of *kasuuiš* is a real problem. Bartholomae (1906: 158) already notes Justi's reading *kēnxūn* 'angry-blooded' (transcribed *kīnḫūn*) 𐭪𐭩𐭭𐭧 which he could not find in the manuscripts, which have 𐭥𐭪𐭥 and 𐭠𐭪𐭥 or 𐭥𐭥𐭯𐭩𐭲.

I have found the form in L4 and ‹wdzwytk› in 4711 B1, where it is translated (in Persian) as رنجیده *ranjīda* 'hurt' (maybe based on a reading **‹r(n)dčytk›*). I am unsure of the reading of those words, and they could be read in many various ways. It is true that F10 has ‹kynyk› 'vengeful' rendered as كينه كش *kīna-kaš* 'idem' in the Persian translation. In general it seems that this word (*kasuuiš*) was not correctly understood by the commentators (or at least, some later ones). Perhaps the translation 'vengeful' was influenced by a folk-etymology based on *kasu-* 'small' (as in 'petty'), as was proposed by Justi (1905: 95).

[113] Perhaps etymologically related to *kand-/kan-* 'to dig' (so Skjærvø 1994: 277) as in 'the digging one' because of the hole-like spots it leaves in the skin, if the scabs remain untreated, or, if the meaning 'leech' is primary, perhaps because of the animal's "blood-digging" habits? Perhaps also it is somehow related to Waxi *kanek* 'mosquito' (cf. Morgenstierne 1938: 526 for the Waxi word; cf. Hasandoust 2014: 2281 for the Waxi connection to the Persian word).

It is not difficult to imagine that an original Pahlavi *ﯟﮔﯟﮔ *kanagīg
'mangy'[114] was corrected in ﯟ-اﯟ   kēnīg 'vengeful' and translated as such
in Persian. The reason could be that the Middle Persian word *kanag, from
which NP kana, itself rare, must derive, is not otherwise attested as far as I
know, and the copyists perhaps did not know it. This "correction" also hap-
pened with the Avestan word kapasti- 'name of a disease or a poisonous
plant (?)' also translated by NP kīna 'vengeance' (cf. Shapira 1998: 220[93]).
Positing the existence of the predecessor of the New Persian word kana in
Middle Persian further solves another problem: if the form (*)ﯽﯘﮔ kēnxūn as
Justi read it (cf. fn. 112) indeed exists, then one could easily see it as a de-
formation of *kanxūn 'blood-digger' (of which kanag is perhaps originally a
back-formation), rather than *kēnxʷāh 'rachsüchtig' as Justi (1905: 95) pro-
posed. That kasuuiš was seemingly less and less clear to the commentators is
also evident from the various translations and unclear paraphrases in various
manuscripts (cf. fn. 112).

If my explanation of the deformation of an original Pahlavi translation as
*kanagīg 'scabby' is to be accepted, then it can be proposed that kasuuiš
indeed meant 'scabby' and the original stem *kasū- 'scabies' would thus be
a perfect cognate of Sanskrit kacchū- 'scabies'. In that case, they would both
need to go back to *kaśćū-. Since this word does not have any Indo-
European cognate, it could perhaps be of BMAC origin.[115]

If PIIr. *kaśćū- is indeed a borrowing from the BMAC language, one
could perhaps imagine that TB kāswo was borrowed from the same source.
However, this is difficult, since whatever sequence of sounds gave rise to
PIIr. *-sć- would not likely be rendered with -s- in Tocharian. For instance,
what we reconstruct as BMAC *ć on the basis of Indo-Iranian, is rendered as
PT *ś (PIIr. *ćaru̯a- : PT *śer(ə)we). Furthermore, an accented vowel *a has
not been found in other BMAC loanwords in Tocharian (see chapter 3). The
word is thus unlikely to be a direct BMAC loanword in Tocharian.

Alternatively, one could also imagine that TB kāswo was borrowed from
Old Steppe Iranian. A Proto-Indo-Iranian form *kaśćū- should become
*kasū- in Old Iranian (cf. Lubotsky 2001a). However, an Old Steppe Iranian

---

[114] Mange is a category of diseases that includes scabies. People sick with scabies
are usually called mangy.

[115] Mordvin kośkilda / kośkelda 'scabies' might perhaps be connected to this Indo-
Iranian etymon by means of borrowing.

form *kasū- would be borrowed as †kes or perhaps †kos < *kesu in Tochari-an. Therefore, Old Steppe Iranian origin can be excluded for the Tocharian B word kāswo.

Yet another option would be to consider Bactrian origin. A Pre-Bactrian form *kaswā (< feminine OIr. *kasu-ā-) would have given Pre-Tocharian B *kas(u)wa, readapted as kāswo, like Pre-Bactrian *kōškā → TB koṣko (see Bernard & Chen 2022), following the same model as maiyya '(supernatural) power', remade into maiyyo. From the Iranian point of view, this would be an -ā- abstract on an adjectival stem *kasu- 'mangy'. Reconstructing such an adjectival stem is nevertheless gratuitous, and there is no evident Middle Iranian candidate with such a formation in this word. A Pre-Bactrian origin is thus also unlikely for TB kāswo.

Dragoni (2022: 89f.) has proposed to see in kāswo a borrowing from Pre-Khotanese *kasūwa-, more specifically from its accusative singular *kasūwu. As to the exact meaning of this Pre-Khotanese word, it might have designated skin eruptions or inflammations, like Late Khotanese kasaa- (cf. Dragoni 2022: 90).

The semantic shift from 'scabby' to 'leprous' is not uncommon, because leprous skin is very often scabby. However, 'scabies' on one hand and 'lep-rosy' on the other, as well as 'leper' and 'mangy' are often carefully differ-entiated, as far as I could find. This is not trivial, as leprosy and scabies, especially in some variants of each disease, present similar alterations of the skin. This is notably true of the many scabs and reddish spots found on the skin of sufferers of both diseases. Because of their never healing lesions, lepers are also prone to mange.

The fact that these two diseases, scabies and leprosy, are not easily con-fused must be due to the fact that their consequences are very different. Both scabies and leprosy are contagious (although scabies is much more conta-gious), but it is leprosy that has terrified the imagination of European, Cau-casian, Indian and Chinese peoples. It is leprosy with which the highest so-cial stigmata are attached: the falling of the limbs and the neurological dam-age it causes are irreversible. Therefore, the semantic change between 'sca-bies' and 'leprosy' could in my view not have happened language-internally. Nevertheless, it is not impossible to imagine a setting where mange was uncommon, such as Central Asia, and upon borrowing such a word from Iranian tribes, Tocharians used it to refer to skin disease in general, but to the most fierce leprosy in particular. This would explain the various meanings

that can be attributed to it in the various texts, where it corresponds both to Sanskrit *kuṣṭha-* and *kilāsa-*. Perhaps also the Iranian donor language had a more general meaning for this word, such as 'skin disease'. Cf. Sinhalese *kas* 'itch, skin disease' < OInd. *kacchū-* 'scabies' (cf. CDIAL p. 203).

To conclude, although the inherited Indo-European etymology from the root *$\hat{k}h_1$-s-* 'to be grey, white' is possible to explain the Tocharian B word *kāswo* 'skin disease, in particular leprosy and alphos', I suggest that the Iranian etymology which was already proposed in the past might function as well. I suggest an emendation of the Pahlavi translation of the Avestan word *kasuuiš* on the basis of a New Persian word which designates the animal that causes scabies. With this emendation, the Avestan word *kasuuiš* may be taken to mean 'mangy'. If Tocharian B word *kāswo* was borrowed from Iranian, it should for phonetic reasons have been borrowed from Pre-Khotanese, because of the regular correspondence between Pre-Khotanese **a* : Tocharian *a* and Pre-Khotanese *-a*, acc.sg. *-u* regularly corresponding to Tocharian B *-o*, as proposed by Dragoni (2022).

2.5.5 TB *tāno* 'seed, grain'

It is generally assumed that Tocharian B *tāno* 'grain', obl.sg. *tāna* is inherited from Proto-Indo-European (cf. DTB: 303 with references). That, however, seems difficult in regard of the fact that the only cognates of this word, in my view, are Indo-Iranian: Vedic *dhānā́-* f.pl. 'geröstete Getreidekörner' (cf. EWAia[1]: 787), Avestan *dānā** in *dānō.karš* adj. '(Getreide)körner (ver)schleppend' (cf. AiW: 734), Khotanese *dānā-* f. 'grain, corn', Sogdian *δān* 'seed', Middle Persian *dān* and *dānag* 'seed, grain' (CPD: 24), New Persian *dāna* 'grain, unit',[116] Xorāsāni *deng* 'corn' (Monchi-Zadeh 1990: 50). Bailey also noted some assimilated forms (going back to **nānā* and **nānā-ka-*): Yidgha *nānoγō*, Pashto *nīne* 'parched grain', *nānga* 'blackberry' (DKS: 156).

The connection with Lithuanian *dúona* 'bread' and Latvian *duõna* 'slice of bread, heel of a loaf' (cf. EWAia[1]: 787) seems rather weak to me. We have a group of close-knit words meaning 'grain' in Tocharian and Iranian, and this meaning is not even found in the Indic cognate. On the other hand,

---

[116] A meaning 'seed' can also be seen in the New Persian compound *wan-dāna* 'seed of the *wan*-tree' (for which see Maggi 2003: 123[37]).

the Baltic words designate bread, and there is a long way from the whole grain to bread. Even more so: the Latvian form also designates the crust of a loaf of bread, which Peyrot (2018: 258) connects with an unaccented word *duona*, which has various meanings relative to an edge (see details in Peyrot, loc. cit.). He suggests an evolution 'edge' > 'edge, heel of a loaf' > 'slice of bread' (> 'bread' in Lithuanian) (2018: 259), which I find more convincing than an evolution, to my knowledge unparalleled, from 'grain' to 'bread' (even with multiple intermediary steps).

Even if the Baltic words were cognates of the Indo-Iranian ones, which is far from evident, it seems difficult to believe in a Baltic – Indo-Iranian – Tocharian isogloss, where Tocharian shares the exact same meaning as Iranian, and not Indic: in no modern Indo-Aryan word does the descendant of this word mean 'grain', cf. CDIAL: 387. All Middle and New Indo-Aryan languages show a different meaning for this word, such as 'parched grain' and 'growing rice' (cf. CDIAL: 387), which implies that the Proto-Indo-Aryan meaning of *dhā́nā́- was different from simply 'grain, cereal', a meaning which can on the other hand safely be reconstructed for Proto-Iranian.

Logically, then, Tocharian B *tāno* should be an Iranian loanword in Tocharian (cf. already Duchesne-Guillemin 1941: 180; later Schmidt *apud* EWAia[1]: 787; Peyrot 2018: 258 and Del Tomba 2020: 151[227]). A problem with this hypothesis, which otherwise seems rather evident, is, as Peyrot mentions (2018: 258-59), that the class to which *tāno* belongs, namely the *kantwo*-type, is rather small and comprises no loanwords. Nevertheless, Del Tomba recently (2020: 126) adduced a strong argument in favor of this hypothesis: Tocharian B *mālo* 'alcohol', obl.sg. *māla*, which doubtlessly was borrowed from an Iranian language, namely Pre-Bactrian, and also belongs to this same class. I have also suggested two more members of this noun class to be loanwords from Iranian:

a) Tocharian B *maiyya*, Archaic TB *meyyā* (s.v. *maiyya*), which was later remade into TB *maiyyo* obl.sg. *maiyya* (cf. Peyrot 2008: 99f.). This word thus shifted to the *kantwo*-type in historical times, which shows that this class was more productive than previously assumed.

b) Tocharian B *kāswo* 'skin disease; in particular leprosy and alphos', also has its oblique singular in -*a* and belongs to the same class as *tāno* and *mālo*. I have argued (s.v. *kāswo*) that this is also likely to be an Iranian loanword.

In case my argumentation is to be followed, there are at least four loanwords in this class: *maiyyo*, *tāno*, *kāswo* and *mālo*. Dragoni (2022) has identified even more loanwords in this category, such as TB *kātso* A *kāts* 'stomach, belly, womb' and TB *tvāṅkaro* 'ginger'.

If Tocharian B *tāno* is borrowed from Iranian, the exact source must be determined. The main clue here is the ending and class appurtenance of the Tocharian word. I have shown that Old Steppe Iranian final *-ā is in Tocharian rendered in the same way as in Old Steppe Iranian *ā in other positions: Proto-Tocharian *a > Tocharian B *a. This correspondence is seen in Archaic TB *meyyā* and TB *newiya*. In the case of Archaic TB *meyyā*, it can also be observed that a new form *maiyyo* was made, to follow a more productive model than a nom.sg. -*a* : obl.sg. -*a* declension. Based on this, I have also suggested that TB *raimo** 'dust, dirt' originally comes from **reyma* (Old Steppe Iranian **ra̤imā-* 'dirt') but was transferred to the productive *okso*-type (nom.sg. -*o* : obl.sg. -*ai*).

Thus, if an Old Steppe Iranian word **dānā-* f. 'grain, seed' was originally borrowed as PT **tana*, it could at any period of time have been remade into a noun *tāno*, of the *kantwo*-type.[117]

On the other hand, an Old or Pre-Khotanese (or Proto-Khotano-Tumshuqese) form **dānā-* could also have been borrowed as PT **tano*. Even a Pre-Khotanese or Old Khotanese form **dāna-* m. could in theory have yielded Tocharian B *tāno* (cf. Peyrot, Dragoni & Bernard 2023 for a discussion of this form).

There is yet another argument that can be used to determine the source of borrowing: there is a diminutive form *tanākko*. The suffix -*kko*, as Dragoni (2022) demonstrated, is likely to be of Khotanese origin. Thus, this diminutive renders more likely the hypothesis that the base word *tāno* derives from some stage of Khotanese as well.

Finally, it is striking that there is a high number of botanical terms among Khotanese borrowings in Tocharian, where *tāno* would fit well, while there are not that many from Old Steppe Iranian, the only plausible candidate being *welke** 'stalk' or 'petal' (s.v.).

---

[117] However, one would have to exclude a source form OSIr. **dāna-*, which could only have led to TB †*tāne*. The protoform **dāna-* needs to be reconstructed for the Middle and New Persian forms: only **dānaka-* or **dānakā-* can be the source of Middle and New Persian *dān(ag)*.

Apart from these two sources, there can hardly have been another Iranian donor language, both for historical and morphological reasons. The word-final vowel warrants an earlier stage, which excludes Sogdian, Parthian and historical Bactrian. Pre-Bactrian is also impossible here: it would have given TB †*lāno* in the same way it gave Tocharian B *mālo* < *madu-*, and, as I argue together with Ruixuan Chen, in the same way that Pre-Bactrian *kōškā-* was rendered as TB *koṣko* (Bernard & Chen 2022). It could also hardly have been a smaller, undetermined Middle Iranian language: how could a word like 'seed, grain' have been borrowed from a smaller, distant language of little political importance?

In conclusion, Tocharian B *tāno* 'seed, grain' can hardly be inherited, as non-Iranian cognates with this precise meaning are lacking. It is very likely an Iranian borrowing, as has been proposed multiple times before. More precisely, it can be either an Old Steppe Iranian loanword, which implies that it was first borrowed as a noun of the *maiyya*-type (nom.sg. -*a* : obl.sg. -*a*), and later shifted to a different declension class, the *kantwo*-type (as *maiyya* became *maiyyo*), or it was borrowed from Pre-Khotanese, Proto-Khotano-Tumshuqese, or Old Khotanese (cf. Dragoni 2022). In favour of the Khotanese hypothesis is the existence of the diminutive *tanākko*, whose suffix -*kko* ist best derived from some stage of Khotanese.

