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.

Citation

Bernard, C. B. A. S. (2023, April 20). *Like dust on the Silk Road: an investigation of the earliest Iranian loanwords and of possible BMAC borrowings in Tocharian*. Retrieved from https://hdl.handle.net/1887/3594385

Chapter 3: BMAC words in Tocharian (a selective survey)

3.1 Introduction

The study in the previous chapter of Old Steppe Iranian loanwords in Tocharian led to both the inclusion and the exclusion of a number of Tocharian words from the list of Old Steppe Iranian borrowings. Some of them cannot be of Indic or of Iranian origin, but they are not inherited either, or at least, an Indo-European etymology has never been found for them, such as TB *witsako* 'root'. Besides, a number of Tocharian words have been analyzed as of substratal origin, and this origin has usually been ascribed to the same substratum as that of Indo-Iranian, also tentatively called the language of the BMAC (for Bactriana-Margiana Archaeological Complex). Carbondating dates the civilization of the BMAC to 2250 – 1700 – 1500 BCE (cf. Parpola 2022: 26 with references).

The pioneering study of Tocharian loanwords of BMAC origin is that of Pinault (2006). In that paper, Pinault extends Lubotsky's (2001) substratal study on the Indo-Iranian vocabulary to some elements of the Tocharian lexicon. As a conclusion, he deduces that the BMAC language had definitely enough influence on Tocharian speakers to provide them with a number of important words. Some of these elements are already present in Pinault (2002).

A number of points need to be expounded before delving further into this issue: 1. it is unsure whether this so-called substratum language was indeed a substratum; 2. even if it were a substratum for Indo-Iranian, it is not certain at all that it was a substratum for Proto-Tocharian. Lubotsky (2001) himself writes that he is unsure whether this language was a substratum or rather an adstrate. Since its contribution to Indo-Iranian was apparently mostly lexical, it is more likely that it was an adstrate or a superstratum (cf. e.g. Kümmel 2020: 255); 3. for practical purposes I will call it "BMAC language", in the case of Tocharian contact, and "Indo-Iranian substratum" to refer to the variety in contact with Indo-Iranian in particular. In the discussion at the end of this chapter, I will investigate whether we can tell if the BMAC language

that influenced Proto-Tocharian was the same as the one that influenced Indo-Iranian, or whether they were for instance sister languages.

In the present chapter, I will discuss a selected number of Tocharian words that appear to have substratal features, or have been claimed to be BMAC words. In general, an Ir anian origin has been ascribed to them, although it can, in my opinion, no longer be maintained, in light of the phonetic correspondences discussed in the previous chapter. This study is selective in the sense that I was not able to accomplish a full survey of all potential BMAC words in the Tocharian corpora, and, unlike for Old Steppe Iranian, I did not discuss every possible BMAC loanword in Tocharian mentioned in scholarship, because it was not the initial purpose of my research. For instance, I have not discussed the proposed BMAC loanword *$\bar{a}ni$- 'hip' (cf. Pinault 2003; 2005) because its etymology is rather complex. In general, there are not many proposed BMAC loanwords, and this topic deserves, in my view, greater consideration.

Some common features can be recognized for most of those words: 1. apart from TB *etre* TA *atär* 'hero', they designate realia: plants, animals, or construction material. 2. Most of them are trisyllabic, and, when visible, the stress is constantly on the second syllable. This reminds us of the feature of the BMAC loanwords as described by Lubotsky (2001: 303), namely: an unusual syllable structure, mostly trisyllabic nouns with a long vowel in the middle syllable. I will discuss these features in greater detail in the discussion at the end of the chapter.

3.2 Analysis of potential BMAC loanwords in Tocharian

3.2.1 TB *iṣcäke* 'clay or brick', TB *iścem* 'clay'

The Tocharian B words *iṣcäke* 'clay, brick' and *iścem* 'clay' are related, both formally and semantically. Their etymology, however, is complex and intricate. I will discuss various etymological proposals concerning these words, and try to put forward my own. In 3.2.1.1, I will discuss Pinault's and Adams' proposal. In 3.2.1.2, I will discuss cognates of these words in other languages, which will permit me to propose another solution for *iṣcäke*. In 3.2.1.3, I will discuss the origin in Tocharian of *iścem*, which I believe to be related but not identical to that of *iṣcäke*. Finally, as an annex, I add a philo-

logical study of the Avestan word *zəmōištuua* and related forms in 3.2.1.5, which will be useful to the present discussion.

3.2.1.1 On the meaning and etymology of TB *iṣcäke* 'clay'

The word *iṣcäke* occurs only once, in the form *iṣcạke*, in a bilingual St.-Petersburg manuscript of the Petrovsky collection. The line where it appears reads *tokharika* : k_u*cañ̃e iṣcake*. The sequence ‹ṣca› stands for ‹ṣcạ›, i.e. *ṣcä* (cf. Peyrot 2008: 179). Since the text in which *iṣcäke* occurs is archaic, it is impossible to determine whether the schwa was accented or not.

This line has been much discussed, but it seems that a definitive break-through was made by Pinault (2002), according to whom *iṣcạke* designates a type of clay, and translates *tokharika*, a pseudo-Sanskrit form corresponding to Sanskrit *tūbarika*, a word designating a type of clay. As Adams (DTB: 191-92) suggests, a meaning 'a sort of' is probably to be posited for k_u*cañ̃e*. Thus we could translate k_u*cañ̃e iṣcạke* as "a type of clay". Theoretically, "a type of clay" could metonymically also designate a specific sort of brick based on the stuff of which it is made.

Pinault (2002: 325-335 and 2006: 171) was the first to connect TB *iṣcäke* etymologically to TB *iścem* 'clay', sometimes translated as 'clay brick' (on which see below). Pinault suggested that there existed a noun **iśc*, a cognate of unattested TA **iśäc*, borrowed in Old Uyghur *išič* (*ešič* 'Kessel, Topf', cf. Wilkens 2021: 264). As a paradigmatic analogy with TB *āśce* 'head' : obl.sg. *āśc* and other nouns having the same pattern, a nominative **iśce* was formed. This noun **iśc* would derive from a BMAC word **išti-* with **-ti* > **-cä* as in words from PIE (cf. Pinault 2002: 330).

The form **iśce* would have been enlarged with the suffix *-maṣṣe* "servant à dériver des adjectifs de relation, surtout avec valeur technique" (2002: 328), to form *iścemaṣṣe* 'earthenware; (thing) made of clay', and from which *iścem* 'clay brick' was extracted. However, it was later demonstrated by Peyrot (2008: 94) that the suffix *-maṣṣe* is both late and colloquial, while *iścem* appears in classical texts. It is thus unlikely that *iścemaṣṣe* derives from *iśce**, rather than from *iścem*.

The form *iṣcäke* itself would have, according to Pinault (2002: 331), de-rived from this same **iśce*, through the addition of a suffix *-ke* of Iranian origin. Pinault interprets this suffix as possibly indicating "une sorte de", and thus glosses *iṣcäke* as "une sorte de terre" or "une sorte d'argile" (2002:

331). Nonetheless, this suffix only occurs in Iranian, Indic or BMAC loan-words, where it never has the meaning 'a sort of', as one can see in the examples cited by Pinault: TB *kattāke*, TA *kātäk* 'maître de maison'; *rṣāke*, TA *riṣak* 'sage, ascète', either from Indic or from Indic through an Iranian intermediary; *ṣecake* (s.v.) 'lion' from BMAC, TB *ainake*, TA *enāk* 'bas, vil' (s.v.) from Old Steppe Iranian. Pinault suggests this meaning 'a sort of' because it is a meaning this *-ka* suffix can bear in Indic, but there is no need for that, since the meaning 'a sort of' is already provided by *kᵤcaññe* (cf. DTB: 191-92).

There is no other example of a borrowed noun following a pattern similar to that of *āśce* in Tocharian, and no positive evidence for BMAC **-ti-* > PT **-cə-*, nor is there negative evidence against it, I have to admit. Although Pinault's interpretation is very enlightening on many points, I believe a more straightforward scenario can be presented to account for the etymology of both *iṣcäke* and *iścem*. Before presenting my own theory, I wish to discuss another etymology of *iṣcäke* and related words, namely, Adams'.

Adams (DTB: 72) wrongly writes that Pinault takes *iṣcäke* to "represent an earlier Iranian **iṣtyaka-*." Adams further (DTB: 73) takes *iścem* to derive from an eastern Iranian language form "*ištyám*". In his opinion these Iranian words derive from **h₂eis-* 'fire (clay)' → **h₂isti-*.

There are some problems with this proposal, the most important being the semantic aspect of it. Indeed, the meaning of **h₂eis-*[144] (cf. Av. *aiiaŋha-* 'cauldron' < PIIr. **aịasa-* 'cauldron' < **'metal(-ware)'*) does not mean 'fire (clay)', but refers to the process of heating metals, cf. YAv. *aiiah-* n. 'metal', Vedic *áyas-* n. 'metal, Nutzmetall' (EWAia¹: 104). This is a totally different meaning than that of 'brick'. Admittedly, it is conceivable that words for brick refer to the process of brick-baking. However, in most Indic and Iranian languages, and in the Burushaski cognates discussed below, these words refer to sun-dried bricks, i.e. simple mud-bricks, which involve a completely different preparation process than metal. Both the objects (clay pots, clay bricks vs. metal-ware, metal pots) and the preparation processes (burning, branding vs. sun-drying) are fundamentally different. I therefore do not believe that this etymology is possible on semantic grounds.

---

[144] The **h₂* here is not assured by means of reconstruction, but based on the hypothetical connection with **h₂eidʰ-* (cf. LIV²: 229).

Rather than looking for a complex, unclear Indo-European origin for *iṣcäke* as Adams does, I believe, like Pinault (2002), that TB *iṣcäke* can be explained as a borrowing from the BMAC language, although my solution differs from his in the detail. I also explain *iścem* directly from Old Steppe Iranian.

3.2.1.2 Cognates and further etymology of TB *iṣcäke* 'clay'

Cognates of the Tocharian B word *iṣcäke* are found in Indic and Iranian languages and in Burushaski. I believe it is useful to cite them here, before discussing the etymology of TB *iṣcäke* in more detail.

Some of the Indic cognates are: Vedic *íṣṭakā-* f. 'brick', Sanskrit *iṣṭikā-* f. 'id.', Buddhist Sanskrit *iṣṭā-* (BHSD: 115: "[p]erhaps loss of suffixal ka [...]"), Prakrit *iṭṭhakā-* 'tile, brick', Nepali *ĩṭ*, Assamese *iṭā*, Hindi *ī̃ṭh*, Gujarati *ī̃ṭi* f. 'brick'. Although it could be inherited, Balochi *īt* 'brick' could also be a borrowing from an Indic language (Korn 2005:137).

Some of the Iranian cognates of these words are: Old Persian *išti-* 'sun-dried brick' (Kent 1953:175), YAv. *ištiia-* n. 'Ziegel, Backstein' (AIW: 378), YAv. *zəmōištuua-* (AiW:1691) 'Lehmziegel' (= Eng. 'adobe'), Khwarezmian *štyc*, < *(i)šti-c(y) pl. *'štyc* (Benzing 1983:99, 601), MP *xišt* (CPD: 94), NP *xišt*. The latter was borrowed into Pashto as *xaxta*, f. Geiger (cited by EVP: 98) was the first to propose it as a borrowing from Persian. Cheung (2013: 618-19) considers the borrowing to be quite old.

Burushaski, a language isolate, has the forms *diṣçík* (sg. and pl.) and in the Yasini dialect *giṣṭék*, *giṣṭík*, *kiśtíki*, *diṣṭík*, cf. Shina (a Dardic language) *diṣṭík* (Berger 1998: 121), all meaning 'sonnengetrockneter Ziegel'. Since neither *di-* nor *gi-* are nominal prefixes in Burushaski, this "élargissement avec occlusive dorsale" (Pinault 2002: 330 concerning the form *diṣçík*) cannot be immediately explained.

In any case, as one can see, Indo-Iranian forms go back to a cluster *-št- while Burushaski (at least dialectally) and Tocharian have forms that go back to a cluster *-šč-. This might suggest that Indo-Iranian languages borrowed this word separately from Tocharian and Burushaski. Perhaps Indo-Iranian adapted an original cluster *-šč- as *-št-, or perhaps this variation was found in different BMAC dialects.

In passing, I would like to exclude another etymology for Tocharian B *iṣcäke*. Because of the Yidgha sound change *št > šč through *štʸ (Morgen-

stierne 1938: 73), I wondered if this word was borrowed from Pre-Yidgha into (Proto-)Tocharian. In that case, it was possible to consider such a Pre-Yidgha or Proto-Yidgha-Munǰī word as an Old Iranian loanword into Tocharian.

However, this hypothesis is weakened by two major points: if *iṣcäke* goes back to a Proto-Yidgha descendant of an earlier *ištika-, then the *-i- of the Old Iranian form was reduced to schwa while the word final -a was preserved, which is chronologically difficult, if not impossible, because the sporadic reduction of unstressed short vowels to schwa (notably of *i* and *u*, cf. Morgenstierne 1938: 95) is a much later phenomenon in Yidgha than the loss of word-final stem short vowels. This is especially true of the -aka-suffix, which was reduced to -ë̆ ~ -ïy in Yidgha (Morgenstierne 1938: 114), although it was retained longer after -u and consonant (Morgenstierne, op. cit.). We can thus expect TB †iṣcike or †iṣcäk, but a derivation of *iṣcäke* from Pre-Yidgha seems to be a chronological problem in itself.

The second difficulty with a Yidgha origin for the Tocharian B form *iṣcäke* 'brick' is that the words for 'brick' in Yidgha (and Munǰī) are *uštu* (← Khowar) and *xišt* (← Persian), thus later borrowings. This does not imply that an inherited word for 'brick' could not have existed in Yidgha, but since it was replaced, it was possibly not a prominent technological feature of the Pre-Yidgha people. Both these arguments make the hypothesis of a borrowing from Yidgha or Pre-Yidgha for this word very unlikely. An independent borrowing from a BMAC language remains the most likely option for TB *iṣcäke*.

The -äke ending of the Tocharian word, cannot derive from the Iranian -aka- suffix. Indeed, the latter should be reflected as †-eke if from Old Steppe Iranian and there would not be a final -e if from Middle Iranian (for Khotanese, see Dragoni 2022). I propose that this -äke element goes back to the *-ka- suffix of BMAC, discussed in Lubotsky (2001) and seen for example in the Indo-Iranian words *atka- 'cloak', *stuka- 'tuft of hair'. It is likely that the word reflected by TB *iṣcäke* 'clay' also contains a reflex of this suffix.

Indeed, there is no trace of a suffix -ka for the 'brick' word in any Iranian language. The use of the -ka- suffix is extremely common in Iranian languages, which means that it could have been added to this word in an unattested language, although the fact it is lacking from Iranian languages in this

word, coupled with the other arguments cited above, make the hypothesis that *iṣcäke* derives from a posited form **ištika-* rather unlikely.

One could imagine that the Tocharian *-äke* represents the *-ika-* part of **ištika-*, built on PIr. **išti-*, but, if my etymology of TB *epastye* 'skillfull' (< Pre-B **epəstiye*) is accepted, OSIr. **-ti-* did not become **-cə-*, although it can be argued that the preceding **-s-* blocked the palatalization. Another argument is that **i-* should have been rendered as **ye-*, were the word of Old Steppe Iranian origin (as in *yentuke* s.v., *yetse* s.v.). A counterargument to that is that TB *iścem*, which I argue below is of Old Steppe Iranian origin, does not start with *ye-*. One could suggest the *i-* in *iścem* was influenced by the *i-* in *iṣcäke*, of identical meaning, or that its initial *i-* remained because it was not accented. Why would Proto-Tocharian have borrowed two different words, with the same meaning, from the same language?

As to the initial *d-* and *g-* or *k-* of Burushaski in this word, I take it that they represent attempts by Burushaski speakers at reproducing an initial sound which was not readily available in their own phonological system. The presence of this initial sound can also be seen from the fact that the word was likely borrowed with an initial laryngeal in (Indo-)Iranian. Indeed: it was likely rendered as **ḫ* (Kümmel's notation), hence the *x-* in Middle and New Persian *xišt* (cf. Kümmel 2018: 166), but this initial sound was *a priori* not reflected in Tocharian.

3.2.1.3 TB *iścem* 'clay'

If Tocharian *iṣcäke* is a direct borrowing from BMAC into Tocharian, one might be tempted to believe that this is also the case for Tocharian B *iścem* 'clay' and its derivative *iścemaṣṣe* 'made of clay'. However, we do not know of a BMAC suffix or ending **-am* (cf. Lubotsky 2001). A more straightfor-ward etymology consists in taking it from OSIr. **išti̯ám* (neut.), as done by Adams (DTB: 73). The root **išti-* 'brick' in Indo-Iranian is ultimately a BMAC borrowing (cf. Kümmel 2020: 257).

An obvious counter-argument is that, in regard of TB *yetse*, TA *yats* 'skin' (s.v.) and TB *yentuke* 'Indian', both from Old Steppe Iranian, we would expect Tocharian B **yeścem*, as per the sound law PIr. **i-* > OSIr. **i̯a-* (or **i̯e-*), cf. p. 166f. Three solutions can be proposed here: first, one can imagine that, if the word was accented on the last syllable (as a neuter noun), a different rule applied: unaccented **i-* would then remain **i-* in Old Steppe

Iranian. Another solution consists in suggesting that, because the Iranian word started with a consonant (*h_2- or *$h̥$- in Kümmel's notation), see above (also Kümmel 2018: 166), the *i of the Old Steppe Iranian form remained as such, as it was not in absolute word-initial position. The third solution is that the initial i- of *iścem* would have been influenced by that of *iṣcäke*, of identical or very similar meaning.

3.2.1.4 Conclusion

In conclusion, I suggest that Tocharian B *iṣcäke* 'brick' was borrowed from BMAC directly, as proposed by Pinault (2002). TB *iścem* 'clay brick', however, was borrowed from Old Steppe Iranian *$*išti̯ám$, a neuter accented on the second syllable. This is similar to *weretem** 'debt, surety' (s.v.), also borrowed from an Old Steppe Iranian neuter *$*u̯rátam$.

3.2.1.5 Annex: a philological discussion on Avestan *zəmōištuua*

An analysis of the Avestan form *zəmōištuua*, often cited in conjunction with the Tocharian forms cited above, seems in place, although not essential to the discussion. The word *zəmōištuua* appears in a passage of the Vendidad (VIII:10), namely:

> *duua dim nara isōiϑe vīzōištąm vīzuuārəntąm maγna anaiβi-vastra*
> *zəmōištuue vā zarštuue vā upaskanbəm*
> "deux hommes, vaillants et habiles, nus, sans vêtements, devront déposer
> le corps sur la brique d'argile ou de la pierre" (Lecoq 2017).

Here *zəmōištuue* is translated by 'brique d'argile', and in the Pahlavi translation '*zamīgēn*', lit. 'earthen', the intended meaning being probably 'earthen (brick)', with a basic meaning similar, if not identical, to Tocharian *iścemaṣṣe*. Bartholomae (AiW: 1691) suggests the etymology to be *$*zama$ + *$*išti̯ah$, but he notes "[m]an erwartete freilich bei dieser Et. die Schreibung *zəmō.iš°*." which is indeed true. The absence of such a spelling shows that the word *zəmōištuua* was not understood as a compound by the scribes. Duchesne-Guillemin writes: "[u]ne autre série a en 1er t[erme] un nom de matière. […]" and further quotes "*zəm.varəta-* « motte de terre » et l'obscur *zəmōištva* qui n'est même pas sûrement un c[om]p[osé]." (Duchesne-Guillemin 1936: 137).

The form *zəmōištuua* is not often found in manuscripts: Ave976 has *zəmōište*, (which stands for [+]*zəmōišti*), while all other Iranian manuscripts available in the Avestan Digital Archive[145] read *zəmōišti*. There is more variation in Indian manuscripts: some (such as ML630) read *zəmōišti* or *zimōisti* (M2), two manuscripts (G106 and Bh3) have *zimōištuuō* ("*vā zarštuue vā*"), B4 has *zimōistevā* ("*zarastavaevā*") G112 and Fires1 have *zəmōišta* ("*vā zruuašta*"), K10 has *zəmōištai*, corrected in *zəmōištuui*, ("*vā zarštuui vā*"). This kind of variation is common with hapaxes.

Avestan manuscripts with Pahlavi commentary are more unanimous: MU1, E10, G34, T44, L4, F10 *zəmōištuuō*, K1, Ml3, B1, *zəmōištuue*, M3 *zimōištuue*.[146]

Not a single manuscript has a separation point indicating a compound inside this word, and it is translated in Pahlavi as *zamīg gōn* 'earth coloured (thing), earthy (thing)' or *zamīgēn* 'earthen' and in Persian as "(on) the dry soil". Although much more common in Pahlavi manuscripts, the forms *zəmōištuuō*, etc. could be anticipated from the next word: *zarštuue*, and variants. Indeed: one notes that the ending often "rhymes" with that of the next word here, as can be seen from the list I made. For instance, *zimōistevā* precedes and rhymes with *zarastavaevā*, while the variant *zəmōišta* is followed by and rhymes with *zruuašta*.

If the word indeed designates dried bricks made of clay,[147] then it is unclear why it was not understood as a compound, since *zəmō-* is a normal first member of compounds. It is also unclear why the form was not *zəmō.ištiia-*. If the form *zəmōišti*, found in the Iranian manuscripts as well as in some Indian ones, is authentic, then the word is indeed a cognate of Old Persian

---

[145] Ave977/978, Ave991, Ave1001, ML16226, RSPA230, HM2, YL1, ML15283, VJ.

[146] The Pahlavi translation of F10 has ‹zmyg gw› (maybe originally [+]‹zmygyn›, since the points on the second ‹g› were added later) 'earthen, earth coloured', corrected by the Persian translator in ‹zmyg gwn̲› and translated in Persian as زمین گُون را *zamīn gūn rā*, while the interlinear Persian translation under the Avestan text has زمین گُون چیز را صاف کنید *zamīn gūn čīz rā ṣāf kunēd* that is "make you pure an earthen thing". The word was thus interpreted as a substantivized adjective.

[147] It is unclear to me what other types of bricks could be opposed to these clay bricks: golden bricks?

išti- 'sun-dried brick'. It seems to me that comparative grammar can help us establish that *zəmōišti* is indeed the original form in the Avestan passage.

On the basis of comparative evidence, (cf. section 3.2.1.2 of the current chapter) what is reconstructible for Indo-Iranian is **ḥišt-*, often, but not always, suffixed with **-i-*. The form **ḥišt-* could either derive from **ḥišt-* or from **ḥist-* with RUKI, which also applied to BMAC loanwords in Indo-Iranian (Lubotsky 2001: 304).

In conclusion, if I am correct, the Avestan word *zəmōištuua* and variants should no longer be cited as a cognate of Old Persian *išti-* and Vedic *íṣṭakā-* f. both meaning 'brick', but rather its variant *zəmōišti*,[148] found in most Iranian manuscripts, should be cited. This variant seems to go back straightforwardly to **zəmō.išti* that is, literally, 'earth-brick', but it is unclear why it was no longer understood as a compound by Avesta copyists and Avestan grammarians.

3.2.2 TB *eñcuwo*, TA *añcu** 'iron'

To the Tocharian B word *eñcuwo* 'iron' corresponds unattested Tocharian A *añcu** 'iron'. The latter can straightforwardly be extracted from the derived adjective *añcwāṣi* 'made of iron', cf. also TB *eñcuwañ(ñ)e* 'made of iron', with a rarer variant *iñcuwaññe*.[149] These words have remained difficult to etymologize. In the present discussion I will argue that this word is of Iranian origin. However, I chose to study it in this BMAC chapter because it has frequently been suggested that Tocharian B *eñcuwo* and its Tocharian A equivalent are of BMAC or substratal origin.

TB *eñcuwo* and TA *añcu** have no clear Indo-European origin. They have been connected to Iranian words of similar meaning, for example by Winter (1971: 222) who links them to Ossetic *ændon* 'steel'. Schwartz (1974: 409[33]) connects the Tocharian words to Khwarezmian *hnčw* 'spear-head', but suggests they are both of non-Indo-European origin. Isebaert (1980: 191-92) connects them to PIr. **spana-/*safna-* 'iron', deriving both from the Proto-Indo-European root **ḱu̯e* (sic) "from **ḱeu-n-/*ḱu̯-en-* 'to be

---

[148] See also Vendidad VIII:8 *ištiiehe vā zarštuuahe vā* where we find the element *ištiia-*.

[149] On which see Peyrot (2008: 60). He convicingly arguments that *eñcuwo* is likely to be the oldest form, and *iñcuwaññe* a later variant.

bright, to shine' ['stralen, schitteren']". I could not find this root in the LIV^2, probably as it is mostly found in nouns.

Tremblay (2005: 424) assumes an Iranian origin, and further assumes that the Iranian word itself is inherited from Proto-Indo-European. He writes: PIE *h₂ék-u̯on- 'cutting edge' > Old Sakan *anću̯an- (a "nasalized variant" of Old Sakan *aću̯an-), which was later borrowed into Tocharian.[150] This "nasalized variant" would have no reflex in Sakan languages, because Khotanese hīśśana- 'iron' shows no trace of the first nasal.