2.5.6 TB *waipte* 'separately, apart'

Tocharian B *waipte* 'separately, apart' is generally considered to be of Iranian origin (e.g. DTB: 668), after Schmidt (1985: 760f.). Schmidt's etymology consists in deriving *waipte* from *waiptay* ← Old Iranian *hwai-patayai* '(dat.) self-master', "für sich".

From the semantic point of view this derivation is not fully convincing. It is, of course, difficult to exclude that *hwai-patayai* 'für sich' could eventually come to mean 'for itself' = '(each thing) for itself', but the assumed semantic development is far from compelling.

Formally, the problems with Schmidt's etymology seem insuperable. According to the correspondences established, "*hwai-patayai*" would yield †*waiptyai* or †*waipcai vel sim*. There seems to be no way to derive the final -*e* of TB *waipte* from the sequences -*ay*- or -*ai* of *hwai-patayai*, nor from a combination of the two.

To conclude, Schmidt's etymology of Tocharian B *waipte* 'separately, apart' is impossible formally. For the moment, this adverb remains without a clear etymology.

2.5.7 TB *waimene* 'difficult'

The etymology of the Tocharian B word *waimene* 'difficult' is arduous. As recalled by Adams (DTB: 668), Van Windekens (1976: 643) suggested to view it as a borrowing from Sanskrit **vaimanas-* 'tired, discouraged', a derivative of *vimanas-* 'perplexed, discouraged'. As Adams writes, "the putative Sanskrit morphology would be unparalleled". The semantic change from 'perplexed, discouraged' to 'tired' and 'difficult' is not evident either.

Isebaert (1980: 184) derives the word from Proto-Indo-European and sees in it a lexicalized locative form of an unattested **dṷoi-mo-* (derived from 'two'), but, as Adams writes (DTB: 668) there is no other example of a frozen locative becoming an adjective in Tocharian. Further, the semantics are not evident either.

Pinault (2015: 170) proposes to derive this word from **wai-me-ne* 'woe on us': **wai* 'woe' + *-me*, the plural suffixed pronoun + *-ne* locative ending. He explains the functional change as being perhaps the result of a reinterpretation of the interjective phrase as a predicate in deontic sentences, and gives a possible example. However, in Tocharian pronoun suffixes cannot be attached to interjections, so that this explanation is syntactically problematic.

Indeed, although it is tempting to see in this word a trace of the Proto-Indo-European interjection **ṷoi* 'woe', I believe a real problem lies in the fact that woe and difficulty are concepts that are quite far apart. Woe relates to the curse of destiny on man, and to the sufferings that ensue. Difficulty relates more to the obstacles man has to overcome. Apart from that, it is not that easy, I believe, for an interjection to become frozen, to acquire a locative suffix, and then become an adjective.

Schmidt's explanation (1985: 762) from Old Iranian **hṷai-manah-*, with the semantic development 'self-willed' to 'difficult' ("wer oder was seinen eigenen Sinn hat", d.h. "eigensinnig", "schwierig" […]") is also unlikely. Indeed, in French, in English, in German, and in many other languages, someone who is self-willed is called 'difficult' (*don't be difficult*, French *ne fais pas le difficile*, etc.). However, in this case the development would be precisely the other way around. That something difficult would be deemed

'self-willed' seems to require a very specific phraseology in a specific language, of which we have no trace in Iranian, and which seems thus unwarranted.

In conclusion, although the Tocharian B word *waimene* 'difficult' seems to have an Old Steppe Iranian appearance, its etymology remains unclear for the moment.

2.6 Discussion of the features of Old Steppe Iranian

I hope that the multiple arguments developed throughout this chapter have been consistent enough to convince the reader that there was an Old Iranian language, of a quite archaic nature, that came in contact with both Tocharian languages at a very early stage or, more likely so, with Proto-Tocharian. The loanwords from this layer are coherent in nature, and the sound correspondences I have been able to establish are regular. The lexical categories these loanwords belong to are also limited and give a coherent idea of the type of contact that existed: military terminology; various objects; animals and animal products; grammatical words; abstract notions. A logical conclusion from this fact is that this layer actually only represents one single language, the so-called Old Steppe Iranian language.

This layer of Old Steppe Iranian loanwords in Tocharian is homogeneous, and almost all – if not all – of these loanwords were borrowed according to well-established borrowing patterns, implying a high degree of fluency or knowledge of this language among Tocharian speakers.

Although few words played an important role in the delimitation of this language, one in particular played a decisive role. It was previously believed that an Old Persian word had spread as far as Tocharian, or to the Old Iranian language in contact with Tocharian, namely TB *peret*, TA *porat* 'axe'. By determining that the predecessor of TB *peret*, TA *porat* must have been regularly Old Iranian **paratu-* and was not an indirect borrowing from Old Persian as previously assumed, I hope to have demonstrated that no Old Iranian loanword in Tocharian was demonstrably borrowed from a Persic language.

The question that ensues is of course the following: to which branch of the Iranian tree did this language belong? To which known (i.e. written) Iranian language was this language closer (or closest)?

Answering this question is not easy. The Iranian branch is by far the most geographically extended and, in terms of the number of languages, the rich-

est of all the Indo-European branches. Out of an unknown number of branchings,[118] only two Old Iranian languages are better attested, Avestan and Old Persian, and a third one, Old Scythian, is attested in an extremely fragmentary and mostly indirect fashion. Until now, no other Old Iranian language has been attested, although we know they must have existed, since very few of the many existing Iranian languages descend directly from the three mentioned above.[119] Attested Middle Iranian languages are more numerous. We can count eight of them: Middle Persian (in fact two main varieties: Pahlavi and Manichean Middle Persian), Parthian, Bactrian, Sogdian, Khotanese and Tumshuqese, Khwarezmian, and Alanic.

After the Middle Iranian period, the number of attested New Iranian languages is much greater, and too uncertain to be enumerated here. We are thus presented with the figure of a funnel, where – traditionally – all the modern Iranian languages are supposed to descend from a group to which one of these older languages, both Middle and Old, belonged. It has been demonstrated recently that this model is impossible (Korn 2016; 2019). At least several Middle and New Iranian languages go back to unattested Old Iranian languages, which were not necessarily in a closer relationship with other known branches. Such is the case, for instance, of Balochi (see Korn 2005).

Based on this state of affairs, it cannot be stressed enough how crucial the discovery of the Old Iranian layer of borrowings (dating back to Isebaert 1980, Schmidt 1985, Tremblay 2005, detailed and made explicit in Peyrot 2015a; 2018a) is for Iranian studies. Not only do we have one more Old Iranian language to count on for drawing the phylogenetic tree of Iranian languages, one that is not hypothesized for the sake of it, but attested through another language, but we also have the confirmation and rejection of a num-

---

[118] They are unknown for two reasons: 1. Branches may have existed of which we have no trace or knowledge, and 2. Even for the attested languages, there is no consensus as to how many inner-Iranian branches have to be posited.

[119] More precisely, Avestan has no known (direct) descendant, and an undocumented dialect of Old Persian gave birth to Middle Persian, another to New Persian (through its unattested direct predecessor, cf. Korn 2021), and possibly to one or two other dialects. Little is known about Old Scythian (or Sarmatian), for a summary on our knowledge thereof, see Novák 2013: 9f., but it seems to be the ancestor of the Scythian or Alanic languages, which constitute two closely related groups of languages.

ber of predictions made by Iranists over the past two hundred years (see below).

A very important step in the present study consisted of excluding a group of words which were considered Iranian, or Old Iranian, and prevented scholarship from establishing straightforward facts (for example: *witsako* 'root', s.v.). I also had to review a number of etymologies, and revise various facts concerning these etymologies, for example by proposing a Proto-Iranian **paratu-* and not **paraću-* for the word for 'axe'. As a result, I have gathered the features of the Old Iranian language in question, as found in Tocharian A and B, in a consistent and systematic comprehensive manner.

An important feature (see Schmidt 1985; Peyrot 2018a) is that PIr. **ć* in that language is reflected by **tˢ*, and that PIr. **j* (PIIr. **j* and **jʰ*) is reflected by **dˣ*, both rendered as *ts* in Tocharian. The realization **[ts]* and **[dz]* of the PIE palato-velars at a very early stage of Iranian has been assumed by Iranists for a long time, but Old Steppe Iranian offers concrete evidence for this realization.

Before discussing the features of Old Steppe Iranian in connection to the identification of that language, it should be mentioned that there will always be two difficulties that will make this identification arduous. The first is the difficulty in deciding whether the relevant sound changes occurred within Old Iranian or in the borrowing process to Tocharian. The second is that this Old Iranian language is evidently old and archaic, and that consequently some of the innovations found in supposedly related or even descendant languages might not be found in it.

In the following, I will first present a table of Old Steppe Iranian - Tocharian vowel correspondences, based on the word-studies of the present chapter (2.6.1). I will then discuss all the relevant phonetic innovations of that Old Iranian language, that is, those that differentiate it from Proto-Iranian in a clear manner. Those innovations are divided between weaker and stronger ones, from two points of view: on the one side we have clear innovations, seen in multiple Tocharian words, such as PIr. **rd* > OIr. **ld* and on the other side we have possible innovations, such as the *u*-umlaut that can perhaps be seen in *mot* 'wine' (from **madu-*), for which I was not able to find any other example or counter-example. In section 2.6.3., I will present some lexical isoglosses, although with the caution that, not all Iranian languages being as well documented as each other, there is an inevitable bias in terms of lexical isoglosses. In section 2.6.4., I will discuss relevant seman-

tic changes. These will often be more certain than the phonetic changes in the sense that their being Iranian or Tocharian developments is easier to establish.

Both in section 2.6.2 and 2.6.4, when discussing those changes, I will try to systematically compare them with a number of Iranian branches or languages. This will inevitably be impaired by a few difficulties, the greatest of which being the fact that some languages are much better studied than others. Ideally one should compare the Old Steppe Iranian language with every Iranian language, but this would lead to a much longer and not necessarily more useful discussion, as many Iranian languages are both attested too late and situated too far away to yield any significant result.

The languages I will compare this Old Iranian language with are: Old Persian (and the Persic group in general), Sogdian, Yidgha (or Yidgha-Munǰī when necessary and possible), Balochi, Bactrian, Khotanese (and Tumshuqese), Khwarezmian, Pashto, and, very importantly, Ossetic, since Ossetic is a direct descendant of a Scythian steppe language. I will also compare it when possible with Wanǰī, an extinct modern Iranian language spoken up to the end of the 19th century in the Wanǰ valley of Tajikistan (in the center south part of the country). The data we have on Wanǰī is scarce, so comparison with this language will be very limited.

The goal of this comparison is to permit us to ascertain whether an ancestorship status can be given to Old Steppe Iranian for any specific known Iranian language, on one hand, and what type of linguistic proximity Old Steppe Iranian has with other Iranian languages – notably, to which branch of the Iranian tree does Old Steppe Iranian belong. This cannot be answered within the section, and in fact, cannot be fully answered within the scope of the thesis, as it is a very complicated topic. To have a better idea of the answer, one has to combine data from this section and from the two next sections, which discuss lexical isoglosses of Old Steppe Iranian and other Iranian languages (2.6.3) and the semantic innovations of Old Steppe Iranian compared to Proto-Iranian reconstructible meanings of words (section 2.6.4.).

2.6.1 Old Steppe Iranian - Tocharian vowel correspondences

This table is not exhaustive, in the sense that it does not include exceptions, or problematic and secondary sound changes. It is thus not to be used as a reference for these sound correspondences, but rather as a succinct convenient summary. All the relevant details and problems are discussed in the following pages.

Old Steppe Iranian re-construction	Proto-Tocharian render-ing	Tocharian B reflex	Tocharian A reflex
*a	*e	e	a
*ā	*a	a	ā
*ai̯	*ey	ey, ai	e
*au̯	*ew	ew, au	o
*i (or *ə)	*ə	ä	ä
*í	*ə́	a	ä
*-íi̯V-	*-íi̯-	-iy-	-i-
*-ii̯n-	*-in-	-in-	n.a.
*-u-	*-u-	-u-	-u-
*-u	*-Ø	Ø	Ø
*r̥/ár	*er	er	ar

2.6.2 Phonetic changes of Old Steppe Iranian

In order to make inferences about the phonetics and phonology of Old Steppe Iranian, we have to differentiate sound changes from sound correspondences. For instance, Old Iranian *a was rendered as Proto-Tocharian *e, and Old Iranian *ā was rendered as Proto-Tocharian *a. This different rendering strongly suggests that there was an important phonetic distinction in timbre between Old Steppe Iranian *a and *ā, on top of the difference in length.

One option is to assume that Old Steppe Iranian *a was more front [æ] and *ā was more back [ɑ:]. Indeed, Tocharian a was perhaps also rather

back, i.e. [ɑ] or [ɒ] (cf. Pinault 2008: 421). These values would fit the phonetics of these two phonemes in Persian very well,[120] and could perhaps reflect the values of Proto-Iranian without further changes, or with only slight changes.

Another option is that originally Proto-Iranian *a was closer to schwa ([ə]), as can be found in the literature (e.g. recently Holopainen 2019: 30 for Proto-Indo-Iranian). In case Proto-Iranian *a was indeed phonetically a schwa, I would like to suggest that Old Steppe Iranian underwent a chain shift, where *a became more front, and perhaps *i and *u more central.

a. The treatment of *-θi̯- and *-di̯- and the phonetic value of *-θ- in
   Old Steppe Iranian

An interesting problem concerning Tocharian adaptations of Iranian sounds is the correspondence between Proto-Iranian *-θi̯-, *-di̯- and PT *-cc- seen in TB *waipecce* 'possession' and TB *kercci* 'palace', and possibly in TB *iścem** 'clay'. Here we have the Proto-Iranian reconstructed forms *hu̯ai-paθi̯a-*, *gr̥di̯a-*, *išti̯am* and the Tocharian borrowings from the Old Iranian language: *waipecce*, *kercci*, *iścem*, but we cannot reconstruct the Old Iranian forms easily on this basis. We have three choices here: α. either the Old Steppe Iranian forms were *hu̯ai-paθi̯a-*, *gardi̯a-* and *išti̯am*, and Tocharian palatalized them to *waipecce*, *kercci(ye)* and *iścem* or β. they had become *-čč- and *-ǰǰ- already in the Old Iranian language, or γ. in the sequences *-θi̯- and *-di̯-, the initial consonant was a palatalized sound, with an outcome *tʲy and *dʲy, which were further palatalized in Tocharian.

There are arguments in favour of all three options, and one has to propose a coherent solution before proceeding to the comparison with any documented Iranian language.

α. The first solution is the simplest, but it relies on a supposed Tocharian change *tyV > -cc-* that did not occur as far as we know (cf. Pinault 2008: 551). As is known, the PIE cluster *-ti̯-* became *tˢ* in Proto-Tocharian, while later *-ty-* sequences were borrowed as such (e.g. *pratyaikapañäktäññe* 'per-

---

[120] A great variety of languages use ‹e› to transcribe [æ] and ‹a› to transcribe [ɑ] or [ɒ], e.g. most foreign notations of Persian *a* ([æ]) ‹e› and *ā* ([ɒ]) ‹a›. There exists also a tendency, among languages, for [æ] to go to [e], and for [ɑ] or [ɒ] to go to [a], as in, for example, Xurāsān varieties of Persian (Monchi-Zadeh 1990: 1).

taining to a buddha who reserves his enlightenment to himself' DTB: 443). That this supposed Tocharian sound change *tyV > -cc-* would have occurred only in Old Steppe Iranian borrowings seems very unlikely, but not completely impossible: it would have been a punctual sound change, in order to render a sequence in Tocharian that was impossible at the time. This is still difficult, because this replacement of *-ty-* to *-cc-* would have only occurred once and for a short time. One can now add one more example, not from Old Steppe Iranian, but nonetheless from an early borrowing: TB *cowo** 'theft', which Federico Dragoni (2022) interprets as deriving from Pre-Khotanese **dyūa-* 'theft', through its accusative singular **dyūwu*. If this etymology is correct, this particular sound substitution may have taken place at a very early period, preceding later Middle Iranian and Indic loanwords.