Later, Pinault (2006: 184-89) proposed to derive both the Tocharian words and Indo-Iranian *anću- (see Lubotsky, 2001: 304, 310), meaning 'Haoma, Soma plant' (the sacred plant and substance of both Zoroastrian and Vedic religions) from a substratic form *anću-. He argues that the substratum word designated a reddish, rusty colour, an attribute of both the twigs of the Haoma plant, and of iron.

Finally, Adams (DTB:84-5) proposes an Indo-European derivation of the Tocharian word (*h₁n̥-ǵʰeu̯eh₂(-n)- 'what is poured in'). He suggests that the word was borrowed from Proto-Tocharian into Proto-Iranian, but does not exclude a transfer in the other direction.

All of the etymologies stated above have their weaknesses: what Isebaert could not have known, since it was established much later, is that the Old Steppe Iranian reflex of Proto-Indo-European *-k̂u̯- is reflected in Proto-Tocharian as *tsw rather than *ñcw (see for instance etswe p. 26f.). This also contradicts Tremblay's proposal. As to Adams' proposal, it is improbable that the Proto-Tocharian form *eñcuwo spread to even one Iranian language (cf. Del Tomba 2020: 147), let alone so many, as he proposes. Indeed, this would be the only Proto-Tocharian word in Iranian. Furthermore, we see that the Iranians were in all aspects more technologically advanced than the Tocharians. That they would have borrowed a metal name from the latter is highly unexpected.

Together with Federico Dragoni and Michaël Peyrot, I have developed an alternative etymology in order to explain the Tocharian words for 'iron'. This etymology is presented in detail in Peyrot, Dragoni & Bernard (2023).

---

[150] According to him *anću̯an- (or rather a descendant thereof) was also borrowed by Khwar. hnčw 'spear-head', hnǰw 'iron-tip', because of the treatment of PIE *k̂u in Sakan languages. "The initial h in hnčw is either a 'cockney aspiration', as in hrs- 'bear' < *r̥ša- […], or a closer assimilation to *handāna- 'alloy'."

In short, we propose that Proto-Tocharian *eñcuwo derives from the Pre-Khotanese predecessor of Khotanese hīśśāna- 'iron'.

It was demonstrated that Khotanese -śś- may correspond, at an earlier stage, to Tocharian -ñc- (Chen & Bernard, forthc.). Indeed, the Tocharian word śāñcapo, which, as we show in Chen & Bernard, forthc., means 'mustard', corresponds regularly to Pre-Khotanese *śanźapa- (or *śanźapāna), which yielded Khotanese śśaśvāna- 'mustard'; compare the reconstruction of an Iranian pre-form *sinšapa- 'mustard' for the rest of Iranian, in Henning (1965: 44).

For the -ī- in the first syllable of the Khotanese word hīśśana-, we argue that it is due to an independently proven Khotanese sound change: when followed by a *yod* in the third syllable, accented *a* regularly became -ī-, as in Khotanese ysīrra- 'gold' < PIr. *jaranịa- (Skjærvø 2004: II, 331). We thus suggest that Proto-Khotanese-Tumshuqese *hám-ću̯anị̄a- became Pre-Khotanese *henśu̯anị̄a- first. This form *henśu̯anị̄a- was then borrowed as Proto-Tocharian *eñcuwañña or *eñcuwañño, with subsequent reinterpretation as an adjective: *eñcuwaññe became an adjective 'made of iron', from which the form *eñcuwo was then extracted by back-formation.[151]

We further suggest with caution that it was the speakers of early Khotanese who introduced iron into the Tarim Basin (cf. Peyrot, Dragoni & Bernard 2023, with references). At the time, these speakers of early Khotanese were perhaps part of the Aqtala Culture, and their language might thus perhaps be conceived of as the language of the inhabitants of Jumbulaq Qum, one of the major sites of this culture, as presented by Debaine-Francfort and Idriss (2001). This is compatible with the current archaeological evidence.

According to us, the Khotanese word is ultimately a derivative from *ham- 'together with' + *ću̯aH-, which would in the context of iron have meant 'strike iron'. Iron would thus be the metal that is "struck together".

In conclusion, Tocharian B eñcuwo and Tocharian A añcu* 'iron' derive from Pre-Khotanese *henśu̯anị̄a- 'iron'. This fits with the archaeological data, which indicates that iron was introduced from the west towards the east.

---

[151] Another possibility that we suggest is that the Tocharian forms were borrowed from a pre-Khotanese form *henśu̯ana- when the *yod* had already disappeared (cf. our paper for more detail).

3.2.3 TB *etre*, TA *atär* 'hero'

A number of proposals have been put forward to explain this word as Iranian (most were summed up in Isebaert 1980: 191). I will discuss these to determine if they could possibly justify to consider TB *etre*, TA *atär* as loanwords from an Iranian language.

Pinault (2006: 171-175, reproduced in 2020a: 381-82) proposes to link these words to the Indic word *átharvan-*, Avestan *āϑrauuan-* 'priest' (on the Avestan form see de Vaan 2003: 65). The Indic and Avestan correspondence is irregular: Sanskrit *-ar-* corresponds here to Avestan *-ra-*. This irregularity is typical of BMAC loanwords borrowed into Indo-Iranian, cf. Lubotsky (2001: 303).

Although Pinault's connection is tempting, it is not assured, since a priest and a hero differ in many respects. Furthermore, a priest is not associated with the same type of strength or guile that is most often associated with heroes.

Adams proposes that this word is from PIE *h_2ot-ro-*, related to TA *ātäl* 'man', which he then derives from *h_2et-lo-*. The semantic shifts 'man' > 'hero' and conversely, although attested, are not evident and require a certain cultural context to work. More importantly, it is far from assured that this root described by Adams can be reconstructed for Indo-European at all.

A derivation from the otherwise unknown Iranian root *aϑ-* 'be violent' (Bailey 1975: 7[16]), or from Ir. *atara-* or *ātara-* (cf. Winter 1971: 218-19) have also been proposed. However, the Iranian root *aϑ-* 'to be violent' does not seem to exist; in any case I could not find any valid cognate, or any traces in the scientific literature. I do not see which words *atara-* or *ātara-* (*atara-* 'one of both'?) Winter refers to, but without a cognate, or an explanation, his theory remains weak.

Pinault's proposal that this word TB *etre*, TA *atär* 'hero' is of BMAC or Central Asian origin is the most convincing proposal made for these words. Whether one agrees with Pinault's connection of these words with Vedic *átharvan-* and Avestan *āϑrauuan-* / *aϑaurun* 'priest' or not depends on one's view of the underlying notions that make up the character of a priest and that of a hero. One could perhaps think of an original meaning *'leader': with the semantic shift *'leader of the sacrifice' on one hand, and the semantic shift *'leader' > 'hero' on the other hand. The latter shift has a parallel in Irish *nēath* 'hero' from the root *ni-* 'to lead' (cf. Buck 1949: 712). Even if the

connection of PT *etre* to Vedic *átharvan-* and Avestan *āϑrauuan-* were to be rejected, it would not preclude a BMAC origin for this word. In lack of a convincing Indo-European or Iranian cognate, this remains the most plausible hypothesis to date.

In conclusion, Tocharian B *etre* and Tocharian A *atär*, which go back to Proto-Tocharian *etre*, are probably of BMAC origin, as proposed for the first time by Pinault (2006).

3.2.4 TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric'

Pinault (2001: 128-129)[152] has shown that the Tocharian A word *kanak*, and its Tocharian B cognate *kenek*, correspond to Old Uyghur *böz* 'Baumwollstoff'. He established its exact meaning as being 'étoffe de coton' (2001: 129). It has variously been proposed that these Tocharian words derive from an Iranian form *kanaka-* (Pinault 2001: 129; Isebaert 2003; Tremblay 2005: 425; Peyrot 2018: 270-71).

However, there is – to my knowledge – not a single Iranian form that goes back to *kana-ka-*. Almost all attested Iranian forms either go back to *kanā-*, for instance Ossetic Digor *gænæ*, or to *kan(a)ba-* ~ *kanafa-*, as do Khotanese *kaṃha-* 'hemp', New Persian *kanaf*, Khwarezmian *knbynk* 'linen', Sogdian *kynpʾ* 'hemp, flax').[153] On the basis of the attested forms, *kanaka-* cannot be reconstructed for Proto-Iranian or Old Iranian.

Peyrot (2018: 270) noted this difficulty. He further wrote (2018: 271) that "in view of the TB vowels *e_e* for Iranian *a_a*, the borrowing must be relatively old; a Proto-Tocharian reconstruction *kenek* would theoretically be possible. The word is clearly a Wanderwort originally from the Middle East so that a unified reconstruction for Iranian cannot be given […]."

It is important to examine this hypothesis, and to see whether the origin of Proto-Tocharian *kenek* is Iranian or not, and, if not, what else it could be. If it is Iranian, it should indeed come from Old Steppe Iranian, because of

---

[152] Although Pinault (2001: 129), more specifically, connected it to Khwarezmian *kcynyk* 'Seidenstoff', deriving the Tocharian word from a form *kcenek* < *käcenek*. For a criticism of this etymology, see Lubotsky & Starostin (2003: 260).

[153] According to Gharib (SD: 203), the Sogdian word was borrowed from Syriac *gnbʾ*.

the vocalism *e* for Iranian *a. It has been proposed that the Iranian proto-forms **kanā-*, **kanaba-* ~ **kanaf-* were borrowed from Uralic **kana-* 'hemp (narcoleptic)' (Katz 2003: 143f.), but this is doubted by Kümmel (2020: 255) who believes that both the Iranian and the Uralic forms were borrowed from the same source.[154] Katz (2003: 143) further suggests that "Alanic *kænæ" was borrowed from Mari *kəne.

If Katz' hypothesis of a Uralic origin is nevertheless correct, an Old Steppe Iranian form **kanaka-*, which would be a simple *-ka-* enlargement on this **kana-*, could have existed. Nevertheless, an Old Steppe Iranian form **kanaka-* would have yielded Proto-Tocharian **kenke TB †*keṅke* TA †*kaṅk* as per the established syncope rules (see section 2.6.2.g). Thus, even if such a word ever existed in Iranian, it could not have yielded TB *kenek* and TA *kanak*.

Indeed, as noted above, this word is a typical Wanderwort, found in some Semitic languages, in Proto-Germanic **hanipa-* 'hemp' (cf. Šorgo 2020: 440), Proto-Slavic *konopʲa, Greek κάνναβις 'hemp', but also in Kartvelian (Svan *kan* 'hemp') and in Abkhaz *akonə* 'hemp', etc. Witzel (1999: 55) proposed that Greek κάνναβις, Old High German *hanaf*, Dutch *hennep* all were borrowed from Scythian. This is perhaps possible (although one would need to explain the gemination of the Greek), but in that case the Scythian word needs to have been different from its Ossetic cognate *gænæ*.

As discussed below (see section 3.2.1.2), the *-ka* suffix that was observed by Lubotsky (2001: 304) in Indo-Iranian words of BMAC origin, is also found in a number of BMAC Tocharian words. A number of words presented in this chapter share the structure *CaCáCa*, unlike most Old Steppe Iranian words seen in the previous chapter, which rather appear to have a *CáCaCa(Ca)* structure.

As mentioned section 2.6.2.g, Proto-Tocharian got rid of trisyllabic loanwords with identical vowels through the apocope of the middle syllable (type OIr. **rataka-* → PT *retke 'army'). I have also tried to show that Old Iranian loanwords into Tocharian, with a few explicable exceptions, had fixed first-syllable stress. In my view, it is possible that fixed middle syllable stress prevented such a reduction in this word: while **rátaka-* could easily

---

[154] Although the forms starting with *g-*, such as Ossetic Digor *gænæ* Iron *gæn* 'hemp', Kabard *ǵanä* 'shirt', would fit well with a language that does not have a voiced/unvoiced phonological opposition.

become *retke, it would have been more difficult for *kanáka- to become †keṅke. Inherited examples of a similar syncope rule are found in, for instance, TB *sonop-* < *sonopo- and TB *tsetserñu* < *tsetsereññu. This is why I think that the apparently unsuitable structure with three syllables with identical vowels was in this case resolved differently, namely by dropping the final vowel, yielding PT *kenek, which does not otherwise receive an easy explanation. Thus, on the basis of the apparent accentuation of the word and of its formation (with the BMAC nominal suffix -ka-), I believe that this word was borrowed, not from an Old Iranian language, but from the source language I call BMAC for convenience.[155]

For the semantics, I suggest that the word originally meant 'hemp cloth', and, possibly because cotton was more common in the Tarim basin or in Tocharian material culture, the meaning of these words shifted to 'cotton cloth' in Proto-Tocharian, or otherwise independently in Tocharian A and B.

In conclusion, Tocharian B *kenek* and Tocharian A *kanak*, both going back to Proto-Tocharian *kenek, cannot be of Iranian origin, simply because there is no plausible Iranian form from which they could derive. Even an Old Steppe Iranian *kanaka- should have yielded a different result, namely, Proto-Tocharian *kenke. We could think of a consonant stem *kenek- or *kanak-, but there is no support for it among Iranian languages. Instead, the word is more likely to come from the Central Asian BMAC language, the vowels of which were rendered in Indo-Iranian as *a and in Tocharian as *e and *ə; which also possessed a suffix *-ka-, and which seems to have had second syllable stress. These features together converge towards a possible borrowing from BMAC *kanaka- 'hemp' into Proto-Tocharian at first as *keneke, which became *kenek through apocope.

---

[155] Carling (2005: 55) writes that this word is "obviously a Turkish borrowing". I cannot see how this is obvious, especially since TB *kenek* and TA *kanak* do not have a Turkish structure or appearance, and, in my view, no certain Turkish words in Proto-Tocharian have been recorded to this day (the best candidate would be PT *kaun 'sun, day', TB *kauṃ, TB *koṃ 'id.', but even this case is controversial, cf. Lubotsky & Starostin 2003: 257f.).

3.2.5 TB *kercapo* 'donkey'

The etymology of Tocharian B *kercapo* 'ass, donkey' (Archaic TB *kerccäpo*, in THT 343 b7), also *kercapaññe* 'pertaining to a donkey', *kercapiśke* 'PN' (with a name-forming affectionate diminutive) is not evident. This word is mentioned in multiple works, but I will only cite two extensively, Winter (1971) and Adams (DTB), because they represent two important and common views concerning this Tocharian B word.

Winter (1971: 222) writes:

"Urverwandtschaft mit aind. *gardabha-* ist ausgeschlossen; das *-c-* ebenso wie das *-a-* der tocharischen Form würden unerklärt bleiben. Gleichermaßen unmöglich ist die Annahme einer direkten Entlehnung aus dem Indischen: weder *-e-* noch *-c-* noch *-o-* wären begründbar. Anderseits kann B *kercapo* kaum auf ein echtiranisches Wort zurückgehen: aind. *gardabha-* scheint keine iranischen Entsprechungen zu haben. [...]"

He further proposes that Sanskrit *gardabha-* was borrowed into an Iranian language, which should also be the source language of Tocharian B *eñcuwo* 'iron', and possibly also Tocharian B *witsako* 'root'. He suggests that this Iranian language is close to Ossetic ("scheint dem heutigen Ossetisch sehr nahezustehen").

Adams (DTB: 210) adds:

"[r]eflecting a PTch **kercäpā-*, which, except for the stem class, is the exact equivalent of Sanskrit *gardabhá-* (m.) 'donkey, ass' (< **gordebho-*) with the same **-bʰo-* which appears in other Indo-European designations of animals […]. If, as has so often been suggested […], *kercapo* is a borrowing from Indic *gardabhá-*, the borrowing must be very early, before the merger of the non-high vowels in Indo-Iranian (otherwise we would expect **kertepo* or the like with the first and second vowels alike and no palatalization). […] Anreiter (apud Thomas, 1985b: 134) suggests that the Indic and Tocharian words are both borrowed from some third (and unknown) language."

Winter's argumentation is weakened by the fact that no cognate of Vedic *gardabhá-* has been found in Iranian, as he himself admits, and the idea of this borrowing is completely *ad hoc*. Besides, it is very unlikely that such a

language would have had -*č*- where Indic has -*d*- (cf. Pinault 2008: 394). Further, I believe that it has been demonstrated (s.v. *eñcuwo*; cf. Peyrot, Dragoni & Bernard 2023) that *eñcuwo* is a pre-Khotanese loanword into Tocharian. Therefore, Winter's hypothesis is no longer acceptable. As to Tocharian B *witsako*, s.v. Adams's interpretation is flawed as well, because of the simple fact that no other borrowing from Pre-Proto-Indo-Iranian, before the vocalic merger of **e* and **o* into **a*, into Proto-Tocharian is known so far. In Adams (2017: 1368) we also find the supplementary proposition that Tocharian B *kercapo* and Vedic *gardabhá-* are both inherited from Indo-European. No other cognate can be added to this comparison, and it is very difficult to see how it would have functioned formally (cf. Pinault 2008: 394).

Anreiter's proposal is more enticing. Tocharian B *kercapo* cannot derive from Sanskrit or any Indic language, for the reasons evoked by Winter, namely: the vocalism and consonantism do not match. They can hardly both be inherited, as no other cognates are found, and, more importantly, as the phonetics do not match either. Furthermore, *gardabhá-* is not necessarily analyzed as **garda-* + *-bha-* (< **-bho-*), although one can suppose that the element **garda-* is due to a secondary recharacterization (see below). Below, I will investigate Anreiter's proposal, and try to propose a scenario for the borrowing.

An obvious candidate for Anreiter's third language is the BMAC language, which counts a number of words "borrowed independently by Common Tocharian and Indo-Iranian in the late Bronze Age", to cite Pinault (2006: 170).

However, precisely on this point, Pinault (2008: 392f.) has a different opinion. He believes TB *kercapo* to be related to Tocharian B *koro**, which he translates as 'mulet'. Pinault takes *koro** as deriving from Old Steppe Iranian **xara-*. According to Pinault, the expected pre-form **kere* 'mulet' would have taken the *-o* ending, in analogy with words such as *okso* 'oxen', and, as Pinault suggests, would have undergone umlaut **kero > koro*. I do not know of any parallel to such an analogy. To note, only Old Steppe Iranian *ā*-stems, not *a*-stems, were made into *o*-stems in Tocharian (see previous chapter) and no Old Steppe Iranian word shows umlaut of **e* to **o* in To-

charian B.[156] According to Peyrot (2016: 328), there is no *o*-umlaut of To-charian B *e* at all. Pinault's etymology of *koro** can thus not be accepted.

Pinault goes on to propose that the element *ker-* in *kercapo* also derives from **xara-*. Since donkeys are "shouting" animals, he suggests that *gar-dabhá-* was reanalyzed within Sanskrit as being related to the root *gard-* 'to shout'. The Tocharians, according to Pinault, would have taken this as an example and built the word *kercapo* on the basis of the Sanskrit form: he thus sees in *capo* a reflex of the root *tap-* 'proclamer, annoncer à haute voix', which has palatalized allomorphs.

This etymology leaves us with too many difficulties: if Tocharian B had a form *kere, borrowed from Old Steppe Iranian **xara-*, why would it not have made a compound word **kerecäpo instead? The form *koro**, can cer-tainly not be Old Steppe Iranian in origin, for the reasons evoked above. If *koro** was used, we would expect ***kor(o)cäpo* instead. An Iranian etymol-ogy for *koro** can perhaps be suggested: either a word related to Sogdian *γwr* 'wild ass', or to PIr. **xara-* 'donkey', through Khotanese *khara-* 'donkey' : acc.sg. *kharu* → TB *koro**. If the meaning of *koro** was 'camel' as tentative-ly suggested by Adams (DTB: 218), then these Iranian etymologies are im-possible.

A more important argument against Pinault's proposal, perhaps, is the fact that the very existence of the root **gard-* 'to shout' is doubtful (cf. EW-Aia[1]: 493). As professor Lubotsky informs me, the traditional etymology of Vedic *ráśabha-* 'ass; donkey' from *ras-* 'to shout' is also very doubtful: the root *ras-* occurs very rarely, and the form *ráśabha-* implies a form **ráśa-* 'screamer', with wrong accentuation (instead of expected **rāśá-*).

I propose to see in Tocharian B *kercapo* 'donkey, ass' and Vedic *gar-dabhá-* 'idem' two different reflexes of a BMAC word: possibly the *-d-* in Sanskrit and the *-c-* in Tocharian go back to a single phoneme, perhaps a palatalized *d*, or perhaps they both go back to BMAC **gardeba-* or **gardepa-*, with inner-Tocharian palatalization.

The ending *-bhá* of Sanskrit can be analogical after multiple other animal names. It is also possible that it reflects the pronunciation of a BMAC pho-neme, rendered as *p* in Tocharian. Perhaps even, the original BMAC word

---

[156] Witzel (1999: 54) has also tried to connect *kercapo* "ker-ca-po" to **khara-* ~ *xara-* 'donkey', and mentions the Proto-Dravidian form **garda-* 'donkey'.

was something like **gardyapa-* or rather **gardepa-*, with the BMAC *-pa* suffix, cf. Lubotsky (2001: 305). The suffix *-pa* was then replaced in Indic by the common animal names suffix *-bhá* found in, for example, Vedic *vṛṣabhá-* 'bull'.

As to the final *-o* of the Tocharian form, it is quite commonly found in substratum words: *mewiyo* 'tiger', *peñiyo* 'splendor', *witsako* 'root', and is most probably due to them being remade according to the morphological classes of Tocharian nouns.

Professor Lubotsky has also suggested to me a possible connection of the BMAC etymon **gardepa-* 'donkey, ass' to BMAC **gṛda-* 'penis' (Lubotsky 2001: 307), certainly due to the fantastic size of donkey male instruments.

In conclusion, Tocharian B *kercapo* 'donkey, ass' seems to be related to Vedic *gardabhá-*, of identical meaning. There is no consensus as to the etymology of the Tocharian B form. Following Anreiter, I propose that both the Sanskrit and the Tocharian words derive, with different phonetic adaptations, from a common substratal (BMAC) source.

3.2.6 TB *kroṅkśe* TA *kroṅśe* 'bee'

For the Tocharian B word *kroṅkśe* 'bee' (with variants such as *krokśe* and *kroṅśe*), borrowed into Tocharian A as *kroṅśe*, several different etymologies have been proposed. We will examine them below.

First, one can cite Isebaert (1980: 148) who proposed to explain the Tocharian word as related to Sanskrit *krauñca-* 'crane'. He saw it as a contamination of **kroṅke* and **kroṅśe*, from Prakrit **krŏṅ* and **krŏñca-*, hypothetical forms which would be, according to Isebaert, related to "Sanskrit *kruṅ*" and *kruñca-* (respectively). As to the semantics, it is according to him related to an onomatopoeic root "**kruñc-*, *kūj-*" 'agreeable to listen to' (which Isebaert cites from Thieme 1974: 295). This is ultimately the etymology I will follow, although differing in the detail.

Later, Hilmarsson proposed (1986: 34f.) to connect the Tocharian word *kroṅkśe* to the Germanic word for 'honey', Old Icelandic *hunang*, OHG *honang* < **hunanga-* (cf. Kroonen 2013: 255-56, who connects Greek κνῆκος f. 'safflower' to the Germanic root, cf. Beekes 2010: 722-23). In Germanic we find an alternation between **hunanga-* (OHG *honang*, Dutch *honing*, etc.) and **hunaga-* (e.g. English *honey*, Finnish loanword from Germanic *hunaja*).

Hilmarsson (1986: 35-7) further proposes to connect the Germanic and Tocharian words to Sanskrit *kāñcaná-* 'gold, golden' and Lat. *canicae* 'bran', among others. As Hilmarsson reminds us (1986: 35), the second nasal in the Germanic words has often been viewed as secondary by previous scholarship. The variant of the type **hunaga-* was often considered primary. The reason evoked by scholarship was generally that the common Germanic suffixes *-ing-* and *-ung-* influenced **hunaga-* and that it thus produced **hunanga-*. Hilmarsson objects to this view that the suffix *-aga-* was productive in Proto-Germanic, so that the forms of the type *hunanga-* need to be original. Hilmarsson thus convincingly argues that no influence from the suffixes *-ing-* and *-ung-* has to be assumed for the Proto-Germanic word for 'honey'.

Hilmarsson further postulates a Proto-Indo-European form **kn̥Honko/e-* (p. 36) which would be the source of the Germanic root **hunanga-*, of Sanskrit *kāñcaná-* and of Pre-Proto-Tocharian **knonko-* > Proto-Tocharian **krænko* > **kronko*. Finally, "the *ōn*-suffix was apophonic, its *e*-grade causing palatalization of preceding susceptible consonants. Through generalization of the *o*-vocalism and the palatal *-ś-* the attested Tocharian forms were reached." Hilmarsson also explains the retention of *-k-* as ("perhaps") a reflex of the velar pronunciation of the *-ṅ-*.

Hilmarsson's arguments concerning the etymology of *kroṅkśe* do not seem very compelling to me. First, it is difficult to believe that both generalization of the palatalization due to the *e*-grade of the *-ōn* suffix and generalization of the *o*-vocalism took place. If this word were really suffixed with *-ōn* we would not expect a final *-e* but a final *-o* (as demonstrated notably by Hilmarsson 1987). The retention of *-k-* as a reflex of the velar pronunciation of the *-ṅ-* seems completely *ad hoc* and without parallel. How could the **n* have remained velar after the palatalization of the *k*?