β. The second solution, namely **(-θi̯-) > * -ti̯-*, **-di̯- > -čč-* and **-ǰǰ-* within Iranian, has the advantage of explaining the Tocharian forms in a very simple way. It would imply that the gemination was preserved in Tocharian. An argument in favor of this option is that we find *kercci* rather than **kelcci*, whereas this Old Iranian language had undergone **rd > *ld*, at least in some contexts. In this scenario, **-di̯-* would have become **-ǰǰ-* before **rd* had become **ld*, thus yielding a cluster **-rǰ-*. A difficulty with this scenario lies in the fact that no known Iranian language – to my knowledge – presents such a change.

γ. The third hypothesis, namely assuming that **θi̯-* and **-di̯* had become **ti̯y* and **di̯y* in Old Steppe Iranian, presents no particular drawbacks, in my view. In its favor: it necessitates no sound change from the Iranian side other than **-θ- > *-t-* which is needed for the previous hypothesis as well (see below), since an interdental fricative would need to become a stop before in order to become a postalveolar affricate. It has two major advantages: it explains the palatalization and provides a coherent system for the Iranian side. The posited Old Steppe Iranian **ti̯y* and **di̯y* could also have been rendered at first as **cy* in Proto-Tocharian, assimilated to **cc* in Pre-Tocharian B, and turning to **ci* in Pre-Tocharian A.

If scenario α. were to be chosen, then nothing could be said about the nature of **-θ-*, as both **-θi̯-* and **-ti̯-* would yield *-cc-* in Tocharian. Again, Proto-Tocharian would have borrowed **-θi̯-*, **-ti̯-* and **-di̯-* as such and palatalized them into **-cy-* "itself".

If scenario α. is discarded, no matter whether scenario β. or γ. is chosen, a sound change of *-θ-* to *-t-* would seem to have happened first, at least be-

fore *yod*. Since I was unable to find a parallel for a conditional defricativiza-tion of -*θ*- in front of *yod* only, it is possible that -*θ*- > -*t*- was a general shift in Old Steppe Iranian. However, I admit that there is no other evidence for this shift than the palatalization under discussion here.

If a general shift -*θ*- > -*t*- occurred, Old Steppe Iranian is closer to Osset-ic, which had *-*θ*- > -*t*- (Cheung 2002: 21), but farther away from Old Per-sian, Yidgha-Munǰī, Khotanese, Khwarezmian, Sogdian and Bactrian, for all of which *-*θ*- rather than *-*t*- needs to be reconstructed.[121] Balochi also un-derwent a change *-*θ*- > -*t*- (Korn 2005: 81), but it would be illusory to compare it here, as the sound change in Balochi is obviously secondary (since *-*θr*- became -*ss*-, for example, see Korn 2005: 89), and is part of a single phenomenon of occlusion of fricatives, which is posterior to the time period we are speaking of (cf. Korn 2005: 323f.). Pashto is more difficult to evaluate. In intervocalic position *θ* does indeed merge with *-*t*- (> -*l*-), as in *plən*, Waziri *plan* 'broad, wide' from **paθana-*, but most probably this change went through an intermediate **δ*, not through **t* (cf. already Geiger 1901: 209). The only other word from our list which may go back to a form with *-*θ*- is Tocharian A *nātäk* 'lord' (← OSIr. **nātaka-* or **nāθaka-* 'pro-tector').

If scenario β. is favored, then the proximity with Ossetic is even more prominent, as Ossetic is the only Iranian language known to have undergone **θi̯* > **-č-* > -*ts*- (transcribed as ‹c› in the transcription based on Cyrillic), cf. Iron *syfc*, Digor *sufcæ* 'shoulder' < **sufθi̯a-* (Cheung 2002: 34); Iron *fælmæcyn*, Digor *fælmæcun*, 'to get tired', from **pari-maθi̯a-* (Cheung 2002: 185). There is no trace of gemination in Ossetic here, but there is no clear indication against it having been there at an earlier stage either.

The words TB *epiyac*, TA *opyāc* 'memory' could also belong here, given their final -*c*. These words correspond to PIr. **abi-i̯āta-* but the final -*c* finds no easy explanation. It has been usually admitted that the proto-form from which Tocharian derives was **abi-i̯āti-* but there is no evidence for an *i*-stem for this word. More importantly even, there is no proof that **-ti* became **-c* in Old Steppe Iranian, or in Tocharian. If my derivation of TB *epastye* 'skill-ful' < PT **epəstiye* is accepted, it would rather show that Old Steppe Iranian **-ti-* was not palatalized in Tocharian.

---

[121] Khwarezmian even seems to preserve PIr. **θ* in some positions.

b. Old Iranian *čį or *šį → Proto-Tocharian *ś ?

The Proto-Tocharian word *śate, TB śāte TA śāt 'rich', was borrowed from a word that goes back to Proto-Iranian *čįāta- 'happy'. To explain this we can assume that the palatalized phoneme *č + yod or *š + yod was rendered as *ś- or as *tś- (with secondary "simplification" *tś- > *ś-) in Proto-Tocharian, where word-initial gemination was not possible. This is more economical than suggesting a sound change within the Old Iranian language to a palatal sound,[122] which corresponded to the palatal phoneme of Proto-Tocharian. The main question is to determine whether the Tocharian *ś- derives from a palatalized *š, as we have it in Avestan and many other Iranian languages, or from a palatalized *č.

Proto-Iranian *čįāta- has an Indo-European etymology: it derives from *kʷieh₁- 'to be quiet, calm' (Lat. quietus 'calm'), and the Proto-Iranian root can be set up as *čįaH- (cf. EDIV: s.v.). However, from the Iranian point of view, the Common Iranian form is often reconstructed as *šįāta- (e.g. in Korn 2005). This is erroneous: the Common Iranian form is also *čįāta-, as we can conclude among others from the Ossetic form cadæg 'quiet, slow' (Cheung 2003: 38).

First, a look at Khotanese and Tumshuqese may be useful. The Khotanese and Tumshuqese languages, which constitute, together with Wakhi, their own branch of the Iranian languages family, have the cognate form tsāta- 'rich', which also goes back to *čįāta- (cf. Bailey 1961: 54). Since the Khotanese-Wakhi-Tumshuqese branch is the first to have branched out, our notion of Common Iranian excludes these languages. Thus, the form tsāta- < *čįāta- does not directly help us for the reconstruction of the Common Iranian descendant of Proto-Iranian *čįāta-, but it shows with certainty that the Proto-Iranian form was not *šįāta- but indeed *čįāta-.

Ossetic, however, is included in Common Iranian, and the form cadæg 'quiet, slow' mentioned above thus shows that there was no Common Iranian sound change *čį- > *šį-. Furthermore, the Khwarezmian form hncy- 'to rest, relax', Sogdian ʾnčʾʾy 'to stop, to cease, to rest' and the Ossetic forms

---

[122] Avestan ś- originally derives from a palatalized *š (*šʲ), in the sequence of sound changes *čį- > *šį > OAv. šy- (‹šii-›) > YAv. ś-, and Old Persian š(i)y- simply shows *čį- > *šį, so that an intermediary step with a palatal *ś is not needed for the reconstruction of the word in Avestan.

Iron *æncajyn*/*æncad* Digor *æncajun*/*æncad* 'to rest', all deriving from **ham-či̯ā-*, show that **či̯* did not become **ši̯* in word-internal position in the ancestors of these languages (so Bailey 1961: 54).

If one accepts that PT **cy* or **cc* derives from Proto-Iranian **θi̯*, **di̯* by means of further palatalization in Tocharian of an already palatalized consonant in Old Steppe Iranian (see above), then I also believe it is more economical to assume that a palatalized sequence **či̯* would be borrowed as **ś*- in Proto-Tocharian. With the caveat that the evidence is circumstantial, I thus assume that the phonetic shape of the word was **č̓i̯āta-* in Old Steppe Iranian.

A form derived from Proto-Iranian **či̯āta-* was also borrowed into Proto-Permic **šud* 'luck', as well as into Khanti and Mansi (Katz 2003: 159; Holopainen 2019: 265f.). In every case, the forms suggest an original **š-* and no trace of a *yod*, which must have disappeared quite early in the source language. Nevertheless, as Holopainen (2019: 266) writes, "[i]n theory, Mansi *š* could also reflect earlier *č*, so the word can also be borrowed from an Iranian variety which preserved the affricate *č*." I thus wonder if Old Steppe Iranian or a closely related language could have been the source of the Mansi borrowing. The difference in meaning would not be an obstacle to this etymology, as the Tocharian meaning is due to a later evolution (see section 4.4.c.).

It seems difficult to establish any specific isogloss here, both from the Iranian and from the Tocharian side. It is possible to speculate on a form with a palatal **śāta-* in the Old Steppe Iranian language, but, although it fits the Avestan data, this would not be warranted by any other Iranian data, and would serve no other purpose than to explain the Tocharian form. An important point to note is that there is no parallel for **cy-* > **ś-* in Tocharian. This seems to suggest that, indeed, the Old Steppe Iranian word showed some degree of palatalization of its initial sound or sequence of sounds.

c. Old Steppe Iranian **aha* > **ā*

A very important feature of Old Steppe Iranian is the phonological change of **aha* (possibly **[æhæ]*) > **ā* (possibly **[ɑː]* or **[ɒː]*). This change is only seen in one Old Steppe Iranian word in Tocharian, TB *ainake* (Archaic TB *eynāke*, TA *enāk* 'evil, bad', but it is quite clear from it. Indeed, Old Iranian **a* was borrowed as Proto-Tocharian **e*, and Old Iranian **ā* was borrowed as Proto-Tocharian **a*. The two Old Iranian phonemes **a* and **ā* were dif-

ferent, not only quantitatively but also qualitatively according to the data at hand. There is thus very little place for the possibility of a sequence of two Old Iranian vowels *aha to be borrowed as *a in Tocharian, as it would imply the backing *within* Tocharian of two front vowels, to render a length that would be, in Tocharian, non-phonemic. In other words, it is not plausible that *e.e would have become *a in Tocharian, cf. TB *pernent*, the obl.sg.m. of TB *perne$_u$* 'glorious, worthy (one)', which derives from *perne.ent < *pernewent (and is not †*pernant*).

Once this fact is established, two main consequences can be proposed for our Old Iranian:

1. Despite the fact that those two vowels *a and *ā were certainly phonetically articulated differently (as they were rendered quite differently in Tocharian), it is possible that they were phonologically the short and long counterpart of each other.

2. It is possible, if not likely, that this Old Iranian language had no phoneme *h* in intervocalic position, at least clearly not in this specific position (between two *a*s).

Other Iranian languages have undergone the same change (*aha > *ā) at a relatively early stage. Sogdian is one of them, e.g. *sāk* 'number' < *sahaka- cf. Gharib, SD: s.v., but so is Ossetic, e.g. *wat* 'room' < *u̯aha̯θa-, cf. Cheung 2002: 24, and Khwarezmian too (*sʾk* 'hare' < *sahaka- cf. Benzing 1983: 567; *rāk* n.f. 'vein' < *rahakā-, cf. MacKenzie 1990: 121), Yidgha-Munǰī (*sīγ* 'hare' < *sahaka-, cf. Morgenstierne 1938: 57) and Pashto (*wéra* f. 'fear' < *wāryā < *au̯ahari̯ā-, cf. Cheung 2011: 192). Khotanese, Tumshuqese, Parthian, Balochi, Old Persian, Bactrian, all preserve the sequence *aha (e.g. MP *rahag* 'vein',[123] Parthian *rhq* 'vein', Bactrian πιδοοαυανο 'request' < *pati-u̯ahana-).

d. Loss of PIr. *h- in Old Steppe Iranian

Apart from the contraction of *aha to *ā, there is another fact which suggests that *h- was dropped in Old Steppe Iranian: the Tocharian B name of

---

[123] Note, however, MP *afsānag*, NP *afsāna* 'fable' of which one of the possible etymologies derives it from *abi-sahanaka- (see Gershevitch 1971: 274; cf. Hasandoust 2014: 242f.), or from *upa-sahanaka-, if it is not borrowed from Parthian.

India, *yentuke**, which shows the same initial vowel as TB *yetse*, TA *yats* 'skin', ultimately from PIr. **iĭa-* (Av. *izaēna-* 'made of skin'). There are Tocharian loanwords with *i-* in initial position: see for example TB *iṣcäke* 'clay'. Even if word-initial *i-* were impossible to pronounce for Proto-Tocharian speakers, why would they not have rendered it with *yə-*, which existed in the language (cf. TB verb *i-* 'to go' /yə/), rather than with *ye-*?

It is more plausible that initial *i-* in that Iranian language was rendered as **i̯a-*, **i̯i-* or a variant thereof, such as **i̯e-*, adapted in Proto-Tocharian as **ye-*. And, again, it is much more economical to suggest that the initial *h-* was dropped in that Iranian language, like in intervocalic position, because **hi-* in **hinduka-* has the same representation in TB *yentuke* as **i* in **idza-* 'skin' in TB *yetse*, TA *yats*. I assume that only after the loss of **h-* was initial *i-* (both primary and secondary) turned to **i̯a-* or similar.

For the loss of *h-*, one can cite Sogdian, Pashto, Ossetic, Khwarezmian and Yidgha-Munǰī, against Bactrian, Balochi, Avestan, Old Persian, Khotanese. Here we need to address two questions of relative chronology. First, the loss of *h-* in Sogdian and Khwarezmian is relatively old, since it precedes the apheresis of *a-* in both languages (**haC-*, **aC-* > *C-*).

Another argument suggests that **h* was lost in initial position: the apparent reconstruction of Old Steppe Iranian **u̯ai-* (see below). Indeed, in Sogdian, Ossetic and Khwarezmian, the loss of *h-* is later than the sound change **hŭ-* and **hu̯-* > *xu-*, *xw-*. Compare, for instance, Buddhist Sogdian *ʾγwšh* 'sister-in-law' < **hu̯asru-* (cf. Gershevitch 1961: 51), Ossetic Iron *x₀y*, Digor *xu* 'pig, boar' < **hū-* (cf. Cheung 2002: 16, 25), or Khwarezmian *xubisk* ‹xbsk› 'eigen' < **hu̯ai-paθi̯a-ka-* (cf. Benzing 1983: s.v.). Thus 1. **hŭ-* and **hu̯-* > **xw-* and then 2. **h-* > **ø-*.

The importance of this observation is due to the fact that, in Old Steppe Iranian, there was no apheresis nor was there any sound change **hu̯-* > **xu̯-*, as is seen in Tocharian B *waipecce* (from **xu̯aipaθi̯a-* one expects TB ****kwaipecce*, cf. TA *kump* 'bowl' ← Old Steppe Iranian **xumba-*, unless this is an Indic loanword).

This seems to imply that, in this Old Iranian language, *h-* was dropped first before any type of vowel, unlike what happened in Sogdian, Ossetic and Khwarezmian. This lowers the chances that this change occurred as part of a single areal phenomenon, and brings to zero the chances of this change having occurred as part of a common innovation of these languages.