I therefore have to agree with Adams (DTB: 235) that Hilmarsson's demonstration requires "some very complicated phonological changes in Tocharian". Besides assuming an *ad hoc* dissimilation of *n* to *r* in **knonko- > *krænko*, **krænko* (**krenko* in our notation) should have become †*kreṅke* in TB. The retention of *-k-* as a reflex of the velar pronunciation of the *-ṅ-* before a *ś* is completely unprecedented in Tocharian.

I also disagree about the semantic connections made here: firstly, the connection with Sanskrit *kāñcaná-* 'gold, golden' is doubtful, as honey is not always yellow. To support it there needs to exist a certain Indic phraseology

connecting honey with gold, or with a golden colour. No such parallel is adduced by Hilmarsson.

Now, bees too can be yellow, at least partially, but they would not be called 'the yellow ones' (perhaps, at the most 'the yellow insects'). There does not seem to be another parallel, at least not among Indo-European languages (cf. Buck 1949: 192).

Hilmarsson is right in that bees are often lexically associated with honey. He cites English *honey-bee* and Icelandic *hunangsfluga* as examples. One can also adduce the Chinese example *mìfēng* 蜜蜂 [honey-bee] 'bee', and Greek μέλισσα 'bee' < *μέλιτ-i̯α 'the one of the honey'. It should nevertheless be noted that compounds like English *honey-bee* are only needed if in that language the word for bee is also used to designate related insects that do not produce honey. English, for instance, calls a number of insects *bee*, such as the bumble-bee (French *bourdon*). In French, for example, *abeille* only designates honey producing insects, e.g. the definition from the Nouveau Petit Robert (2007: 4) "abeille, n.f. […] Insecte social hyménoptère (*apidés*) dit *mouche à miel* vivant en colonie et produisant la cire et le miel […]."

As a French speaker, I would thus find it absurd to add the word *miel* 'honey' to *abeille* 'bee'. Some dialects that do not have the word *abeille* call them *mouche à miel* 'honey fly' (e.g. Bourbonnais [muʃ a mjɛl]), identical to the Icelandic compound evoked by Hilmarsson. As Peyrot and Meier (2017: 11) write: "although 'bee' and 'honey' are obviously contingent semantically, a direct change in meaning from the one to the other is not at all trivial." According to these observations Hilmarsson's etymology of the Tocharian word for 'bee' needs to be abandoned, both on formal and semantic grounds.

In my opinion, Isebaert was partially correct in seeking a connection to Sanskrit *krauñca-* 'crane'. In Bernard (2020: 33f.) I proposed, on the basis of a wide range of Indic and Iranian comparanda, to reconstruct a substratum root **kra̯u-* alternating with **kru-*, variously suffixed in Indo-Iranian to designate loud animals. The original sense of all suffixed forms seems to be 'loud animal', since that is the smallest common denominator of all the species of animals derivatives of this root designate. One of these suffixed forms contains the suffix *-ñc-* added both to the base **kra̯u-* and to the base **kru-*. I also suggested that this suffix is of substratal origin (2020: 34). This suffix can be seen in Skt. *krauñca-* and *kruñc-* 'crane'.

It is known that Tocharians were very familiar with honey and bees, since Chinese borrowed its word for 'honey' from Tocharian (see for instance Behr 2001: 359). I can see two reasons for calling bees the 'loud animal' *par excellence*. The first one is evident: groups of bees make a very loud continuous noise. This noise can be deafening, especially if one gets too close to a beehive. The Hebrew word דבורה (*dḇōrā*) 'bee' is probably related to the verb 'to speak' (root *d-b-r*).[157]

The second possibility is that the word initially designated the fly, the sound of which is not even necessary to describe. The semantic change 'fly' > 'bee' is attested. For example in most Normand dialects *mūk* (< Lat. *musca* 'fly') is used to designate bees. It is nevertheless unclear to me whether this meaning of *mūk* is likely to be back-formed on the compound *mūk a myèl* 'honey fly', also attested in Normandy. This weakens very much the latter hypothesis. Another possible example is Proto-Iranian *makš-* 'fly', borrowed into Proto-Finno-Ugric as *mekše* 'bee' (cf. Holopainen 2019: 139f.; van Sluis 2022: 5).

The former hypothesis somewhat finds an echo (although chronologically much later) in Buddhist phraseology, even in Tocharian: one can think of THT 571b4 *krokśāṃts weśeñña māka* "the voice of many bees". In my opinion PT *kroṅkśe* designated the swarm of bees initially, and later came to designate the bee in general, and the individual bee in particular. Such a semantic development is attested, for example in Rumanian *albină* 'bee' < late Lat. *aluīna* 'beehive' (cf. Buck 1949: 192).

The Tocharian combination of *ṅk* and *ś* occurs in this word only. Furthermore, only one other Tocharian word contains the combination *k + ś*: *lyekśiye* 'millet', and it has no known etymology, cf. Peyrot (2018: 245). This naturally suggests a borrowing, rather than inheritance. I believe that in TAB *kroṅ(k)śe* the cluster -*ṅ(k)ś*- (which is variously written *ṅkś ~ ṅś ~ kś*) is a (Proto-)Tocharian rendering of the "substratal" sound which Sanskrit writes as ‹ñc›.

If one accepts this etymology which, as with all these non-Indo-European matters, cannot reach a high level of certainty, then Tocharian B *kroṅkśe*

---

[157] I thank Dr. Benjamin Suchard for informing me of this parallel. If this word is a cognate of Aramaic *zibbūr* 'hornet', Classical Syriac *debbōrā* 'wasp, hornet' and Arabic *dabbūr ~ zanbūr* 'id.' (all from Proto-Semitic *ᵈzanbūr-*), then it was perhaps remotivated on the basis of the root *d-b-r* 'to speak'.

must go back to a proto-form *krVÑCa- 'loud animal'. I write ÑC in capital letters to indicate that sound which is rendered as -ñc- in Sanskrit and <ṅ(k)ś> in Tocharian. One can evoke Armenian kṙunk 'crane', also of difficult etymological derivation (Martirosyan 2010: 377), however Armenian k- points to *g- and this word must thus be left out of the discussion.

As to the -o- corresponding to -u- in the Sanskrit and Armenian, it is not an expected correspondence. Nevertheless, since we do not know the original value of the "substratal" vowel, it is possible that it was neither *o nor *u but a sound in between. Furthermore, perhaps the alternation krauñca- ~ kruñc- also found as *krau̯- ~ *kru- in Iranian […] could reveal different strategies to adapt that original substratal phoneme.

Since the etymology of krauñca- and of kroṅkśe relies, I believe, on the notion that these were noisy animals, one could think that they were independently built on an onomatopoeia "kronk" or the like. This onomatopoeia, however, is very far away from the type of noise made by cranes or bees, and rather reminds us of a falling or breaking object.

As a conclusion, I wish to underline that, if the Tocharian and Indic forms are borrowed from a non-Indo-European language, it is possibly a different language than the one which yielded most other words of this list. The structure is disyllabic or perhaps monosyllabic with a thematic vowel, cf. Sanskrit krauñca- ~ kruñc- is quite different from that of other words presented and discussed here, which are mostly trisyllabic with second (middle) syllable stress, and usually contain a schwa. It is also unusual to see this word reflected in the Indic branch and apparently in Tocharian, without any reflex in Iranian. As far as I know, there is indeed no Iranian correspondent to the Sanskrit substratal -ñc- suffix. However, other reflexes of the root *krau̯- are found in Iranian languages (cf. Bernard 2020: 31f.) and a BMAC origin cannot be excluded for all these words.

3.2.7 TB witsako 'root'

The Tocharian B feminine noun witsako 'root' is of great importance. For a long time it has been claimed to be of Iranian, and generally more precisely of Old Iranian origin. The first scholars to suggest an Iranian origin for witsako were Karl Bouda (apud Krause 1971: 37) and Winter (1971: 222),

and they were followed by many others.[158] However, with the establishment of the features of Old Iranian loans throughout the present work, and in particular the first chapter, such claims should be thoroughly reviewed. This is what I intend to do below.

Indeed, the *communis opinio* about the Tocharian B word *witsako* 'root' is that it is a borrowing from an Iranian form, either from a (pre-)Proto-Ossetic preform of Ossetic Iron *widag* 'id.', or from a form closely related to it. The first to propose this was Winter (1971: 222); it was then accepted by most if not all scholars (see for example Abaev 1989: 106; Hilmarsson 1986: 227; Kim 1999: 124; Adams DTB: 658; Del Tomba 2020: 130).

The Ossetic word *widag*, and its Scythian cognate Βιδάχης, have been connected before that to Av. *vaēⁱti-* 'willow, willow-twig' (cf. NP *bēd* 'willow') by Abaev (1949: 186), and every other scholar has since repeated this. It seems that the Avestan word for 'willow' should rather be seen as a cognate of Ossetic Iron *xæris*, Digoron *xærwes* 'willow', perhaps from **xara-* 'donkey' and Iron *wis*, Digoron *wes* 'rod; copse, brushwood', which Morgenstierne (1938: 264; 1942: 269) derives from **u̯aitsa-*. Donkeys enjoy stripping the bark from willows, and they also enjoy eating willow in general.[159] Cheung (2002: 51), however, connects the Ossetic words for 'willow' to *xæræ* 'grey', which I find surprising because willows are not grey.

To explain the Tocharian B sequence *-ts-* in this word, it is mostly assumed that the Tocharian form derives in some way from an Old Iranian **u̯aitikā-* or **u̯aiti̯akā* or the like, although it is also usually noted that these forms would not straightforwardly yield *witsako* (e.g. Isebaert 1980: 97, who suggests that TB *witsako* is a borrowing from its unattested TA cognate).[160] Some scholars, such as Winter (1971: 222), Tremblay (2005: 426) and Adams (DTB: 657-58), propose that the Tocharian B form derives directly from a Middle Iranian **wīδikā-*, itself derived from **u̯aitikā-*.

Kim (1999: 124-126) proposes that the Proto-Ossetic predecessor of Iron *widag*, Digoron *wedagæ*, which he reconstructs as **wędəga*, was borrowed

---

[158] Although this idea is not mentioned as originating from Bouda in the scholarly literature.

[159] See for example https://donkeywise.org/2017/08/01/what-treats-can-i-give-my-donkeys/ (consulted on the 7th of October 2021).

[160] So far, there are no commonly accepted Tocharian A loanwords into Tocharian B, which makes this hypothesis very unlikely.

into Tocharian before the assibilation of *d* to Proto-Tocharian **ts*. It would subsequently have undergone devoicing from **dz* to **ts*. At the same time, he derives this Proto-Ossetic **d* from Proto-Iranian **t*. While somewhat ingenious, his reasoning does not work. Proto-Ossetic **d* would not be reflected by a **ts* in Proto-Tocharian. For one thing, Old Steppe Iranian **d* corresponds to Proto-Tocharian **t*, for example in Old Steppe Iranian **paribanda-* → Proto-Tocharian **perpente* 'burden' (s.v. TB *perpente*) or in Old Steppe Iranian **spaldaka-* → Proto-Tocharian **speltke* > TB *spel(t)ke*, TA *spaltäk* 'zeal' (s.v.). It never corresponds to Proto-Tocharian **ts*. It would thus be very difficult to imagine how Proto-Ossetic **d*, which would certainly be later than Old Steppe Iranian, could yield Proto-Tocharian **ts*.

I believe it is now necessary to briefly discuss the etymology of the Ossetic words and related Iranian words. Iron *widag*, Digor *wedagæ* point to Proto-Ossetic **wedagæ-*. These Ossetic words are cognate with Sogdian *wytʾk* 'string' and Yaghnobi *wita* 'cord' (Cheung 2002: 242), but also with Pashto *wuláɣ* f. 'root, root-fibre', which, however, Cheung (2017: 42) cites as a possible Ossetic loanword into prehistoric Pashto. In any case, the forms cited above point to a reconstructed form **u̯aitāka-*, which is incompatible with Tocharian B *witsako* (/witsə́ko/, see below) because this rather points to a short middle vowel.[161]

These phonetic details concerning the vocalism and the consonantism of *witsako* disagree with the theory that TB *witsako* 'root' is borrowed from Iranian. Besides, I believe that the Iranian forms cited above (Iron *widag*, Digoron *wedagæ* and their Sogdian, Yaghnobi and Pashto cognates) need to be separated from Avestan *vaēⁱti-* 'willow, willow-twig' and New Persian *bēd* 'willow', due to their semantic distance: a willow has roots, and roots can be willow-roots, but the semantic proximity does not get any closer.[162]

---

[161] An interesting form is provided by Wanjī *wisk-* < **u̯aitika-* in the toponym *Wiskroɣ* 'grape vine(yard?)' (cf. Lashkarbekov 2008: 76), with -*roɣ* being a borrowing from Tajik *roɣ* 'field'. Naturally, this cannot be a descendant of the donor form of Tocharian B *witsako*, for formal reasons (the intermediary form between **u̯aitika-* and *wisk-* is reconstructed as **viϑ-k* by Lashkarbekov) as well as semantic ones. On the other hand, grapevine is quite different from willows, and perhaps resembles roots, in the way that it develops and grows. This topic is quite intricate.

[162] Although one could argue that the original meaning is 'string, cord' (as in Sogdian) which is somewhat closer.

Furthermore, going back to phonetics, I have established that Old Steppe Iranian *-ai̯- is reflected in Proto-Tocharian as *-ey- (s.v. *waipecce*), preserved in Archaic Tocharian B as *ey* (TB *eynāke*, *meyyā*), and reflected as Tocharian A *e* and classical Tocharian B /ay/. Old Steppe Iranian **u̯aitāka- would thus become Tocharian B **waitke*, Tocharian A **wetäk*, or perhaps Tocharian B **waitake* /wáytake/, Tocharian A **waitak*. In any case, it would not have become Tocharian B *witsako*.

Moreover, as is well known, the Tocharian word is actually /witsə́ko/, as can be seen, for example, in PK AS 9Ba6 *witsäko*. As I hope to have shown, Old Steppe Iranian had fixed initial accent, with a few exceptions (see section 2.6.2.g). This accent was reflected in Proto-Tocharian. If one wishes to make the word "Proto-Ossetic" or the like, this does not solve the problem, as Ossetic also had fixed initial accent (Cheung 2002: 123). The Tocharian B word *witsako*, with its middle syllable stress, therefore does not seem to be of Old Steppe Iranian or Ossetic origin.

A Middle Iranian language could not serve as a better source, since it would imply both the preservation of initial *u̯-, which excludes Khotanese, and a sound change *ē > *ī, which would exclude all known Middle Iranian languages but Khotanese and Tumshuqese. However, most importantly, it would not be of any use to explain the aberrant Tocharian -ts- corresponding to Old Iranian *-t- and Middle Iranian *-d-. This is because we know that Proto-Tocharian *ts had already arisen when Tocharian was in contact with Old and Middle Iranian languages, and Tocharian simply devoiced Iranian *d to *t in all cases, while it reflects Old Steppe Iranian *ts and *dz (< PIr. *ć and *ǰ) as *ts. All the Iranian words that are close in meaning and form to our words for 'root', for instance Pashto *wulə́y* 'root, root-fibre', go back to a *d or *t.

One could think, however, that the etymon *u̯aitsa-, suffixed with -ka-, could have yielded the relevant Tocharian form. It has yielded, as mentioned above, Ossetic Iron *wis*, Digor *wes*, *yes* 'rod; copse, brushwood' (cf. Morgenstierne 1942: 269). These Ossetic words are, however, far removed semantically, and this would also imply that the *-sa- in *-saka- would first become a schwa in that given language. However, the stress would then remain on the first syllable since it is difficult for an *a vowel in Iranian to both be weakened to schwa and take the accent. Both obstacles, coupled with the fact that the Ossetic words are quite isolated in Iranian, render this etymology for Tocharian B *witsako* quite unlikely.

The Iranian words cited above with the meanings 'root' or 'string, cord' have not yet found a satisfying Indo-European etymology, and lack identifiable Indo-European cognates. It thus seems that they are potentially borrowings themselves. One could think of a Pre-Khotanese borrowing, but the expected cognate is not attested in Khotanese or in Tumshuqese, and furthermore, the accent would also be a problem as Tocharian words of Khotanese origin usually show initial accent (Dragoni 2022).

No known language can be mentioned as a potential donor, but in this particular case, the BMAC language could be evoked. Other substratic words designate flora, for example *u̯r̥ćša- 'tree' (Lubotsky 2001: 313). An interesting point here is that the Iranian forms mentioned above all point to a shared proto-form *u̯aitāka-, as if the -ka- was part of the root itself. This suffix -ka- cannot be understood synchronically at any stage (as far as our knowledge goes) as an addition to an otherwise known word *u̯aitā-, mostly because such a noun does not otherwise exist. This makes it plausible that it is the BMAC suffix *-ka- described by Lubotsky (2001: 304).

It is also possible that both the Iranian and Tocharian words derive from the BMAC language. Other such examples can be found, as Pinault (2003; 2006) proposed, corroborated by further examples (s.v. *kercapo*, *kroṅkśe*). In the case I am right and it is a BMAC loanword found both in Iranian and in Tocharian, one can think that, in the same way that BMAC *-dy- became Proto-Tocharian *-c- in TB *kercapo* 'donkey, ass' (s.v.), BMAC *-tV- could have become Proto-Tocharian *-ts-, for instance if the following vowel was *-ē-. One could, very speculatively and very cautiously, suggest a BMAC form *u̯ai̯tēka- 'root'. Alternatively, perhaps the form was originally *u̯ai̯ti̯āka- or the like, and *-ti̯- became Proto-Tocharian *-ts- while *-ai̯ti̯ā- was dissimilated to *-ai̯tā- in Iranian.

In conclusion, after having shown the difficulties with the traditional etymologies of TB *witsako* 'root', which make it impossible to derive it from an Iranian language, I have presented a new possible etymology of this word. I suggest it is a BMAC word, connected with the Iranian forms Iron *widag*, Digor *wedagæ* 'root', Pashto *wuləy* 'root, root-fibre', Sogdian *wytʾk* 'string' and Yaghnobi *wita*, and thus, ultimately, a cognate of these Iranian forms that previous scholars wanted to derive it from.

I am aware of the fact that my BMAC derivation is tentative. Although it seems a good option to me, I should stress that if this derivation is not accepted, my rejection of the traditional derivation of the Tocharian word di-

rectly from Iranian still holds true. The supposed correspondences contradict the secure patterns of the adaptation of Old Steppe Iranian words in Tocharian, and no other plausible Iranian candidate for the borrowing is attested.

3.2.8 TB śerwe, TA śaru 'hunter'

Tocharian B śerwe 'hunter' (adj. śerwäññe) and Tocharian A śaru 'idem' point to a Proto-Tocharian reconstructed form *śerwe or *śerəwe. From this noun, a verb *śerw- was built, with a subjunctive stem *śerwəy°/ₑ (cf. Peyrot 2013: 826), as found in the Tocharian B infinitive śerītsi 'to hunt' (cf. Pinault 2008: 588f.).

Pinault (2006: 179-181) connects these words to a BMAC word reconstructed as *ćaru̯a by Lubotsky (2001: 310). Lubotsky does not give a particular meaning to this word, but assumes it is the name of the deity from which Vedic Śarvá- and Avestan Sauruua- 'name of a daēva' come. Pinault suggests an original meaning 'hunting, living in the forest', and notes that Śarva- is depicted as an archer in the Vedas.

Further cognates adduced by Pinault are Ossetic Digoron sorun/surd, Iron suryn/syrd 'to track, hunt'[163] and Khotanese hasura- 'quarry, hunted beast'. This is accepted by Cheung (EDIV: 338) who posits a Proto-Iranian verb *saru̯- 'to hunt' and adds the Yazgholami verb sard/sar- 'to track, hunt' as a new cognate.

Adams (DTB: 695) prefers to connect Tocharian B śerwe, Tocharian A śaru to Proto-Indo-European *ĝʰu̯er- 'wild animal' with a secondary suffix *-u̯o-. However, as Pinault (2006: 179) points out, the function of the suffix is problematic, and "the development of the initial cluster would be unexpected".

Another explanation, based on *k̂ēru̯-o- and a connection to the Proto-Indo-European word for 'horn', is mentioned (with relevant literature) by Pinault (2006: 179-180). This proposal is also found again in Jasanoff (2017: 79). The idea is that the word for horn served to make an animal name 'stag'

---

[163] Miller 1962 [1881-1887]: 106 and 1903: 62 translates it as 'to chase (гнаться, преследовать)' and 'nachjagen, verfolgen'. This might derive, naturally, from a meaning 'to hunt', but I believe 'to chase; to follow (track)' seems more fitted for a primary meaning from which 'to hunt' would derive. I am not sure what the consequences of this could be for the general etymology discussed here.

and from this animal name the name of the 'hunter' was produced: or, as Jasanoff (op. cit.) writes "lit[erally] 'stag man'".

This idea is rightly rejected by Pinault (2006: 179-180) for obvious semantic reasons (Indo-Europeans did not hunt stags more than any other animals; as Pinault writes: "hunting was never limited to stags or horned animals, especially in Central Asia").

Pinault's explanation makes much more sense than any Proto-Indo-European connection made until now. He suggests a semantic path in Indo-Iranian from *'hunting, hunter' to *'living in the wild' > 'god of the wild, killer of living beings'. It is not necessary for this semantic development to have occurred within Iranian, as the source of Tocharian loanwords from BMAC is probably not the exact same as for Indo-Iranian, as I argue further below. The theonymic aspect of this word could have already arisen in the BMAC source language of Indo-Iranian, or, possibly, it both meant 'hunter' and designated the god of hunters, or a hunter god in the source language (both of Indo-Iranian and Tocharian, or only of Indo-Iranian).

In conclusion, I follow Pinault (2006: 179-181) in viewing the origin of Tocharian B *śerwe* 'hunter' and Tocharian A *śaru* 'id.' as BMAC, ultimately related to the theonym Vedic *Śarvá-* and the *Daimonym* Avestan *Sauruua-*. The Proto-Indo-European connections that were proposed for the Tocharian words are rather weak, and the connection to the Indo-Iranian names are justified (at least for the Indic side) by the identification of *Śarvá-* to a hunting character.

3.2.9 TB *ṣecake*, TA *śiśäk* 'lion'

Tocharian B *ṣecake* and its Tocharian A equivalent *śiśäk*, both meaning 'lion', do not have a clear etymology. It is for instance impossible to reconstruct a single prototype for both forms, as Proto-Tocharian *e does not yield Tocharian A *i*, but *a*, and PT *c does not yield Tocharian A *ś* but *c* (see for example TB *epiyac*, TA *opyāc* 'memory'). On the basis of the Tocharian B form, one needs to reconstruct Proto-Tocharian *ṣecəke, but *ṣecəke would not yield Tocharian A *śiśäk*.

To explain the *ś-* in the Tocharian A form, influence of TA *śiśri*, for which Sieg (1944: 16) proposed the meaning 'mane' has been evoked (e.g. DTB: 723). This is of course speculative, but not unlikely, if the meaning is correct, as few other animals have manes, and they are thus often seen as

stereotypical of lions. Note, however, that *śiśri* is a hapax and that its meaning could be different. Indeed, in the context where it occurs, the lion is lying down and his *śiśri* "up" (*koc*), which is unexpected. Furthermore, as Pinault notes (2022: 525) the word for the lion's 'mane' is attested in Tocharian A, it is *kesār-śiśäk* (*kesār* is a borrowing from Sanskrit *kesara-* 'mane' and the compound is a calque of Sanskrit *kesari-siṃha*, cf. Pinault, op. cit.). Pinault (2022: 525) thus suggests a different meaning for *śiśri*, namely, 'whiskers'. He also analyzes (op. cit.) the form *śiśri* as a dual and proposes to derive this word to the Proto-Indo-European form **dei̯k̂-ero-* 'indicateur' < **dei̯k̂-* 'to show, to indicate'. This new meaning does not preclude the solution suggested above: if cats and other feline animals were not so common among Tocharians, whiskers could have been viewed as typical of lions, and thus **ṣecəke* 'lion' could have become TA *śiśäk* by contamination of TA *śiśri* 'whiskers'.

In lack of an evident Indo-European etymology, TB *ṣecake* and TA *śiśäk* have often been compared to Middle Chinese *srij-tsiX*, modern Chinese *shīzi* (cf. DTB: 722), or to Ved. *siṃhá-* 'lion' and Classical Armenian *inj*, Middle Armenian *unj* 'panther, leopard', (cf. e.g. Witzel 1999: 56; see the literature in DTB: 722 and Blažek & Schwarz 2017: 69). If these forms are related, it must be in a very complex and intricate way.

It has been variously argued that the Tocharian words were borrowed from Chinese, or the other way around. Since lions were originally present in the West and were only later introduced to the Chinese (cf. Blažek & Schwarz 2017: 69), and there were no lions in China proper until the first centuries of our era, at least, (cf. Behr 2004: 6), I follow Pulleyblank (1962) in observing that the borrowing direction from Chinese to Tocharian for this word does not make much sense and that the other direction is preferable. One problem for the comparison is the fact that Tocharian *-ke* seems, superficially, to correspond to Chinese *-X*, which is not a consonant, but designates a tone.