In Pashto, *$h\underset{.}{u}$- became *xw- (e.g. *$h\underset{.}{u}ásru\underset{.}{i}ā$- 'mother-in-law' > Pashto $xwā́x̌e$), but *hu- seems to have become *u- (*$hufta$- 'asleep' > Eastern Pashto $ud\acute{\partial}$, Waziri $wewd$ 'id.'). In any case, the change *$h\underset{.}{u}$- > *xw- of Pashto excludes it here as well. Because this change also occurred in Yidgha-Munǰī (Morgenstierne 1938: 68), these two should be excluded as well.

The Persic group represents a difficulty here: we have, on the one hand, Middle and New Persian which show very clearly *$h\underset{.}{u}$- > *xw- ($xwad$ 'self', for example) and on the other hand, there is documented Old Persian, which did not undergo this sound change: rather, in Old Persian *$h\underset{.}{u}$- > *huw- > uw- (cf. Hübschmann 1895: 217f.; Korn 2021: 100f.).[124]

To note, Proto-Iranian final *-ah became *-i in Ossetic (cf. Cheung 2002: 56f.; Kim 2003: 57f.), which very likely indicates that *-h was lost later in Ossetic. In Old Steppe Iranian, there is no trace of such a change, with, for instance, PT *$perne$ and *$entse$ pointing to Old Steppe Iranian *$a(h)$ as a reflex of Proto-Iranian *-ah. On the basis of the other elements here, it is possible to suggest that the Old Steppe Iranian ending was simply *-a, without final *-h, as that had disappeared in all other positions, as far as we can tell.

e. Old Steppe Iranian reflexes of PIIr. *j and *$ć$

A useful feature to look at is the result of PIIr. *j and *$ć$ in Old Steppe Iranian, that is, the Proto-Indo-European palatovelars. Both phonemes are rendered as *ts in Proto-Tocharian, and as ts in Tocharian A and B, in the words TB $etswe$ 'mule', $ñyātse$, TA $ñātse$ 'danger', TB $entse$, TA $emts$ 'envy, greed', TB $yetse$, TA $yats$ 'skin', TB $tsain$ 'arrow', TB $tsaiññe$ 'ornament', and, if my etymological proposals are accepted, also in TB $wertsiya$ 'community, assembly', TB $t^s\partial w$-, TA $t^säw$- 'to adhere, to fit, to put together', TB $m\partial tstsa$- and TA $nätswā$- 'to die of hunger, to starve', TB $tsetke$ 'centaine (?)', TB $tsere$ 'a measure of liquid volume of half a $lwāke$ (?), cup (?)' and TB $tseriteke$ 'young grass (?), green animal (?), turtle shell (?)'.

The reconstruction with *t^s and *$d^{\bar{z}}$ was made by scholars of Iranian since a few decades (e.g. Klingenschmitt 1975; Mayrhofer 1983; 1989: 6f.). Ever

---

[124] In her article, Korn argues that the loss of *h- in documented Old Persian is one of the features that exclude the possibility that it is the direct parent of Middle and New Persian.

since it was noted that the earliest Iranian loanwords in Tocharian show *ts* for PIIr. *j and *$ć$ (Schmidt 1985), we actually have under our eyes the re-flexes of these Old Iranian sounds, which were until then only reconstructed.

It has been postulated for multiple reasons that Proto-Iranian *$ć$ and *j did not become *t^s and *$d^ƀ$ in all positions. The question is too thorny to be discussed here in its entirety. However, the Khotanese-Tumshuqese-Wakhi branch preserved a palatal pronunciation of *$ć$ and *j in front of *$u̯$, that is, this branch never had *tsw (Sims-Williams 1998: 136; 2017: 264; Peyrot 2018a).[125] As Peyrot (2018a) demonstrates, the Tocharian B word *etswe* 'mule' derives from Proto-Iranian *$aću̯a$- 'horse' (< PIE *$h_1ék̂uo$-) through a borrowing from Old Iranian *$at^su̯a$-.

As Peyrot (2018a: 271f.) argues, the Khotanese-Tumshuqese-Wakhi branch is excluded for Old Steppe Iranian because of the outcome *$-tsu̯$- in Old Steppe Iranian of Proto-Iranian *$-ću̯$-. The Khotanese word for 'horse' is *aśśa*-, which goes back directly to *$aću̯a$- with preservation of the palatal and assimilation of the glide.

Some scholars do not see Khot. -*śś*- as going back to *$-ću̯$- directly, but as ultimately going back to *$-tsu̯$-, for instance Kümmel (2007: 234) and Novák (2013: 121-22[186]).[126] However, new considerations about TB *eñcuwo*, TA *añcu** 'iron', which was borrowed from a preform of Khotanese *hīśśana*- 'iron' (s.v. *eñcuwo*; cf. also Peyrot, Dragoni & Bernard 2023), show that the sequence *$-nću̯$- in Pre-Khotanese or Proto-Khotanese-Tumshuqese became a sequence of nasal + palatal, which must have been a palatal sound as *$-ć$- in *$-nću̯$- did not merge with the result of PIr. *$-č$-. This discovery thus in-validates the hypotheses of Kümmel, Novák and Lipp (apud Novák), and confirms the exclusion of the Khotanese-Tumshuqese-Wakhi branch as a close relative of the Old Steppe Iranian language, as already argued in Peyrot (2018a).

Thus, the Old Steppe Iranian language cannot be closely related to Kho-tanese, Tumshuqese and Wakhi, but is phylogenetically closer to the rest of Iranian languages.

---

[125] We cannot, at this point, be sure that Khotanese-Tumshuqese *s*, *z* derive from an intermediary stage *t^s, *$d^ƀ$ (respectively) and not directly from a palatal phoneme as in *j > *$ź$ > *z* instead of *j > *$d^ƀ$ > *z*, for instance.

[126] Including a suggestion by Reiner Lipp.

It is still unclear how Old Persian, and the Persic group in general,[127] which show *s for *ću, arrived at that result (cf. for instance Kümmel 2007: 234). In any case, the Persic group did not preserve *ću as such and cannot be excluded from the comparanda on the basis of this feature.

f. Proto-Iranian *i and *u in Old Steppe Iranian

A thorny question concerning Old Steppe Iranian is that of the treatment of the short vowels *i and *u. As is known, PIE short *i and *u were changed to schwa in Proto-Tocharian (see for instance Ringe 1996: 133f.). New *i and *u phonemes arose from PT *əy and *əw. However, Tocharian speakers seem to have had continuous problems with the pronunciation of /i/ and /u/, as can be seen in, for instance, TB pañäkte 'Buddha' with pa- from *pät 'Buddha', itself a borrowing from *bud or *but (perhaps from Sogdian /but/) compared with the poetic variant pūdñäkte, thus with "learned" pūd-. One can also look at the variation between kuñcit and kwäñcit 'sesame' (cf. Bernard 2020: 52f.), and sakw 'luck' ← Skt. sukha- 'happiness'.

In our list of Old Iranian loanwords in Tocharian, there are a number of words which had a short *i or *u, in several positions. In initial position there are TB yetse, TA yats 'skin' going back to a form borrowed from the Old Steppe Iranian reflex of Common Iranian *idƷa- in Proto-Iranian, and TB yentuke 'India' which goes back to an Old Iranian form *induka- < *hinduka- 'India' (for the loss of the initial /h/, see above). There was thus a sound change *i- to *i̯V- (*i̯a- or *i̯i(/e)- ?) in Old Steppe Iranian.

The change of *i- to *i̯V- is not entirely uncommon in Iranian, but it is not found in Khotanese, Pashto, Ossetic, nor in Old Persian or Bactrian. There is definitely an alternation, in writing, between i- and yV- in Sogdian and possibly in Khwarezmian, for instance in the Sogdian variants ʾync ~ ynch 'woman' (SD: s.v.), and in the Khwarezmian variants ʾyδʾβrc ~ yδʾβrc 'Zwischenräume habend' (Benzing 1983: s.v.).

However, Yidgha and Munǰī do show traces of an ancient development of *i to *y- in initial position (Morgenstierne 1938: 95), including after the loss of pre-vocalic initial *h, as in Munǰī yumar 'to count' < *hišmar- (vs.

---

[127] That is, all Iranian languages that show a number of common innovations, such as *θ and *z for *ć and *j in a number of positions, and *s in pre-labial context.

Yidgha *imar-*); *yimsāl* 'this year' < **ima-sarda-* (perhaps a loanword from Persian?), cf. Morgenstierne (1938: 95).

The Yidgha word *Idγ* 'Yidgha' itself shows a secondary **yV- > i-* (as in *imar-* as well?), cf. the name of the language in both Yidgha and Munǰī, *Yedγǎ* (cf. Morgenstierne 1930a: 443), itself from **yenduka-*. The Yidghas are called *Inǰigān* in both Khowar and Persian (cf. Morgenstierne 1930a: 443).

In this sense, I consider the Yidgha-Munǰī group closer to this evolution of Old Iranian. Possibly this evolution can be connected to the constant first-syllable stress of this Old Iranian language. If **i* had a tendency to be reduced to schwa, and the first-syllable was accented, then initial *i-* would probably change to another sound, but possibly not to initial schwa. We can speculate that word-initial schwa was not allowed in Old Steppe Iranian, as in many languages, including Tocharian. We would, in a parallel fashion, expect **u̯a-* or **u̯e-* for initial **u-*. One can think, for instance, of TB *wertsiya* 'community, assembly', if ultimately going back to the zero-grade form **u̯r̥ďii̯ā-* (*-iya* might also have been added in Tocharian itself to a base **u̯r̥ďā-*). I have also suggested that the *we-* in TB *weretem** 'promise, contract' might possibly go back to an initial **u-* or **u̯-* (see p. 72f.).

A Proto-Iranian **i* has been reduced to naught in the two following cases: TB *perpente* 'burden, load' from **pari-banda-* 'attached around' (but with a much more specific meaning attested already in Iranian), and, if my etymology is accepted, in *eprete* 'resolute, steadfast' from **abi-ratu-* 'according to the resolution'.

It has become *yod* in *ñyātse* 'danger' from **ni-āďah-* 'distress', *-māññe* 'pavilion' from **māni̯a-* 'of the house'.

It may have been reduced to schwa, as shown by **abi-st(H)-i̯a-* → TB *epastye*, if that derivation is accepted.

It has possibly been maintained as **i* only in front of *yod*. However, since **-əyV-* would also yield *-iy-* in Tocharian, these examples are not probative. The examples are: TB *epiyac* (TA *opyāc*) 'memory' ← **abi-i̯āt*… 'idem', *mañiye* 'servant' (← **mānii̯a-* '(servant) of the house') *newiya* 'canal' ← **nau̯ii̯ā-* 'idem', and TB *wertsiya* 'community, assembly' ← **u̯r̥ďii̯ā-*. The word *ekṣineke** is no counterexample, as the *-i-* in it goes back to **-ii̯-* (cf. section 2.6.2.h), which was simplified either in Old Steppe Iranian itself, or in (Proto-)Tocharian.

 There are no traces of word-final **i* in our loanwords.

The TB word *tseriteke*, whatever its meaning (I argue, speculatively, for 'young grass', 'green animal' or 'turtle shell', s.v.) does have an Iranian appearance. It would correspond, if the etymology I developed together with Federico Dragoni is correct, to Ossetic *zældæ* 'young turf', with *-ka-* suffix. I have no explanation for the preservation of the $*i$ here. As a rule, we must assume that Old Steppe Iranian $*-i-$ has become schwa in Proto-Tocharian.

An early $*u$ has been reduced to naught in word-final position: TB *peret*, TA *parat* 'axe' from Old Iranian **paratu-* 'idem' and TB *tsain* 'arrow' from Old Steppe Iranian **ɟ̌ainu-* (the $*-u-$ is still visible in the plural TB *tsainwa*, see below). Possibly it is also lost in *mot* 'wine', if from Old Steppe Iranian **madu* 'mead, wine' and if the original $*-u$ ending was not lost in the umlaut process. Compare further perhaps TB *eprete* 'steadfast' (< **epret* + adjectival *-e*) from Old Iranian **abi-ratu* 'according to the resolution' (> 'steadfast'). For the possible preservation of the final $*-u$ when the words were borrowed in Tocharian, see below.

The phoneme $*u$ has been maintained in contact with consonantal clusters in Tocharian B *yentuke* 'India' (← OSIr. **i̯anduka-* or **i̯enduka-*) and perhaps in TA *kump* 'bowl' (from OIr. **xumba-*), if not from Sanskrit or Middle Indic. This preservation of $*-u-$ is unexpected. In the case of *wertsiya* 'community, assembly', it is conceivable that the initial *we-* derives from $*u-$. In that case initial $*u-$ had become $*u̯e-$ or $*u̯a-$ in that Old Iranian language, **urí̯i̯ā-* or **u̯rí̯i̯ā-* > **u̯ardzii̯a-*.

We do not know whether the reduction of $*i$ and $*u$ to schwa or the loss thereof is Tocharian-internal or already Old Iranian, but there are strong elements in favor of the loss having occurred within Tocharian. One can depart from the plural of *tsain* 'weapon' (← **ɟ̌ainu-*), which is *tsainwa* and seems to indicate that Tocharians borrowed **tsainu*, made a plural **tsainwa*, morphologically originally |tsainu-a|, and only then lost the $*-u$ (cf. Peyrot 2018a: 270).[128] One can also add the adjective *tsainwaṣṣe* as a further argument. A similar argument based on the verb **nətswa-* 'to starve' can be adduced: indeed, this verb implies a Proto-Tocharian noun **netsu-* 'corpse'. In both cases the preservation of the $*-u-$ is due to a following vowel. These elements would suggest that the loss of *-u* occurred within Proto-Tocharian.

---

[128] However, *a*-plurals are rare in Tocharian, and it is also possible that Tocharian speakers chose to make a *-wa* plural, which is more productive, from *tsain*. This would not explain the adjective *tsainwaṣṣe*.

However, this does not preclude that the reduction of short *i and *u, at least in other positions, is an Iranian phenomenon.

As I wrote above, the fact that in Old Steppe Iranian accented word-initial short *i had turned into something else (s.v. TB *yentuke*, TB *yetse* TA *yats*) is possibly connected to a general reduction of *i. Reductions of *i and *u are not observed in the generally phonologically more conservative Old Persian and Balochi groups.[129] Neither does Ossetic reduce short *i and *u: on the contrary, they merged with *ī and *ū (Cheung 2002: 15-16).[130] Reductions of *i and *u are, however, common to multiple languages spoken to the East.

In Sogdian, for instance, PIr. *i, *u and *a, when they are unaccented, have been reduced to schwa (Yoshida 2009: 285; Novák 2013: 124), namely Gershevitch's "indistinct vowel" (Gershevitch 1961: 15-16). In Khwarezmian as well, there is a "reduction of unstressed short vowels" (MacKenzie 1990: 94). This phenomenon is also common to Yidgha and Munǰī for *i and *u (cf. Morgenstierne 1938: 95-96). To some degree *i and *u were reduced in Bactrian as well, but not *a (cf. Gholami 2014: 58f.). I believe that the reduction that took place in Pashto and, more particularly, the conditions for the preservation of *u are of particular interest here.

In Wanǰī, an extinct Modern Iranian language or Tajikistan, of which we only have very scarce documentation, it seems that both short *i and *u merged into a single phoneme, which was reflected as either *i* or *ə*, but was probably /ə/. After this reduction, long *ī and long *ū became /i/ and /u/ (cf. Novák 2013: 36), as there was no longer a length distinction in that language. Theoretically it could fit with the Tocharian data to some extent, but there is no way to know when any of these changes occurred.

In contrast, in Pashto, *ī and *ū were shortened, and thus merged with *i and *u (Cheung 2011: 199). The "new" vowel *u was lost in unstressed open syllable, as in *bar* 'top, summit' < *upára- (in detail in Cheung 2011: 199, also for the following). It became *ə́ in stressed open syllables, for example in *ɣanə́m*, Wanetsi *ɣandə́m* 'wheat' < *gantúma-. So did *u < *-am, as in *zə* 'I' (< *azám < *ajám).

---

[129] This is, of course, a generalization. There are positions where *i and *u were reduced, of old, in Balochi and also in languages related to Old Persian. This is nevertheless not a general tendency, but only relates to specific sound laws.

[130] In Ossetic, *i̯a- becomes *i- (cf. Cheung 2002: 18).

However, in a closed syllable, *u was preserved: *búr* 'having lost a son, childless' < **apúθra-*, *gúta* f., Wanetsi *nəgút* 'finger' < **angúštā-*, etc. This situation can be somewhat compared to the Tocharian one, where TA *kump* 'bowl' ← **xumba-* (if from OSIr.) and *yentuke* 'India' ← **i̯anduka-* show preservation of *u either in a closed syllable or after a closed syllable. The same rules also apply to ancient *i (< *i and *ī), see Cheung (2011: 199f).

In view of this, I would now say that *i was preserved in front of *yod* (**i̯i̯a-* preserved as a group), for instance in TB *mañiye* 'servant', but the fact remains that schwa in Tocharian automatically became *i* in front of *yod*, which means we cannot be sure of the preservation of Old Steppe Iranian *i in front of *yod*. As to the Tocharian B word *epastye* < **abí-st(H)-i̯i̯a-*, it shows a reduced *i (of **abi-*) in a closed syllable. This *i carried the stress, and it is thus normal that it did not disappear giving a form †*epstye*, or, if the reduction occurred in Old Steppe Iranian, ***abst(H)i̯i̯a-*. As stated earlier, the preservation of the *-i-* in *tseriteke* is definitely problematic. Due to syncope rules, one might expect rather †*tseritke* where the *-i-* would be preserved in a closed syllable, perhaps, but this is in any case not what we have.[131]

It should be underlined that, although *a was reduced in Sogdian and possibly in Khwarezmian, in unstressed positions, neither in Yidgha-Munǰī nor in Pashto was the treatment of *a identical to that of *i and *u.

It seems likely that the general tendency in Iranian languages spoken to the East to reduce their vowels, at least *i and *u, was an areal phenomenon. Possibly, more specific changes occurred in parts of that area, and possibly, these changes occurred at different points of time.

I would like to suggest here that the reduction of *i and *u which we find in loanwords in Tocharian, even in later ones, was also due to the influence

---

[131] An *ad hoc* solution to account for the preservation of *i and *u in *tseriteke* and *yentuke* is by supposing that the words where pronounced with voiced intervocalic *d (**dzaridaga) and *g (**i̯anduga). Then, we should suppose that the *-u- and *-i- in contact with a *d or *g were lengthened, as they were at a very late stage of Old Persian or a very early stage of Middle Persian, as per Korn's law (Korn 2009). Since an ancient *rd had become *ld in Old Steppe Iranian, it is probable that *rt had then become *rd, implying that voicing of ancient unvoiced stops in this language may possibly have occurred. Since there is no other reason to assume this solution apart from the fact that it occurred in Middle Persian, I view it as a very doubtful one.

of this areal phenomenon. For the discussion of the features of the Old Iranian language, it is relevant to note that we cannot know for sure whether every point discussed here pertains to Old Iranian or to Tocharian. It could be a Tocharian development, although *yentuke* is definitely a problem. However, it does seem to me that the Old Iranian language itself underwent some reductions of *i* and *u* in some contexts, including when stressed, as seen in the shift of initial *i* and *u* to *i̯V-* and *u̯V-*.

g. The Old Steppe Iranian accent as reflected in Tocharian loanwords

I have suggested at multiple points throughout the thesis that this Old Iranian language had fixed first-syllable stress, with morphological exceptions. It is now the place to examine the evidence in its totality. The stress can be seen in the following secure loanwords:

Tocharian B: *śāte*, *ākteke*, *ainake*, *ñyātse*, *pāke*, *epiyac*, *mañiye*, *-māññe*

The stress in *epiyac* and *mañiye* can be seen from the fact that the *-i-* was preserved, and the second syllable must therefore have been stressed. As far as *epiyac* is concerned, this fits the facts because *abí-* probably carried a fixed accent on the *-i-*, as it did in Indic. Unstressed *-i̯a-* secondarily becomes *-ya-* in Tocharian, as in *epastye* and *-māññe* (for more on this, see below).

   Due to syncope rules, where in a series of three identical vowels a non-accented vowel was reduced, creating a consonant cluster (cf. Winter 1994), the stress can be deduced in TB: *ākteke* ← *ā́gataka-*, *retke* ← *rátaka-*, *speltke* ← *spáldaka-*, *welke* ←*u̯áldaka-* and TA: *nātäk* < *natke* ← *nátaka-* or *ná́θaka-* (although in the latter instance the first vowel was long).[132]

   Where the stress can be deduced from syncope rules, it is always on the first syllable. In the word *kenek* 'cotton cloth', which needs to be a BMAC word because there is no suitable Iranian etymon, the stress was probably on the middle syllable, thus *kenéke > kenek*, with apocope rather than syncope,

---

[132] Tocharian B *perpente* 'burden, load' does not show the expected syncope: this may be because a regular reduction would have led to a form *perpnte*, with a cluster *-rpnt-* which was certainly too difficult to pronounce.

due to the middle syllable stress. If the word had been of Old Steppe Iranian origin, we would have had *kéneke > TB †keṅke. Where the stress can be seen, it is constantly on the first syllable, with two exceptions: 1. it is on the second syllable of the prefix *abi-, 2. it is fixed on the -i- or the -ə- in the adjectival suffix *-ii̯a- (known from comparative Indo-Iranian data to have been accented *-íi̯a-). In -māññe 'pavilion', it shifted from the *-íi̯a- suffix to the preceding syllable (showing thus that unaccented ñíye > ññe within Tocharian). This shift is not surprising, since -māññe, the attested spelling, thus /máññe/, is a second member of compounds, and it is the last syllable of the first member of compounds that carries the accent in Tocharian B. As to 1., the prefix *abi- was originally stressed on the second syllable, which fits with the rest of the Indo-Iranian data, for instance with Vedic abhí-, and explains why we have TB epiyac and TB epastye (and not †epyac and †epstye or rather †epästiye).

We thus have the portrait of a language with fixed lexical stress, always on the first syllable, but where the stress could shift due to already stressed suffixes and prefixes. Resembling this portrait is Proto-Ossetic, which also had fixed initial stress (Cheung 2002: 123), as discussed below.

Possibly, *i- in iścem 'clay' (< OSIr. *išti̯ám) did not become *i̯a (or *i̯e) as expected specifically because it was not stressed, as it was originally a neuter noun ending in -ám (or by influence of iṣcäke, an earlier borrowing with a similar meaning). In any case, the reflex of stressed initial *i- can be seen in yentuke (< *(h)índuka-) and yetse (< *ídⱬa-), see section 2.6.2.d. This hypothesis has not been tested to the full yet due to the scarcity of data, and it remains to be confirmed or disproved.

Sogdian stress is not archaic, and is mostly the result of inner-Sogdian innovation (the so-called "rhythmic law"), which was not even shared by its closest sister-language, Yaghnobi (cf. Novák 2013: 80). It was initially free, and had to be learned for each word or word-class, but then a new system emerged, where stress was fixed on the first long vowel or diphthong, and otherwise on the last syllable of the word (Sims-Williams 1989: 181f.; Novák 2013: 77f.).

Pashto, Old Persian and Balochi all have their own stress rules. As Niels Schoubben informs me, Bactrian stress seems to have been on the first heavy syllable from the last one, as can be seen from syncope rules. Fixed initial accent is excluded for Bactrian: cf. σαζαμανο 'as soon as' < *čii̯át-ǰamā́na- (etymologically 'whenever time'). However, Bactrian had initial accent in

light-syllable words. In the prefix *abi-, the accent was even on the first syllable, unlike in Old Steppe Iranian: αβδδινο 'custom' < *ábi-daiana-.

Morgenstierne (1938: 103) has shown that a great number of Yidgha-Munǰī words go back to preforms with initial stress, similar to Old Steppe Iranian. However, he posits a different position of the stress for some words, like *dram* 'inside' < *antaráhmi, and the situation remains to be clarified.

It thus seems that only Proto-Ossetic has been reconstructed with fixed initial stress. A reason for this fixed initial stress could perhaps be the Uralic tribes to the north, which also had initial stress, or some other substratum/superstratum. Possibly, initial stress is a marked feature of "Steppe Iranian".

My conclusions concerning fixed word-initial accent with exceptions due to prefixes or suffixes seem to contradict Kuryłowicz's conclusions that the accent was fixed on the penultimate syllable in all of Old Iranian (cf. Kuryłowicz 1964; 1975). He was basing himself on Old Persian and Avestan spellings. However, his conclusions were not accepted, as it has been shown that, for instance, Avestan preserved an old stress system (cf. Beekes 1988). One of his main arguments is that the phonological opposition between *a* and *ā* has been neutralized in Old Persian and in Avestan (cf. 1975: 500f.). With Old Steppe Iranian, we have now the example of a third Old Iranian language, and this language clearly preserves the opposition *a* : *ā* of Proto-Iranian (e.g. *maiyya*, *newiya* vs. *perne*, *perpente*), but does not have a fixed stress on the penultimate syllable.

h. Old Steppe Iranian and the name of the 'dove'

Since we know the stress to have been on the first syllable, we can propose that unaccented *ai̯ became *ī in front of *n in order to explain TB *ekṣinek* (possibly *ekṣineke*) 'dove' from OIr. *axšai̯naka- (a -ka- derivative of PIr. *axšai̯na- 'dark blue'). I posit the additional condition of unaccented *ai̯ because we also have *ainake* 'evil' (and not †*inake*) and *tsain* 'arrow', which have both preserved *ai̯ before *n, possibly due to the fact that *ai̯ was accented in these words. In Ossetic, *ai̯ became *ī in front of *n (through *ē̆?) before *ai̯ became *ē in other positions (Cheung 2002: 17). A notable example is *æxsin* (Digor) 'dark blue', a cognate of *axšai̯naka- 'dove'. Nevertheless, in Ossetic this rule was not regulated by the position of the stress. Although it is hard to reject the possibility of unaccented *-ai̯- becoming *-i- in

front of *-*n*- in Old Steppe Iranian, it is also hard to accept it on the basis of one single example, and I thus indicate it with much caution.

A different solution to the problem can be suggested. It differs from the posited sound change of *$a\underset{.}{i}$ to *$\bar{\imath}$ in front of *n mentioned above. Kim (2003: 51[20]) proposed a sound law Pre-Proto-Ossetic *-$a\underset{.}{i}n\underset{.}{i}a$- > Proto-Ossetic *$iyna$. The examples he presents for it are: *abi-$\check{s}ai\underset{.}{}\theta n\underset{.}{i}\bar{a}$ > POss. *$æfsijnæ$ > Digoron $æfsijnæ$, Iron $æfsin$ 'landlady', *$x\check{s}ai\underset{.}{}\theta n\underset{.}{i}\bar{a}$- > POss. *$(æ)xsijnæ$ > Digoron $æxsijnæ$, Iron $æxsin$ 'lady, mistress', *$ax\check{s}ai\underset{.}{n}\underset{.}{i}aka$- > POss. *$æxsijnæg$ > Digoron $æxsijnæg$, Iron $æxsinæg$ 'wild dove, pigeon' and *$kai\underset{.}{n}\underset{.}{i}\bar{a}$- > POss. *$kijnæ$ > Digoron $kijnæg$ 'reward, revenge'. I recently discussed and added an example to this sound law, which I have named Kim's law, in Bernard (2022: 32f.).

Kim's law is a very interesting phenomenon, especially since it could also explain the Tocharian B form, if the sound change had occurred in Old Steppe Iranian as well. From the Ossetic point of view this sound law would explain why a so-called long $\bar{\imath}$, contrasting with i, occurs in Ossetic only in front of *n (cf. Kim 2003: 51[20]). Furthermore, as Kim (op. cit.) points out, Digoron [$\bar{\imath}$] is phonemically /ij/, which supports Kim's suggested sound law.[133]

We are here facing a sound law that operated in both Ossetic and Old Steppe Iranian, namely *$Vn\underset{.}{i}V$ > *$i\underset{.}{i}nV$ (cf. Bernard 2022: 34). Technically, one can suppose that first a vocalic palatalization took place: *$a\underset{.}{i}n\underset{.}{i}V$ > *$in\underset{.}{i}V$, followed by a metathesis, *$in\underset{.}{i}V$ > *$i\underset{.}{i}nV$. Another possibility, which I mentioned in Bernard (2022: 34) is that "there was first palatalization of *n, subsequent depalatalization, and palatalization of the preceding vowel", by which I meant a process such as *$a\tilde{n}$ > *$i\tilde{n}$ > *iyn. No matter the phonetic detail of the intermediary phases, it would be uneconomical to assume two

---

[133] In Bernard (2019), I proposed that New Persian *yak* 'one' went back to Middle Persian *yak* < *$\bar{\imath}yak$, itself ultimately from *$a\underset{.}{i}\underset{.}{u}a\underset{.}{i}aka$-, analogical form based on either *$\theta ra\underset{.}{i}ah$- 'three', or on *$d\underset{.}{u}a\underset{.}{i}a$- 'double' with the original meaning of *$a\underset{.}{i}\underset{.}{u}a\underset{.}{i}aka$- being *'unique', cf. Balochi *ēyōk* 'alone' < *$a\underset{.}{i}\underset{.}{u}a\underset{.}{i}a$- + suffix -ōk, cf. Bernard (2019: 52). To explain the initial *i- of MP *$iyak$, I had some difficulty, and resorted to a dissimilation of the two *-$a\underset{.}{i}$-, as suggested to me by Johnny Cheung, which I still believe is possible. However, if the Ossetic sound law proposed by Kim is correct, we could see here a parallel sound law operating: *$a\underset{.}{i}\underset{.}{u}a\underset{.}{i}aka$- > *$a\underset{.}{i}\underset{.}{u}\underset{.}{i}aka$- > *$\bar{\imath}yak$. This idea is further developed in Bernard (2022).

identical sound laws in both Ossetic and Old Steppe Iranian, given how uncommon this sound law is. Thus, it is an extremely important piece of evidence for the prehistory of these languages, as we have to assume that this sound law happened in the ancestor language of those two languages (since Old Steppe Iranian cannot be the ancestor of Ossetic).

This sound law provides information of various sorts: it must have occurred in the ancestor language of Old Steppe Iranian and Ossetic; before the monophthongization of Proto-Iranian diphthongs (since Old Steppe Iranian had undergone this change before monophthongization); and possibly, I assume, before the shift of both languages to word-initial stress. Indeed, this sound law should rather have operated on a stress syllable, as is more likely typologically, thus *axšá͜in̩i̯aka- > *áxšii̯naka (→ PT *ekṣineke). Since, as far as I could find, no other Iranian language shares this exact sound law, we can safely assume that Old Steppe Iranian and Ossetic were closer to each other than to any other (documented) Iranian language, as other data also suggests (see table 2.6.2.m for a summary).

i. Shortening of *ā before *-u̯ā- and *-(i̯)i̯ā- in Old Steppe Iranian

Another notable innovation of Old Steppe Iranian seems to be the shortening of *ā before *-u̯- and *-i̯-. There are two examples: TB *newiya* 'canal' (< PIr. *nāu̯i̯a- perhaps '(belonging) to boats, of navigation'), and Archaic TB *meyyā*, *maiyya* 'power, might, strength'.