As other examples of interesting similarities between Chinese words and Tocharian ones, one can cite TB *ṣitsok* 'millet alcohol', derived from Chinese *shǔ jiǔ* 黍酒 < MC *syoX tsjuwX*, and TB *śakuse* 'brandy', derived from *sù jiǔ* 粟酒 < MC *sjowk tsjuwX* (cf. Peyrot 2019: 144). There are thus other examples of Chinese *-X* corresponding to Tocharian *-k*. Since, for geographical reasons, a borrowing from Chinese into Tocharian is unlikely, it is pos-

sible that the Chinese word for 'lion' was borrowed from Tocharian, as proposed earlier in the literature.

It has also been proposed that TB *ṣecake*, TA *śiśäk* were borrowed from an Old Iranian **šargu-* ~ **šargawa-* 'lion' (e.g. Kümmel 2020: 259). This is impossible, as Old Steppe Iranian **šargu-* would yield Proto-Tocharian **ṣerk*, and **šargawa-* would yield Proto-Tocharian **ṣerkwe*. Furthermore, the reconstruction of this root for Proto-Iranian seems more than doubtful to me (regarding its internal derivation, formal aspect and geographical distribution).

The only segment for which a Proto-Tocharian reconstruction is fitting for this word is the ending in **-əke* (Behr 2005: 10; Pinault 2015: 188). This ending was already observed for *iṣcäke* (s.v.) and I proposed that it was of BMAC origin in *iṣcäke*, in accordance with Lubotsky (2001: 304). This might suggest that these words were borrowed from a BMAC language. One can refer to Behr (2005: 12), who saw the Tocharian words as borrowed from another language (he also wrote that Chinese borrowed its word for 'lion' from that same third language).

Given the possibility that Tocharian A *śiśäk* was influenced by *śiśri* 'mane' or 'whiskers', we can reconstruct a Proto-Tocharian form **ṣecəke* 'lion', a trisyllabic word with middle syllable stress, like most other words in this chapter.

In conclusion, TB *ṣecake* and TA *śiśäk* 'lion' are problematic words, with internal and external issues. Unless new facts enlighten the situation and help us solve this equation, I do not think a clear etymology can be proposed. However, they are clear borrowings from another language, and their **-əke* suffix seems to suggest the source word was of a BMAC origin.

3.3 Other possible BMAC loanwords in Tocharian

In this section I will discuss three more words of possible BMAC origin found in Tocharian, in less detail than the previous word studies. Two of these words have already been suggested as BMAC loanwords in the literature and one, *lepäś**, has not.

3.3.1 TB *peñiyo*, TA *pañi* 'splendor'

Pinault (2006: 181-183) reconstructs a BMAC word **paṇi* 'wealth', of which PT *peñiyo* would be a derivative. This etymology works formally and

semantically. It should nonetheless be noted that another etymology has been proposed for these words: Beekes (2010: 1546) has connected them to PIE *b^heh_2- 'to shine, appear'. This etymology is formally difficult (cf. Del Tomba 2020: 168) so that Pinault's proposal seems more likely.

3.3.2 TB *mewiyo* 'tiger'

Dragoni (2022: 170f.) has suggested that Tocharian B *mewiyo* 'tiger' was borrowed from BMAC *$mau̯ii̯a$- which, according to him, was also borrowed in Iranian: Sogdian *myw* and Khotanese *mūya*-*. The final -*o* of the Tocharian B form could fit this hypothesis. No better Indo-European or Iranian origin has been put forward (for a critical review of the literature, see Dragoni 2022: 170f.), and this hypothesis would strengthen my suggestion that TA *lepäś** 'jackal' is a BMAC loanword (see below) because they are from the same semantic field.

3.3.3 TA *lepäś** 'jackal'

This word is only attested in the gen.pl. *lepśäśśi*. Malzahn (2014: 92-93) writes that "[…] *lepś* does not have the ring of an inherited word, and the jackal is *not* native to Central Asia". Although it is true that this word does not have the ring of an Indo-European inherited word, it is not scientifically correct that the jackal is not native to Central Asia (the golden jackal for instance is present in Central Asia). She further suggests that TA *lepäś** 'jackal' was a loan from an unattested Tocharian B word (that, for the sake of the argument, would have to be set up as **lepśe vel sim*.), itself from Sanskrit *lopaśa*- 'a kind of fox or jackal' *vel sim*. but that it was borrowed into (pre-)PT early enough to get its vowel -*o*- treated in the same way as PIE *-*o*-. This is hardly possible, as for instance no Sanskrit vowel **o* can be shown to have been treated in Tocharian as PIE *-*o*-, no matter how early the word was borrowed.

 Palmér & al. (2021) suggest that the Indo-Iranian word for 'fox' is a cognate of, for instance, Greek ἀλώπηξ 'fox', and make it go back to **h_2lop-$e\hat{k}$-. Nonetheless, there are many issues concerning the vocalism of this etymology, notably in the root and in the suffix. The fact that the word for 'fox' shows the same vocalic shortenings in Iranian languages as do a number of

other borrowed animal names (and no inherited name) rather suggests that this is a borrowed word.[164]

It would be more plausible to consider TA *lepäś** as a direct BMAC loanword. If *lepäś** goes back to Proto-Tocharian, it points to a pre-form **laipəśe*, which would correspond to the general structure of the "substratum" words as described above. One can further speculate that **laipəśe* was a distant cognate of Sanskrit *lopaśa-* 'fox, jackal' < PIIr. **r(/l)au̯paća-*, itself a BMAC loanword, with a different initial diphthong. If this is correct, the element **äśe* could be interpreted as a reflex of the BMAC suffix **-aća-*, which is often found with words denoting "dangerous animals" (on which see Bernard 2020: 38f.).[165]

3.4 Discussion

3.4.1 Presentation of the loanwords and their phonetics

In this chapter I tried to analyze and discuss thirteen Tocharian words that have no clear Indo-European etymology.

These words are: TB *iṣcäke* 'brick'; TB *eñcuwo*, TA *añcu** 'iron'; TB *etre*, TA *atär* 'hero'; TB *kercapo* 'donkey'; TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric'; TB *kroṅkśe*, TA *kroṅśe* 'bee'; TB *peñiyo*, TA *pañi* 'splendor';

---

[164] Some of the issues found in the paper include the fact that if, as they claim, the suffix **-āća-* spread from the word for 'fox' to the other animal words, it needs to have spread after the thematicization to other words, when every individual language had generalized a short or long variant of the suffix: this poses a problem, given that not every language has a short or long variant for every *āća*-word (cf. Bernard 2020: 37f.). They write that "[a]s the suffix **-ăća-* was unique, **(H)rāpā̆́ća-* may have been analysed as containing the suffix **-ća-* '-like' (cf. Skt. *yuvaśa-* 'youthful') by the speakers of Proto-Indo-Iranian." (2021: 241). The suffix *-aśa-* 'like' is absent from Iranian, and seems to be an Indian, if not Sanskrit, innovation. In any case, they do not show its antiquity, which is necessary for such a claim.

[165] In Bernard (2020: 38) I proposed that **rau̯paća-* (as opposed to the variant **rau̯pāća-*) derived from the substratal word **rau̯pi-* suffixed with **-aća-*. Perhaps **lai̯pi-ća-* or the like would have been borrowed as PT **lai̯pəśe*.

TB *mewiyo* 'tiger'; TA *lepäś** 'jackal'; TB *witsako* 'root'; TB *śerwe*, TA *śaru* 'hunter'; TB *ṣecake*, TA *śiśäk* 'lion'.

Out of these words, it was determined that TB *eñcuwo*, TA *añcu** 'iron' is of Iranian origin (cf. also Peyrot, Dragoni & Bernard 2023). I considered that all the others were borrowed, or possibly borrowed, at the Proto-Tocharian stage. Indeed, when cognates are found in both languages, a Proto-Tocharian prototype can be reconstructed for all these words, except for TB *ṣecake*, TA *śiśäk* 'lion', where the discrepancy can be explained by the influence of TA *śiśri* 'mane' on the Tocharian A form.

On the basis of the Old Steppe Iranian - Proto-Tocharian correspondences established in the previous chapter, I was able to reject an Iranian etymology for all these words when an Iranian etymology had been proposed. Besides, no convincing Iranian etymon has yet been found to explain TB *witsako* 'root' and TB *kenek*, TA *kanak* 'cotton cloth' as Iranian loanwords, nor does an Iranian etymon for TB *etre*, TA *atär* 'hero' exist.

For a number of those words it was proposed that they were from the same language that is otherwise known as the BMAC language/substratum or Indo-Iranian substratum (Lubotsky 2001). For each and one of those words for which it was proposed, I have examined these proposals, and found them, if not convincing, entirely plausible (with the exception of B *eñcuwo*, TA *añcu** for which an Iranian etymology was ultimately found).

It is useful to compare, like Lubotsky (2001) did, the main features of these words in order to establish more general observations about BMAC loanwords in Tocharian. Out of the twelve words I assume to be of BMAC origin, nine follow the same syllabic pattern, namely, they are trisyllabic words with fixed second-syllable stress. These words are: TB *iṣcäke* 'brick'; TB *kercapo* 'donkey'; TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric' (← *kanáka 'hemp cloth'); TB *peñiyo*, TA *pañi* 'splendor'; TB *mewiyo* 'tiger'; TB *witsako* 'root'; TB *śerwe* < *śeruwe, TA *śaru* 'hunter'; TB *ṣecake*, TA *śiśäk* 'lion'. Only three words are disyllabic or monosyllabic: TB *etre*, TA *atär* 'hero'; TB *kroṅkśe*, TA *kroṅśe* 'bee'; TA *lepäś** 'jackal' (which could be from *laipəśe).

This observation is to be compared to the fact that Old Steppe Iranian loanwords in Tocharian had fixed first-syllable stress (section 2.6.2.g of the previous chapter). The stress in these BMAC words needs to be old, or orig-inal, because, according to my findings, Proto-Tocharian preserved the orig-

inal stress when borrowing words. One can also deduce from TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric' which go back to BMAC **kanáka*, that the borrowings precede the syncope law of Proto-Tocharian (see section 2.6.2.g), and they have thus in any case not been borrowed much later than the Old Steppe Iranian loanwords.

The tendency of the Indo-Iranian substratum language to have words with three syllables was observed by Lubotsky (2001: 303). He noted that many of the substratal words in his list were "trisyllabic nouns with long middle syllable". This differs from our list, where the middle syllable is – in a small majority of cases – an accented schwa. If we look at the Vedic accent in those trisyllabic words listed by Lubotsky (p. 305), we see that the accent is sometimes on the first, middle or last syllable, with no easily recognizable pattern. One can imagine that an original middle stress accent "lengthened" the vowel, and that the stress would later be placed on another syllable. Perhaps Tocharian speakers confused length with stress, or perhaps Proto-Indo-Iranians confused stress with length, leading to the different adaptation of the middle syllable in Proto-Indo-Iranian and Proto-Tocharian.

However, it is also possible that the source languages of Indo-Iranian and Proto-Tocharian were different, albeit related, with different stress patterns. In words that are, in my opinion, clear cognates, such as PT **kercə́po* : Ved. *gardabhá-* (not listed in Lubotsky 2001, because there are no Iranian cognates), the stress is clearly different in Vedic and Tocharian, and there is no lengthening of the middle syllable in the Indic word, although the explanation can be that Vedic speakers interpreted the ending as related to the animal name suffix *-bhá-*.

Another striking phenomenon is the fact that the BMAC vowel borrowed as **a* in Indo-Iranian was borrowed as Proto-Tocharian **e* (TB *e*, TA *a*), typically, Tocharian B *śerwe*, Tocharian A *śaru* 'hunter' : Ved. *Śarvá-* 'name of a god', and PT **kercəpo*, TB *kercapo* : Ved. *gardabhá-*. This coincides with the adaptation of Old Steppe Iranian **a* as Proto-Tocharian **e* (see the previous chapter). In PT **kercəpo*, TB *kercapo*, compared to Ved. *gardabhá-*, it also appears that another vowel borrowed as **a* in Indic was borrowed as schwa in Tocharian.

I have argued that there was an unclear phoneme which was rendered by PT **ṅkś* and Indic *-ñc-*, in the word TB *kroṅkśe* 'bee' : Sanskrit *krauñca-* 'crane' (going by, I argue, the notion of 'loud animal'). Further, the initial phoneme (perhaps a pharyngeal?) seen in PIIr. ** ḥišt-* 'brick' was not ren-

dered by Proto-Tocharian. Finally, the palatal phoneme provisionally set up as *ć was rendered by Proto-Tocharian as *ś, as in Proto-Tocharian *śerəwe 'hunter' ← *ćarwa, and by (Proto-)Indo-Iranian as *ć, as in Proto-Indo-Iranian (or common Indo-Iranian) *ćaru̯a- 'hunting god'.