Tocharian B *maiyya* has been borrowed from OSIr. *maii̯ā-, ultimately going back to Proto-Indo-Iranian *māi̯ā- 'magic power', a BMAC loanword (cf. Lubotsky 2001: 314). Originally the word was perhaps accented on the last syllable, as in Vedic *māyá̄-*. In Old Steppe Iranian it was likely accented on the first syllable; see for example the form *maiyyane*, /máiyyane/ 'in the power' found in multiple Classical Tocharian B texts, for example in the following text:

PK AS 17B a2
    cwī maiyyane klye(ma)ne a(kā)lkäntaṃts ta(r)n(ene)
    "in his power standing on top of the wishes" (translation from CEToM).

Naturally, a Tocharian B word cannot be accented on the final syllable (see p. 18), so the accentuation of *maii̯ā- on the first syllable is mostly based on the other Old Steppe Iranian examples showing first syllable accentuation in

unsuffixed, unprefixed words. The Archaic TB form *meyyā* confirms the antiquity of the (Old Steppe Iranian) shortening of long **ā* in this word.[134]

Avestan might show a similar, if not identical sound law: cf. for these same words Avestan *nauuāza-* 'Schiffer', Young Avestan *maiiā-* 'witchcraft power' (vs. OAv. *māiiā-*). In Avestan this is the effect of the sound change **-āu̯ă̆- > *-au̯ă̆-* and **-āi̯ă̆- > *-ai̯ă̆-*, cf. Hoffmann & Forssman (2004: 58). A similar sound law also operated in Sogdian and in Khotanese (cf. Sogdian *nwʾʾz* 'sailor', but also *nʾwzy* 'id.'); cf. for example Gershevitch (1961: 17). It is to be noted that de Vaan (2003: 118ff.) rejects this sound law for Avestan, and rather considers case by case explanations for words which show it, such as *nauuāza-* which he explains as a dissimilation **ā_ā > a_ā* (2003: 124).

Szemerényi (1951: 159) proposed that this Avestan sound law is correlated to stress, that is, the shortening only applied when the syllable was in pretonic position. I am not able to judge the pertinence of this observation for Avestan (strongly doubted by de Vaan 2003: 118), but this would work if the Proto-Iranian stress was the same as in Sanskrit (cf. Vedic *mayā́-*), and remained there long enough for the shortening to occur for the word.

There is no Ossetic example or counter-example of the sound change **āu̯ă̆- > *-au̯ă̆-* (see Cheung 2002: 17), but **-āi̯- > *-ai̯-* is recorded (cf. Cheung 2002: 15). It is difficult to know whether either happened in Yidgha-Munǰī, as many shortenings of **ā* occurred in Yidgha-Munǰī, and the historical phonology of this branch is not sufficiently known. It did not occur in Old Persian, nor did it in Balochi, and as to Pashto, **ā* is only preserved as such in unstressed closed syllables (Cheung 2011: 178), so that it there is no way to know if this law operated here.

Since the sound law as known from Avestan, Sogdian and Khotanese is **-āu̯ă̆-* to *-au̯ă̆-*, and here we have **-āu̯íi̯- > *-au̯íi̯-*, we have to assume that 1. the sound law applied to **nāu̯ā- > *nau̯ā-*, only after which we had **nau̯íi̯a-* (analogically of **nau̯ā-* and possibly before the fixed initial stress pattern was introduced), or 2. that the sound change was extended to or only

---

[134] An important point to note is that the *-aiyy-* sequence in Classical and Late Tocharian corresponds to *-ey-* in Archaic Tocharian B, the only other constant example thereof being Archaic TB *eynake* 'evil, bad', corresponding to Classical TB *ainake*. This suggests that the Old Steppe Iranian diphthong **ai̯* was not borrowed as a diphthong **ay* in Proto-Tocharian.

concerned the sequence *-āu̯íi̯(a)-. If the Old Steppe Iranian shortening is shared with other branches, then we can imagine that it was shared at an earlier level, and thus defines a greater branch. However, it is difficult to believe it was, as the detail of this shortening varies among the different languages mentioned.

In any case, this sound change or variants of it occurred among Iranian languages spoken to the east, and it certainly concerned Old Steppe Iranian. Since it is not a trivial change, it is more logical to suppose that it occurred as part of a continuum, or of an areal feature, rather than assuming parallel developments. In that case, it is easier to explain the fact that the sound law is not identical in all these languages.

j. Simplification of the cluster *dm- in Old Steppe Iranian

Another possible sound change, only seen in one word is the apparent simplification of Proto-Iranian *dm- to *m- in TB -māññe 'pavilion'; TB mañiye (formally the same Old Steppe Iranian word, with different accentuation, see section 2.6.2.g). The word goes back to the PIE root *dem- 'house' (Latin domus, Greek δόμος), and can be reconstructed for Proto-Iranian as *dmāna- (OAv. dəmāna-). The initial cluster *dm- was reduced to m- in several Iranian languages: Old Persian māna- 'house' (MP, NP mān 'house, mansion');[135] Parthian mān 'house, mansion'. It was changed to nm- in Young Avestan nmāna, from Old Avestan dəmāna- (/dmāna/).

There are three possibilities to explain the initial m- in the Tocharian words. The most straightforward consists in suggesting a simplification *dm- > m- in the Old Steppe Iranian source, yielding *māníi̯a-. The second one consists in suggesting a sound change *dm- > *nmāníi̯a- → PT *nmāníi̯a- which could have been simplified to *mañiye in Pre-B. Both are theoretically possible, and unfortunately we do not have a Tocharian A cognate to be sure of it, but possibly the first hypothesis should be favoured, as it is much

---

[135] It has been claimed since long that *dm- changed to b- (through *db-) in Middle Persian, hence bānūg 'lady' < *bān + -ūg; cf. also bānbišn 'queen' < *dmāna-paθnī- (see Benveniste 1966: 29), see also Persian bām, bān, Pahlavi bān 'Dach, Haus' (cf. Horn 1898-1901: 60). If this is correct, which is far from certain (see doubts expressed already in Hübschmann 1895: 25 and in Back 1978: 200) this must be in another dialect or variety than the one which simplified *dm- > *m- (as in Middle and New Persian mān 'house').

more economical. Furthermore, only Young Avestan is known to have undergone *dm-* > *nm-* in this word, to my knowledge.

The third possibility is a simplification of **dm-* to **m-* in Tocharian itself. However, this simplification seems completely unwarranted: initial clusters are not uncommon in Tocharian, and we have a word *tmāne* 'ten thousand' which could theoretically go back to an initial **tm-*, although of course its antiquity is not assured, and it can thus not be used as an example or a counterexample.

Considering the possible simplification of **dm-* > **m-* in Old Steppe Iranian, it should not be compared to other languages which underwent the same simplification, as it is a rather trivial one, which could have occurred independently in many languages. For instance, it occurred in Ossetic and Old Persian as well as in Pashto (Morgenstierne 1927: 44) and others. Rather, we should compare the possible Old Steppe Iranian simplification with the languages that did not undergo it, to exclude common branching. Thus, Sogdian, which has *δmʾ(ʾ)n*, and Avestan are to be excluded here; Khwarezmian is ambiguous but seems to point towards a preservation of **dm-* as *δm-*;[136] Khotanese is also excluded, as it has *damānā-* 'house' (DKS: 152, written *damāna-*). To my knowledge the word is not attested in Yidgha-Munǰī; nor could I find it in Balochi.

In Bactrian I could not find a cognate of this word, however, at least word-internally, **-dm-* was preserved: νιϸαλμο, ναϸαλμο, νιϸλμο 'seat (of the gods), throne, dwelling' < **ni-šadman-* (cf. Sims-Williams 2007: 240).

For the form, one can also exclude Zazaki, which has *bān*, certainly also from **dm-* (possibly through **db-*).

Whether TB *mañiye* derives from OSIr. **nmāníịa-* or **māníịa-*, it is unlikely to derive from **dmāníịa-*, which I believe would have yielded †*tmañiye* in Tocharian B. It is thus a relevant exclusionary isogloss, which leaves Khotanese, Sogdian, Avestan and Zazaki out of the possible lan-

---

[136] In Khwarezmian two forms with *-δm-* are attested, and in both cases it occurs word-internally: one example is *γrδmʾn* 'Paradise'. However, here the word is very likely a religious borrowing (← Av. *garō-dəmāna-*) and it could also have been preserved in word-internal position only (cf. Middle Persian *garōdmān* 'Paradise', New Persian *garuθmān bihišt* '*garuthmān* Paradise'). The other form is (ʾ)*δmʾs-*/*δmʾδy̆k* 'dick werden' < **ā-dmasa-* (Benzing 1983: 28).

guages that would share a direct (immediate) ancestor with Old Steppe Iranian.

k. Old Steppe Iranian sound change *rd > *ld

The study of Old Steppe Iranian has led me to conclude that, possibly apart from the palatalized context *rdi̯i̯, and *rdⁱ (see point a. in the present discussion), the regular outcome of Proto-Iranian *-rd- seems to have been Old Steppe Iranian *-ld-. This was already suggested by Tremblay (2005: 424) who explains it through a very contrived chain of changes: "*rd becoming something like ḍ, δ̣ or ḷ, whereas *rt was retained or did not go further than *rṭ or *rḍ".

There is no need for such a complex scenario: Proto-Iranian *rd in my view had rather directly become *ld in Old Steppe Iranian. There are many arguments one could evoke against Tremblay's reasoning, but simply from the phonetic point of view ḍ, δ̣ or ḷ to ld is very unlikely, although theoretically a substitution of ḷ by ld cannot be totally excluded. Instead, one can view the change from *rd to *ld as an assimilation of the place of articulation. If, for instance, *r was an alveolar flap or trill, and *l a dental liquid, then this sound change would suggest that the Old Steppe Iranian *d was a dental plosive. Such a change is quite frequent cross-linguistically, including among Iranian languages, e.g. Middle and New Persian l < *rd, with compensatory lengthening.

The Old Steppe Iranian sound change *rd > *ld is seen in the following words: TB speltke from OSIr. *spardaka- or *spr̥daka- 'zeal', TB melte, from OSIr. *mr̥da- 'top'; TB welke from OSIr. *u̯ardaka- or *u̯r̥daka- 'petal' or 'stalk'.

The TB word kercci, possibly from OSIr. *gr̥di̯a-, poses an obvious problem here, because it can imply that *rd > *ld was only a phonetic, not phonologized development, that is, */rd/ had an allophone *[ld] in some contexts. If TB kercci is to be removed from the list of Iranian etymologies, or if my explanation of its sequence -rcc- (s.v.) is to be accepted, then the sound change *rd > *ld of Old Steppe Iranian is without exception.

One should however note that TB wertsiya, TA wartsi 'assembly', which I derived from OSIr. *u̯ardz(i̯ā)-, shows that this sound change did not occur in front of the phoneme *dⁱ. This could either imply that the sound change was affecting */r/ + */d/ but not */r/ + */dᶻ/ and thus contradict my

previous statement concerning *kercci*. A possible solution is **dᶾ* was not pronounced at the same point of articulation as **d*. For one, **z* cannot be pronounced at the dental point of articulation, and a simple pronunciation exercise will prove without difficulty that /dᶻ/ needs to be pronounced at least in the apex.[137] A third possibility is that the sound change occurred when **dᶾ* (and **tˢ*) were still the palatal sounds **ǰ* and **ć* of Proto-Iranian when the sound law **rd > *ld* took place. An obvious difficulty with this hypothesis is that it requires the sound change to have taken place very early, at a stage when (most) Iranian languages had not diverged from each other.

Here I wish to suggest another solution, which could enlighten us on the prehistory of both TB *wertsiya* and *kercci*. If **rd* – at least in pre-vocalic position – was actually realized as [rð], because **d* was pronounced [ð], then we could very easily understand why **rdᶾ* is not reflected as **rcc* in Tocharian: **dᶾ* was not realized as [ðz], which is a difficult cluster. Furthermore, one could surmise that **-dᶤ-* was not realized as [ðj] but either as [dj] or as [θj], in both cases **gardᶤa* would be palatalized to PT **kercye*. This would be either because **d* was only realized as **[ð]* in pre-vocal position, or because of a sound change **[ðj] > *[θj]*. There is meager evidence for the latter among Iranian languages, so this solution is rather speculative, but one can nonetheless cite punctual examples (that is, not occurring due to a specific sound law), such as Shughni *yē̆θ* 'nest' < **āhadᶤa-*.[138]

The sound change **rd > *ld*, sometimes accompanied by **rt > *rd*, occurred in many Iranian languages, sometimes as part of shared innovations, and in other cases as part of an areal phenomenon.

Here again, only languages that did not partake in this innovation can be said to be not descending from Old Steppe Iranian. These are: Old Persian (however, Middle and New Persian share this innovation); Sogdian (*wrδ* 'rose', cf. Gershevitch 1961: 44); Khwarezmian (*wrδc* 'Blüten', cf. Benzing 1983: 651); Pashto (*zṛǝ* 'heart' < **ĵṛdaᶤa-*, cf. Cheung 2011: 187); Ossetic (cf. Cheung 2002: 29); Parthian (*wār* 'flower', cf. DMMPP: 335).

---

[137] A dentalized laminal alveolar variant of the /z/ phoneme exists (noted as ⟨z̪⟩ in the IPA), however it is only found cross-linguistically in word-initial prevocalic or in intervocalic position. As Professor Lubotsky points out to me (p.c.), **-d-* might also be apical, as it is usually dental only when opposed to, for instance, retroflexes.

[138] I thank Professor Martin Kümmel for this example.

The situation in Pashto and Ossetic is slightly more complex. Since it might help us shed some light on what happened in Old Steppe Iranian, I wish to enter it in more detail. In Ossetic, *rd* and *rt* merged as *rd* (cf. Cheung 2002: 29), like in Pashto where *rt* and *rd* also merged to *rd*, which later became *ṛ* (cf. Cheung 2013: 622-23). It seems slightly counter-intuitive to suppose that these two clusters simply merged, without any intermediary steps, although this is what seems to be the case at a first look.

For the sake of the argument, one can suppose that *rd* > *ld*, and *rt* > *rd*, both in Ossetic and Pashto: then, *ld* merges again with *rd* later in Ossetic, simply because the phoneme /l/ was rare in preconsonantal position, if not non-existent.[139] Once *rd* shifted to *ld*, a chain shift may have occurred, leading to *rt* becoming *rd*. In Pashto it is possible that either the same phenomenon as the one proposed above for Ossetic took place, or that both *ḷ* (< *rd*) and *ṛ* (< *rt*) coexisted, before merging at a later stage. This situation would perhaps be similar to that of Yidgha-Munǰī, where *rt* > *ṛ*, but *rd* > *ḷ* > *l* (cf. Morgenstierne 1938: 81).

Thus, *rd* could have indeed become *ld* in some earlier stages of Pashto and Ossetic, before it merged with *rd* < *rt*. It should be stressed that this remains hypothetical, but could theoretically explain the merger of *rt* and *rd* in Ossetic.

It is of note that Bactrian ρλ (< *rd*) reverted to ρδ in later stages of the language. This is not identical to Pashto or Ossetic, of course, as it is the λ < *d* which changed, here, but it could be seen as a somewhat parallel sound shift.

Wanǰī apparently underwent the sound change *rd* > *l*, possibly through *ld* as can perhaps be seen from *zil* 'heart' < *ĵr̥daia-* (cf. Lashkarbekov 2008: 83). However, in this language *d* (and *d* < *t*) becomes *l* in many positions, including word-finally: OIr. *pāda-* 'foot' > *pal* 'leg' and OIr. *gātu-* 'moment' > *gād* > *amγal* (< *ima-gātu-*), *γal* 'now', so one could also imagine that *zr̥d-* became *zid* and then, regularly, *zil*.

Of particular interest concerning the question of the fate of *rd* in Iranian languages is the fact that in the Persic group, for instance in Middle and New Persian, both *rd* from PIr. *rd* (< PIE *rd*, *rdʰ*, *ld*, *ldʰ*) and *rd* from PIr.

---

[139] Another possibility is that the "new" *l* (< *rḭ*) was phonetically different from the *l* in *ld*, for instance more retroflex (as it was perhaps originally palatalized), and the existence of these two different *l*s "pushed back" the *l* in *ld* to *rd*.

*rj̑ (< PIE *rĝ⁽ʰ⁾, *lĝ⁽ʰ⁾) have become *l*. This would naturally only have happened after the sound law *dʮ > d* that defines Proto-Persic. In the second case, we have OSIr. *rts* rather than *lts*: TB *wertsiya* 'community, assembly' ← OSIr. *u̯ardʮii̯ā < PIr. *urji̯i̯ā*. Changes of *rd* to *ld* or *l* are thus very old among Iranian languages, and have occurred at multiple stages, in multiple languages, including languages geographically separated from each other. The assumption of a continuum or areal phenomenon cannot justify every change from *rd* to *ld* among Iranian languages, and we have thus found the earliest example of that change in Old Steppe Iranian.

It is noteworthy that Khotanese seems to be the only known Middle Iranian language in the neighbourhood of Tocharian to have without a doubt undergone the sound change of *rd > l*. Wanjī, which is not in the immediate neighbourhood of Tocharian, could also have undergone this change, but it is not possible to determine this with certainty.

l. The Old Steppe Iranian treatment of vocalic *r̥*

To determine the reflex of *r̥* in Old Steppe Iranian, we need to turn to the Tocharian words TB *melte*, TA *malt** 'pile, summit', TB *spelkke* (< *speltke*), TA *spaltäk* 'zeal', TB *welke* 'petal of a flower', TB *kercci* 'palace', TB *wertsiya*, TA *wartsi* 'community' (< Common Iranian *u̯r̥dʮii̯ā-*).

All these words could represent either *r̥* or *ar*, since both are attested among Iranian languages, but in each case *r̥* is original. As examples I can cite *kamr̥da-* 'head', related to *mr̥da-* 'neck', where one has Bactrian καμιρδο 'head; chief-god' < *kamr̥da-*, but Persian also shows a reflex of *kamarda-* in *kamāl* 'face (vulgar, derogatory)'. One also finds Middle and New Persian *gul* 'rose, flower' < *u̯r̥da-*, as opposed to Avestan *varəδa-* 'rose' (AiW: 1369) < *u̯arda-*. The same *u̯arda-* was also borrowed in Arabic *ward* 'rose' (secondary collective from *wardaᵗ* reinterpreted as a singulative).[140]

Having considered this, it appears very unlikely to me that Old Steppe Iranian would have kept only -*ar*- variants for all four words cited above, since these reflect original zero-grades, so that -*ar*- is secondary in any case.

---

[140] Pace Eilers (1962: 207), who writes that "*ward* „Rose(n)": vgl. av. *varəδa-*; np. *gul* (und *vil* „Liebchen") gehen auf ein *u̯r̥da* zurück". Neither the Avestan nor the Arabic forms are likely to come from the zero-grade *u̯r̥da-*.

This could be theoretically possible for one, two or maybe three, but the likelihood of that is very low. Thus, I consider *-ar- to be the Old Steppe Iranian reflex of PIr. *-r̥-, at least before *-d- and *-đ-. It is likely that *-ar- was the phonological development of *-r̥- in all four words.

In order to explain these forms going back to *-ar- instead of *-r̥- in Iranian languages, it has been suggested by scholars that accented vocalic *-r̥- yielded *-ar-. For instance, Cantera (2017: 489) explains Pahlavi *kamāl* from *ka-mŕ̥da-. The same principle could also be invoked here: all the examples we have at hand (*kercci*; *melte*; *spelkke*; *welke**; *wertsiya*) contain a vocalic *-r̥- in the first syllable, which we have reasons to believe was always accented in Old Steppe Iranian, with only few exceptions, (see p. 171f.).

Based on the evidence at hand, there are two possibilities that we can consider for Old Steppe Iranian: (1) either *-r̥- became *-ar- in every position, or (2) *-r̥- only became *-ar- under the effect of stress and had other reflexes or realizations in other positions. An argument in favor of the first hypothesis, I believe, is found in the fact that the cluster *rd, even when going back to *ŕ̥, had become *-ld- in Old Steppe Iranian. This might show that the sequence was phonologically not /r̥+d/ but /ar+d/, or in any case phonetically [ərd] and not [ərəd] or [rəd]. If this is correct, one can observe that such is not the case of Old Persian (see Korn 2021 with discussion of the literature); Bactrian (e.g. καμιρδο 'head; chief-god' < *kamr̥da-); Pashto (cf. Cheung 2011: 187); Khotanese (cf. Emmerick 1989: 211f.); Yidgha-Munǰī (cf. Morgenstierne 1938: 97f.).[141]

Khwarezmian seems more ambiguous and more research is needed to determine the result of vocalic *r* in Khwarezmian. Sogdian and its sister language Yaghnobi also have a different reflex of *r̥* than Old Steppe Iranian: in both languages, *r̥* becomes variously *ər*, *ir*, ur, etc. depending on the context (cf. Novák 2013: 95; Gershevitch 1961: 19f.).

Wanǰī has three different reflexes of vocalic *r*: *r̥* either becomes -i- or *ir* as in *kr̥nau̯- > kin- 'to do' (compare Persian *kun-*) and *kirmɣaz* 'wormy', with *kirm-* 'worm' < *kr̥mi- (cf. Lashkarbekov 2008: 67), and it also becomes *ai* in front of a nasal (cf. Novák 2013: 36). For the latter change, it is

---

[141] Khotanese has multiple outcomes of vocalic *r* depending on the position, as listed in Emmerick (1989: 211-212).

the same reflex as that of *ar and *a, *ā in front of a nasal, as in Wanjī *main* 'apple' < *amarna-, γain 'woman' < *gan(ā)- < *gnā-, kain 'source, spring' < *kāna- (cf. Lashkarbekov 2008: 71). This common reflex for all four sounds or sequences is somehow disturbing, as it would suggest a merger of *r̥ and *ar in front of a nasal (> *arN > *aN). This does not square well with the fact that *r̥ is otherwise reduced to i or to ir (possibly /ər/). Perhaps – but one has to recall this is a dead language of which we have no texts and only a limited amount of words – *r̥ became /ər/ in most positions and /ar/ in front of nasal consonants, or perhaps /ər/ became /ar/ in front of nasal consonants in Wanjī. It is also possible to imagine that the reflex of Proto-Iranian *r̥ was /a/ before nasal consonants in Pre-Wanjī, before it became -i- and -ir- in the other positions. This, however, would not explain why *ar also became *a in front of nasals, and it is thus a less preferred hypothesis, in my view. In any case, the reflexes of *r̥ as -i- and -ir- exclude Wanjī as a possible descendant of Old Steppe Iranian, as one has to reconstruct a vocalic r̥ phoneme for Pre-Wanjī.

The only language of our list which has a consistent reflex of PIr. *-r̥- as -ar- is Ossetic (cf. Cheung 2002: 24). The change of *-r̥- to -ar- could thus be a common development of Pre-Proto-Ossetic and of the ancestor language of Old Steppe Iranian. However, this change might perhaps have occurred independently. Nevertheless, the very fact that it occurred so early in Old Steppe Iranian, when it did not in other Iranian languages spoken in Asia that we know of, is non-trivial.

As Korn (2016: 410-11) notes "as soon as *r̥ yielded ir, ur or ar, this sequence is indistinguishable from old sequences of vowel plus r […]. The development is thus a "point of no return" in dividing Middle Persian plus Parthian from other WIr. languages […]." The same is true about Old Steppe Iranian: PIr. *r̥ > OSIr. *ar is indistinguishable from PIr. *ar > OSIr. *ar.

However, if Old Steppe Iranian underwent the same sound law as – possibly –, other languages, such as Middle and New Persian (according to Cantera and others), namely that accented *-r̥- had become *-ar-, the isogloss question becomes more obscure. It is unfortunate that we do not have an Old Steppe Iranian word with *r̥ in unaccented position, for instance a reflex of *ka-mr̥da- (or another word with vocalic -r̥- in the second syllable), since it is the syllable *ka- that would be the one carrying the accent. This would have been helpful in order to prove or disprove Cantera's suggestion (see above).

In conclusion, although it is clear that all examples going back to vocalic *-ŗ- in Proto-Iranian are reflected with Old Steppe Iranian *-ar-, which suggests a regular development from PIr. *-ŗ- to OSIr. *-ar-, all examples also go back to an *-ŗ- that should have been accented in Old Steppe Iranian, making it difficult to determine if the sound change was triggered by the stress or not.

m. Table of phonetic innovations and isoglosses

Below is a table of all the phonetic innovations of Old Steppe Iranian discussed here, compared to the other Iranian languages I have considered, in order to sum up and give an overview of the types of isoglosses found.

Phonetic innovations of OSIr.	Avestan	OP | MP	Sogdian	Khwar.	Khotanese-Tumshuqese	Bactrian	Yidgha-Munǰī	Pashto	Balochi	Ossetic
Palatalization of *θi̯ and *di̯	No	No	No	No	No	No	No	No	No	Yes
*ai̯ni̯a- > *-ii̯na-	No	No	No	No	No	No	No	No	No	Yes
Preservation of *či̯- or palatalization of *ši̯-	Yes (pal.)	Yes (preservation)	No	No	Yes (preservation)	?	?	?	No	Yes (preservation in post-consonantal pos.)
Contraction of *a(h)a > *ā	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes
Loss of *h	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes

Possible loss of *h in front of *u̯	No	Yes|No	No	No	No	No	No	No	No	No
PIIr. *ću̯ reflected as *tsw	Yes	Possibly	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Initial *i > *i̯a- or *i̯e; *u > *u̯a- or *u̯e-	No	No	Occasional palatalization of *i-	No	No	Tendency to have *i- > ya, yi.	No	No	No	Unclear
Reduction of short *i and *u in word-internal position	No	No	Yes	Un-clear	No	Yes	Yes	Yes	No	No
Fixed word-initial accent	No	No	No	No	No	No	Tendency	No	No	Yes
Shorte-	Yes	No	Yes	No	Yes	Un-	Un-	?	No	Yes

ning āu̯V- > *-au̯V- and *-ā i̯ẵ- > *-a i̯ẵ-						clear	clear			clear
Simplifi-cation *dm- > *m-	No	Yes	No	Un-likely	No	Un-likely	Un-clear	Un-clear	Un-known	Yes
Change from *rd to *ld	No	No | Yes	No	No	No	No	Yes	No (not clear-ly)	No	No (not clearly)
Vocalic *r̥ > ar	No	No	No	Un-clear	No	No	No	No	No	Yes

Based on this table, we can say that Old Steppe Iranian agrees with Ossetic at least nine times;[142] with Sogdian five times; with Yidgha-Munǰī also five times; with Old and Middle Persian four times; with Pashto also four times; with Avestan thrice; with Khwarezmian thrice; with Khotanese-Tumshuqese twice; with Bactrian once or twice; with Balochi once. Naturally this should be taken together with lexical shared innovations, and a number of other elements in order to be significant, so that one cannot say that Old Steppe Iranian is closer to Old Persian than to Avestan or Khwarezmian for instance. Nonetheless, the amount of times Old Steppe Iranian agrees with Ossetic, including for non-trivial innovations, and including for changes which no other listed Iranian language shares (such as the development of vocalic *r̥ to ar) is remarkable, and suggests a real genetic proximity between both languages, as has been suggested in the past (e.g. cautiously, Pinault 2002a: 245).

2.6.3 Lexical isoglosses of Old Steppe Iranian

To have a complete overview of the features of Old Steppe Iranian, one must not stop at the study of its phonological innovations and archaisms, but it is also necessary to look at the lexical isoglosses it shares with other Iranian languages and the semantic shifts, developments and archaisms of this language. I should like to underline a few reasons which might explain the formation of both the lexical and the semantic isoglosses.

First, most languages surrounding Old Steppe Iranian might have disappeared without leaving any trace. Indeed, Old Steppe Iranian itself would have left no trace if it had not influenced Tocharian. This could explain why no closely related variants are known. Second, Persic is by far the best attested Iranian subgroup, since besides Old Persian, we have Middle and New Persian, which have enormous corpora, and have been lexically studied both by native and foreign lexicographers since quite a long time. Other Persic languages such as Kumzari and Bakhtiari also provide us useful information whenever data is lacking from the three languages mentioned above. It is thus not entirely a coincidence that this group is where many of our isoglosses are found.

---

[142] Since in Ossetic *i̯a- becomes i- (Cheung 2002: 18), there is no way to know if word initial *i- had also become *i̯a- in Pre-Proto-Ossetic or not.

An ancient, prehistorical connection between Old Steppe Iranian speakers and Proto-Persic (the ancestor language of Old Persian and all other Persic languages, such as New Persian and Bakhtiari, for instance) is not entirely excluded, by means of the Caspian sea. If one goes to the North East of the Caspian, one encounters the endless steppes that lead, ultimately, to South Siberia and Xinjiang. This hypothesis is not very likely, however, notably because of the geographical distance between Old Persian and these very steppes.

Avestan, in its turn, was more probably spoken on the plains of modern-day Afghanistan, or in the region which is globally situated in the South of the steppes, and I would be more inclined to believe that the lexical isoglosses one sees here are due to chance, in the sense that there is a higher probability that an archaic rare Iranian word is reflected in Avestan, which is a well documented archaic Iranian language, rather than in another language that has undergone more lexical replacement. Another factor here is naturally the extensive study that has been done on Avestan vocabulary since at least two hundred and fifty years in the West (and much longer in the East).

In this section, 2.6.3, I will analyze and discuss a number of isoglosses that include Old Steppe Iranian and a minority of Iranian languages. It will be an overview, and for details I refer to each specific word in the individual discussion above (sections 2.2, 2.3 and 2.4).

a. Old Steppe Iranian – Persic

Old Steppe Iranian and Persic share at least two specific lexical isoglosses:

1. Old Steppe Iranian *rataka- 'army' (TB retke, TA ratäk 'id.'), with its Persic correspondent *rataka- 'row, (army) line', represented by Middle Persian radag 'row, (army) rank, line', New Persian 'line, row'. A cognate of this word that is formally farther away is found in Ossetic Iron rad, Digoron radæ 'row' < *rātā- (cf. Abaev 1973: 338). The military semantics of this word must have arisen in Old Steppe Iranian, possibly due to a pars pro toto formula, since it is not particularly salient in the Middle Persian and New Persian words.

2. Old Steppe Iranian *nāθaka- / nātaka- 'protector, lord' (Tocharian A nātäk 'lord') has Indic cognates but very few Iranian cognates. All those

Iranian cognates are exclusively found in the Persic branch, namely New Persian *panāh* 'protection' (< **pad-nāh* < **pati-nāθa-*) and Middle, New Persian *nihān* 'secret, hiding' and other cognates. The root **nāθa-* 'protection' as such seems to be exclusively shared by Old Steppe Iranian and Persic.

b. Old Steppe Iranian – Balochi – Persic

Old Steppe Iranian shares one isogloss together with Balochi and Persic.

Old Steppe Iranian **gra* u*pa-* or **gra*u*fa-* 'group' (TB *kraupe*, TA *krop* 'id.') which is found in a very limited number of languages: Balochi *grōp*, Middle Persian *grōh*, New Persian *gurōh* 'id.' Dawāni *gorufa* 'ball of fiber or rope' represents a derivative of this word, (with the **-ak* suffix) which has a parallel in New Persian *gurōha* 'globe, bowl, or any spherical figure; ball for a cannon; a kind of sweetmeat; an assembly, multitude of people'.

Dawāni belongs to the Persic group (cf. Mahamedi 1994). It thus seems that this word is only found in the Persic branch, Balochi and Old Steppe Iranian. It is possible that the Balochi word was itself borrowed from an unattested Old Persian **gra*u*pa-*, making this an isogloss between Old Steppe Iranian and Persic. The other direction of borrowing, namely, from Balochi to Old Persian, is unlikely.

c. Old Steppe Iranian – Avestan

Old Steppe Iranian shares at least two exclusive isoglosses with Avestan. These can naturally be due to chance, as Avestan preserves a number of ancient lexical items and verbs that have been lost in the rest of Iranian.

1. Old Steppe Iranian **ra*i*θ*u*a-* 'combination, mixture' (TB *raitwe* and TA *retwe* 'combination, mixture' *passim*), Avestan *raēθβa-* 'to mix'.

2. Old Steppe Iranian **andza(h)* 'greed' (TB *entse*, TA *eṃts*), Avestan *ązah* 'constriction; distress, peril'. While many cognates of the root (PIr. **Hanȷ́-*) are found in Iranian languages, I was unable to find any other exact formal cognates.

d. Old Steppe Iranian – Avestan – Persic – Parthian

Old Steppe Iranian *ʒ́aịnu- 'arrow' (TB *tsain* 'id.'), Avestan *zaēnuš* 'baldric', Avestan *zaēna-* m. 'weapon'. Here we observe an interesting semantic evolution in Old Steppe Iranian, for more detail see section 2.6.4.e.

Cognates are also found in Parthian and Manichean Middle Persian *zēn* 'weapon, sword, armor, arms' (DMMPP: 386). The Parthian (or, perhaps, the Middle Persian) word was borrowed into Armenian *zēn* 'weapon', *-zēn* (as second member of compounds) 'weapon, armor'.

e. Old Steppe Iranian – Avestan – Persic – Sogdian – (Balochi)

Further, Old Steppe Iranian shares one isogloss with Avestan and Persic.

> Old Steppe Iranian *maiịā- 'power, strength, might', Avestan *maiiā-* 'strength, craft', Old Persian *māịā- 'craft, strength', Sogdian *myʾkčyk* /mayākčīk/ 'happy'. The word probably existed in other branches as well, but can only be found in Avestan, Sogdian, and, residually, in the Persic branch (through names attested in Middle and New Persian as well as Elamite).

f. Old Steppe Iranian – Avestan – Khotanese – Ossetic – Balochi – Yidgha

See the discussion of *yetse* p. 62-63 for more detail on this possible (but complicated) isogloss.

g. Old Steppe Iranian – "Eastern" languages – Scythian and Ossetic – Khotanese – Khwarezmian

This is the opposite of the Old Steppe Iranian – Persic isoglosses mentioned above, since the Old Steppe Iranian word for 'axe', *paratu-, is only shared with Ossetic *færæt*, Scythian Φαραδο[υ] (inscription of a proper name, see also fn. 16), Khwarezmian *pdyk*, Khotanese *paḍa-* (for etymological details, see section 2.2.11). Furthermore we have another (probably) Iranian language from which Old Turkic borrowed its word for 'axe': *baltū* < *baltō* 'axe', but that language remains unknown. Dragoni (2022: 144) proposed to view *baltō* as a borrowing from the Old Khotanese acc. sg. *paḍu*.

In Avestan and Sogdian another word for 'axe' is used: Avestan *taša-*, Sogdian *taš*, and in the west, the word **tapara-* and descendants of it are found. The word **parat^u-* is thus quite limited in distribution. Naturally it could have existed in Proto-Iranian and been replaced elsewhere, but it seems more likely that there was no Proto-Iranian word for 'axe', or perhaps there were different types of axes in Proto-Iranian times, each of them having a different name, and these types were generalized among different groups. This question needs to be answered through an archaeological lens, which goes beyond the scope of the present study. This isogloss is nevertheless both relevant and useful, in that it places Old Steppe Iranian in a group of Iranian languages which shared the same word for 'axe', although they were not (all) closely related.

h. Old Steppe Iranian – Ossetic

The Old Steppe Iranian meaning 'sword' of **karta-* (TB *kertte* 'sword') is only to be found in Ossetic (cf. p. 32f.), which also has the meaning 'knife' and 'saber' for this word. In all other Iranian languages, the bare, unsuffixed word **karta-* strictly means 'knife' (New Persian kārd, etc.) so that this convergence of meaning is of particular interest. This isogloss is not entirely assured (see section 2.6.4.f for a different explanation), but of rather important consequence if it were correct. It would suggest that the ancestor of Old Steppe Iranian and that of Ossetic, either areally or genetically, developed the meaning 'sword' or, in any case 'cutting war weapon' for **karta-*. Given the amount of proximity between Ossetic and Old Steppe Iranian (see previous section), the fact that these two languages would have acquired this meaning for this word separately seems rather improbable.

i. Old Steppe Iranian – Sogdian – Persic

The Old Steppe Iranian word **paribanda-* 'load, burden' (TB *perpente* 'burden, load') corresponds exactly, apart from the lack of a reflex of the suffix -*aka-*, to Sogdian *prbnty* 'idem' and Persian *parwanda* 'bundle of clothes, etc.' (s.v. TB *perpente*). Although it is likely that any Iranian language could have created such a word at the Old Iranian stage, or perhaps even later, it is still noticeable that only Sogdian and (Old) Persian did, to our knowledge.

There is a small chance, I believe, that the Persian word *parwanda* is borrowed from Sogdian, or from a hypothetical Bactrian cognate. Indeed, the

expected Persian form is †*parbanda*. However, the shift *rb > rw* is dialectally attested in Persian, so that the word could also be dialectal, which fits well with its material semantics.

j. Summary

Old Steppe Iranian is thus found to share exclusive isoglosses with most groups of Iranian languages, including Persic. Interestingly, no Old Steppe Iranian – Khotanese-Tumshuqese(-Wakhi) isogloss has yet been found. Since, on the one hand, some Iranian languages (such as Persic ones) have been much more documented than others, and on the other hand, Old Steppe Iranian is an archaic language, these isoglosses should naturally not all be taken at face value. An important isogloss is the semantic isogloss for **karta-* shared by Old Steppe Iranian and Ossetic. If we bring this together with the observations of sound changes in the previous section, it seems to strengthen the idea of kinship between Old Steppe Iranian and Ossetic. All in all, these isoglosses do represent important departure points for further research on the phylogenetic identity of Old Steppe Iranian, and perhaps for research concerning its geography and the migratory history of its speakers. This data should be coupled to data on semantic innovations of Old Steppe Iranian (see next section).

2.6.4 Semantic innovations of Old Steppe Iranian

I believe it could be useful to discuss a few semantic developments found only in Old Steppe Iranian, besides the lexical isoglosses mentioned in the previous section.

In the same way that it was sometimes difficult to determine whether a specific sound change had occurred within Old Steppe Iranian or in Tocharian, or in the process of borrowing, it is sometimes difficult to determine whether a semantic shift occurred within Tocharian or already in Old Steppe Iranian, although it is generally easier than for sound changes as, naturally, less examples are needed to propose a semantic change than to establish a sound law.

a. Old Steppe Iranian *anǯa(h) 'envy, greed'

While multiple Iranian languages have a word āz, āzi with the meaning 'greed' (cf. p. 21f.), the Proto-Iranian form *Hanjah- has only one other known cognate: Avestan ązah-, with a very different meaning: 'constriction; distress, peril'. I wonder if Buddhist influence could be the source of the semantic change, that is, after the Proto-Tocharian period, and quite some time after the word was borrowed.

b. Old Steppe Iranian *čⁱāta- 'happy, content'

Together with Federico Dragoni I was able to establish that the meaning 'rich' of TB śāte, TA śāt is probably due to Buddhist influence. Indeed, such a meaning is found in Khotanese tsāta- and exclusively in Buddhist Sogdian šʾt /šāt/ and not in Christian or Manichean Sogdian. Also, in Tumshuqese, the sister language of Khotanese, and a Buddhist language, tsāta- means 'possession' (possibly this is a backformed noun, derived from an unattested adjective *tsāta- 'rich'). No other Iranian language attests a meaning 'rich' for any cognate of these words.

This apparently purely Buddhist semantic shift has presumably occurred in a literary context, due to the influence of Buddhist Sanskrit bhoga- and derivatives, which ambiguously meant 'enjoyment, happiness' and 'possession, property, wealth' (cf. e.g. M-W: 767), which śāte, tsāta- and šʾt often translate. One can take as an example bhoja- 'bestowing enjoyment, generous', and the adjective bhogavat-, which means both 'happy' and 'wealthy, opulent'. A concrete example of this can be seen from the Suvarṇabhāsottamasūtra (cf. Skjærvø 2004: I, 111), where Khotanese tsāta- (§6.2.64) translates Sanskrit bhoga-.

Since it is impossible that Old Steppe Iranian speakers were Buddhists, it is more than probable that the Old Steppe Iranian language still had *čⁱāta- in the meaning 'happy, content', and that the semantic shift occurred within Tocharian, alongside Khotanese, Tumshuqese and Buddhist Sogdian, due to a calque of Sanskrit bhoga- 'happy, wealthy, opulent'.

c. Old Steppe Iranian *māníia- 'servant'

The Old Steppe Iranian word *māníia- 'servant' (Tocharian B mañiye 'id.'), initially a derived adjective 'of the house', is perfectly identical to the French

word *domestique* 'servant' (< Lat. *domesticus* 'of the house') in its semantic derivation. Typologically, one can also compare Armenian *aɫaxin* 'female servant' derived from *aɫx* which originally meant *'house' (cf. Martirosyan 2010: 24f.).

We find an identical semantic shift in Old Persian *māniya-* (translated as 'household slave(s)' by Kent (1953: 202). However, an interesting, very different semantic shift is found in Ossetic Iron *moj*, Digoron *mojnæ* 'husband' < *māniịa-* as convincingly argued by Cheung (2002: 206). From 'the one of the house' one can easily derive 'husband' and 'servant', but 'servant' is hardly derived from 'husband', and vice-versa. It thus appears that the meaning 'servant' must be the result of a separate innovation in Old Steppe Iranian, and the meaning 'husband' in Ossetic also, both branching out early enough. To sum up: the original meaning, in Proto-Iranian, was simply 'of the house', and the shift to 'husband' and 'servant' must be independent from each other.

Of course, it can be argued that, if the meaning 'of the house' was retained long enough, it would have been possible to remake a new word meaning 'servant' or 'husband' at any point in time. However, given the strong lexical preference that must be given to one or the other meaning, this hypothesis seems quite unlikely to me.

d. Old Steppe Iranian *ʣainu- 'arrow'

In Tocharian B *tsain* means 'arrow'. It derives from Old Steppe Iranian *ʣainu-*. In Avestan there exists a word *zaēna-* m. 'weapon', also YAv. *zaiia-* 'kind of weapon', cf. also Vedic *hetí-* 'missile, weapon', and also Avestan *zaēnuš* which means 'baldric' (cf. Lubotsky 2021: 228). The original meaning of the root was 'to hit' (cf. Lubotsky, op. cit.). The meaning 'arrow' could have arisen within Old Steppe Iranian.

In Parthian and Manichean Middle Persian *zēn*, also going back to *ʣainu-*, means 'weapon, sword, armor, arms' (DMMPP: 386). The meaning 'arrow' is thus a specific specialization of this word. Possibly, Old Steppe Iranians had remarkable arrows, which they, or the Tocharians, designated as the weapon *par excellence*. Old Steppe Iranians can perhaps be associated with a horse-riding group, for archaeological reasons (see p. 252). It is possible, if we combine this with the semantic shift 'weapon' > 'arrow', that a sizeable part of these horse-riders were mounted archers. Perhaps, as

many soldiers of the antiquity, they carried both arrows and a sword (*kertte*), which permitted them to attack both from afar and from near.

e. Old Steppe Iranian **apa* 'neither… nor'

On both typological and historical grounds I inferred that the intermediary meaning between the Proto-Iranian meaning **'away from' of **apa* (documented in Old Iranian languages) and the Tocharian meaning 'or' of *epe* should have been **'neither' (with a secondary change 'neither' > 'or', compare North African Arabic *wala* 'or' < *wa-lā* lit. 'and-not'). We can add Old Uyghur *ap* 'nor', *ap … ap* 'neither… nor'; 'oder … auch, entweder… auch' (cf. Clauson 1972: 3; Wilkens 2021: 55) to the comparison, which was probably borrowed from some early, prehistorical stage of Tocharian.[143]

If my argument is accepted, the Old Steppe Iranian meaning may have been 'neither … nor' rather than 'or' at the time of the borrowing. Such a meaning cannot be found in any other Iranian language. It should thus be a specific semantic development of Old Steppe Iranian.

f. Absence of suffixes – suffixed meanings

I have noted that three different Old Steppe Iranian words are reflected without any suffix, yet have a meaning that in my view would only be possible with a suffix. These are:

1. **atˢu̯a-* 'mule', which, semantically, represents a virtual **atˢu̯a-tara-* 'similar to a horse = mule' as seen in New Persian *astar* 'mule' and Khwarezmian ʾ*sptyr* 'id.' The word for 'mule' in Sogdian and Khotanese rather goes back to **xara-tara-* (e.g. Khotanese *khaḍara-* 'mule'). However, one should keep in mind that the meaning 'mule' is only known through an Old Uyghur gloss (cf. Peyrot 2018a), and the exact meaning of *etswe* could perhaps be refined if the word can be found in a different context.

---

[143] Rather than directly from Old Steppe Iranian, as we have no other trace of such a contact (cf. ongoing research by Hans Nugteren, Michaël Peyrot and Jens Wilkens).

2. *karta- 'sword', which semantically has the meaning found in *karta-tara- 'similar to a knife = sword' as found in Khotanese kāḍara- 'sword' from *kartara, which I believe to come from *karta-tara- with haplology (pace DKS: 58 with *karta-ra-, which would have a suffix -ra- that is otherwise unexplained). One can also compare Sogdian ākartē 'sword' (< *ā-karta-ka-), cf. SD: 6. A different explanation is provided (p. 34, and section 2.6.3.h).

3.*paribanda- 'load, burden', which semantically looks as if derived from *paribanda-ka, which is indeed reflected in Sogdian prbnty and New Persian parwanda (both < *pari-bandaka-). Without the -ka- suffix, I expect the meaning of such a noun to have been simply 'the binding around', although it would then have quickly reached the same meaning. The absence of suffix is thus less striking for this specific word than for the two previous ones.

We could attribute this lack of suffixes to Tocharian: an inner Proto-Tocharian simplification of the hypothetical "too long" forms *kertetere, *etswetere and *peripenteke. However, according to the sound law we have observed, that is *é_e_e > e_ø_e, we should expect †etsu < etsutere etc. Rather, I believe this is an internal Old Steppe Iranian issue, which cannot be explained at this stage, but needs to be mentioned, in any case.