Out of all the substratal suffixes noted by Lubotsky (2001: 304), I can notably recognize the suffixe *-(ə)ka, found in TB iṣcäke 'brick', ṣecake 'lion' and in PT *kenek 'hemp cloth' (or 'cotton cloth') if from an earlier form *keneke as I suggested. The substratal suffix *-pa- (also discovered by Lubotsky), is found as -po- in kercapo 'donkey'. As I wrote above, we can suppose that this BMAC word was remotivated in Vedic, with the replacement of the BMAC suffix *-pa- with the animal name suffix *-bhá- which is found for instance in vṛṣabhá- 'bull' (this could even explain the accent of the Vedic word as different from that in Tocharian).

3.4.2 The semantics of the BMAC loanwords in Tocharian

In order to get a clearer picture of the type of contact that took place between the BMAC language and Proto-Tocharian, it is necessary to analyze the semantic fields of the loanwords. We easily see that they mostly concern realia:

1. Animal names: TB kercapo 'donkey'; TB kroṅkśe, TA kroṅśe 'bee'; TB mewiyo 'tiger'; TA lepäś* 'jackal'; TB ṣecake, TA śiśäk 'lion'.
2. Botanics: TB witsako 'root'; TB kenek, TA kanak 'cotton cloth' (possibly originally a plant name).
3. Legends, myths: TB peñiyo, TA pañi 'splendor'; TB etre, TA atär 'hero'.
4. Construction material: TB iṣcäke 'clay or brick'.
5. Hunting vocabulary: Tocharian B śerwe, Tocharian A śaru 'hunter'.

By far the largest category of BMAC loanwords studied in this chapter is constituted by animal names. This differs somewhat from the main semantic category found in the Indo-Iranian substratum loanwords (Lubotsky 2001: 307): names of body parts. Nevertheless, the second largest category, as per Lubotsky (2001), seems to be "wild animals", or perhaps, rather, animals found in the wild (since camels, donkeys and tortoises, and even pigeons do not need to be wild animals). Equally important for Indo-Iranian is the category "religion, cult", of which we find only a small trace in Tocharian,

namely what I named "legends, myths", the notion of glory and the notion of hero. Then comes "clothing" to which corresponds TB *kenek*, TA *kanak* 'cotton cloth' (if it was not borrowed as a plant name, different from what I hesitantly suggested above). "Building technology" has the Tocharian cognate TB *iṣcäke* 'brick'. Finally, for the categories "artifacts", "water economy and irrigation", "cattle breeding" and "agriculture" I could find no corresponding Tocharian "substratal word".

There are of course much fewer Tocharian words in my list than Indo-Iranian words in Lubotsky's list. As this is a selective study, I did not include every possible example, although I did not find any example of the latter categories, even in other sources.[166] The evident explanation is that Tocharian A and B have a much smaller quantity of attested words than the very well attested and prolific Indic and Iranian branches of Indo-European. This discrepancy in data will naturally lead to a bias in the comparison of the Tocharian and the Indo-Iranian "substratum" words. Nonetheless, the type of BMAC vocabulary in Tocharian in my view seems to indicate another type of contact, namely, contact between a more "primitive" BMAC culture than the one in contact with Indo-Iranians. It is also clear from the animal terms that these BMAC people were more familiar with the fauna of the region than the Tocharians, which might imply that the latter were only arriving in the region when they borrowed the words from the BMAC language.

An interesting point is that, according to my findings, TB *iṣcäke* 'clay' was borrowed from the BMAC language, while TB *iścem* 'clay' was borrowed from Old Steppe Iranian. Iranians, in their turn, had borrowed this word from a BMAC language (cf. Kümmel 2020: 257). This might indicate that Tocharians had contact with a different BMAC people from the one Indo-Iranians were in contact with, and also, perhaps, that the type of clay designated by both words was different, so that Tocharians could have used

---

[166] One could, however, suggest TB *ñemek* 'harvest' which has no clear etymology and resembles TB *kenek* 'cotton cloth, cotton fabric'. However, not only it has no Indo-Iranian equivalent, but it has received a very convincing etymology by Pinault (2020b: 214-215) who derives it from PIE **nem-* 'to take', with the *-ek suffix of Tocharian (for which see Pinault 2020b). Pinault (op. cit.) further connects it to TA *ñomes* 'martingale' (the harnessing part of the plow) which for him originally meant 'pertaining to taking (the animal under control)'. I think one can also propose that it originally meant '(harness) pertaining to harvest'.

both of them in a complementary manner, although there are not enough attestations of TB *iṣcäke* to be sure of this.

An intriguing fact, if I am correct, is that these BMAC speakers called bees 'the noisy ones'. This could suggest that they did not collect honey, as they did not name them after honey, which is their major attribute to honey-collecting peoples (although they perhaps only called swarms of bee so, and had a different word for individual bees). Indo-Europeans did have a word for honey, **melit-*, so they were familiar with honey, but they also did not conceive of bees as 'honey-makers', which implies they were possibly not familiar with the process of honey-collecting (cf. van Sluis 2022).

These BMAC speakers were familiar with hunting, which could imply that they were a hunter gatherer society, but could also simply imply that they hunted besides their agricultural activities. In Indo-Iranian, a cognate of TB *śerwe* is also found, yet we can be fairly certain that the BMAC speakers in contact with Indo-Iranians had an agricultural society. As far as I know, no agricultural society has completely abandoned hunting activities, even though they are sometimes reduced to a hobby, as they are currently in Western Europe. As far as I can see, the Tocharians did not borrow any names of sophisticated tools or weapons from them, as they did from Old Steppe Iranian. This is remarkable, as the BMAC civilization in the West of Central Asia exported weapons, notably decorated weapons (cf. Parpola 2022: 26). Two other interesting words, represented by TB *peñiyo*, TA *pañi* 'splendor'; TB *etre*, TA *atär* 'hero', seem to indicate that Tocharians could also have heard stories from that BMAC people, that this people also influenced their world view, to a more limited extent, of course, than Iranians did. This could perhaps be compared with "mythical" BMAC loanwords in Indo-Iranian, such as Vedic *gandharvá-* : Avestan *gaṇdərəβa-* 'a mythical being' (cf. Lubotsky 2001: 303).

Thus, although BMAC loanwords are also found in Indo-Iranian, the picture of the culture of the source language given by the borrowings in Tocharian is very different from what we can deduce from the ones found in Indo-Iranian. Most importantly, Tocharian cannot be shown to have borrowed words relative to "water economy and irrigation", "cattle breeding", or "agriculture", categories which we find listed in Lubotsky (2001). The same BMAC word which for Iranians meant 'clay brick' designated 'clay' in Tocharian. This perhaps shows that the BMAC speakers in contact with Indo-Iranians had a more evolved culture, with bricks made of clay, whereas

BMAC speakers in contact with Tocharians had a technologically less advanced culture.

Interestingly, since there are no indications of words relative to weaponry or violence and war (unlike from Old Iranian), it is difficult to imagine that Tocharian speakers were in any type of military conflict with speakers of the BMAC language. If Tocharian borrowed its word for 'root' from them, this may show that they had a certain knowledge of botany, and perhaps thus of medicine, as roots often have medical uses.

The BMAC loanwords in Tocharian seem of a more primitive content than the Old Steppe Iranian ones, indicating possibly a less developed culture. One could deduce that these borrowings from BMAC in the Pre-Proto-Tocharian period preceded those from the more militarized and more economically developed Old Iranian civilization. Indeed, it is difficult to accept the scenario that Indo-Iranian should have borrowed their words from this civilization, usually words for simple notions, for realia, then, hundreds of years later, an Iranian tribe came to the Tocharians, probably with military strength, and only then came that unknown people, or was that unknown people encountered, which had kept the same archaic phonology as during the Indo-Iranian borrowings, and that Tocharians borrowed some elements of their lexicon from them.

Furthermore, I do not see why they would not have borrowed the same words from the Iranians, who also possessed a similar if not more developed craftsmanship, and mastering of metals and construction materials, etc. In particular, it is difficult to understand why they would not have borrowed the animal names that they borrowed from the BMAC language from Old Steppe Iranian instead. I nevertheless do not think that the borrowings of BMAC words into Tocharian occurred much earlier than those into Indo-Iranian: they could have occurred more or less at the same time, but from a sister language.

It seems unlikely that all three, the Indo-Iranians, the Tocharians, and the BMAC people were at the same place at the same time. If Tocharians and Indo-Iranians were at the same place at the same time when the latter borrowed their BMAC loanwords, why did Tocharians not borrow any Indo-Iranian words at the same time? Of such words there is no trace. Many peoples borrowed from Proto-Indo-Iranian during the Proto-Indo-Iranian migration (for a survey, see Kümmel 2020). It seems more likely, in my opinion, that the Indo-Iranians were in contact with a specific group of BMAC speak-

ers, much more advanced than the one Proto-Tocharian speakers were in contact with. BMAC contact should thus be parallel, but not necessarily simultaneous, for Indo-Iranians and Tocharians. This would perhaps explain why a word like 'hunter' was deified in a more complex civilization, while it kept its base meaning 'hunter' in the variant that was borrowed into Tocharian.

3.4.3 Western and Eastern BMAC varieties

One can push the hypothesis discussed above further, and propose that there existed an Eastern and a Western BMAC language. Tocharians would thus have been in contact with the Eastern BMAC language and Indo-Iranians with the Western one. This needs of course much more research, but one can already mention a few of the differences that appear between both varieties. This distinction is supported by archaeology (Zhang & al. 2021), which recently found traces of BMAC people in the region where Tocharian speakers where also found. It is possible, if not probable, that a BMAC language spoken so far away from the West was different from the language spoken in Bactriana-Margiana properly. Below, I will try to systematically present the variation we see between Indo-Iranian BMAC loanwords and Tocharian BMAC loanwords, in order to present what type of variation one can find.

Western BMAC *d or *t vs. Eastern BMAC *ǰ or *č

- Vedic *gardabhá-* 'donkey' vs. Tocharian B *kercapo* 'donkey'
- PIIr. **ištika-* 'brick' vs. Tocharian B *iṣcäke* 'clay', Burushaski *diṣçík* 'brick' (but also Burushaski dialectal variant *diṣṭík* etc.).

Western BMAC *d vs. Eastern BMAC *dz (?)

- PIr. **u̯ītāka- ~ *u̯ai̯tāka-* 'root': Iron *widag*, Digor *wedagæ* 'root', Pashto *wuláy* 'root, root-fibre', Sogdian *wytʾk* 'string' and Yaghnobi *wita* 'cord' vs. TB *witsako* 'root'

Western BMAC *-ñc- vs. Eastern BMAC *-nkś-

- Sanskrit *krauñca-* 'crane' vs. TB *kroṅkśe* 'bee' (for the semantics, s.v. TB *kroṅkśe*).

Western BMAC *-au̯- vs. Eastern BMAC *-ai̯-

- PIIr. *r(/l)au̯pi-* 'marten', *r(/l)au̯paća-* 'fox, jackal' (< *'dangerous marten') vs. TA *lepäś** 'jackal' < PT *laipəse.

Western BMAC *-ba- ~ *-fa- suffix (< *-pa-?) vs. Eastern BMAC *-ka- suffix

- Although the suffix *-ka- is known in Western BMAC, it is not found in Common Iranian *kanaba- ~ *kanafa- 'hemp' vs. PT *kenek 'cotton cloth'.

There were also differences in meaning:

- Indo-Iranian *aθrau̯a- ~ *aθaru̯a- 'priest' vs. PT *etre 'hero'.
- PIIr. *išti- 'brick' vs. Tocharian B *iṣcäke* 'clay'.
- PIIr. *ćaru̯a- 'hunting god' vs. PT *śerəwe 'hunter'.
- Sanskrit *krauñca-* 'crane' vs. TB *kroṅkśe* 'bee' (for the semantics, s.v. TB *kroṅkśe*).

One can add a further example which does not directly concern Tocharian:

Western BMAC *paraću- 'axe' vs. Eastern BMAC *paratu- 'id.'

- Indo-Aryan *paraću- 'hatchet' (Ved. *paraśú-*, etc.) vs. Eastern Iranian *paratu- 'axe'.

The examples are too few to make developed conclusions concerning these differences, but they seem to indicate that, although there probably was a dialect continuum between BMAC varieties, they were sensibly different on a number of points.

Usually, Proto-Indo-Iranian borrowed from one single variety and Proto-Tocharian from another. However, the word for 'axe' was seemingly borrowed later, and from two different varieties: *paraću- in Indo-Aryan and

paratu- in Eastern Iranian (cf. section 2.6.3.g). While one cannot be sure that BMAC is the donor language of this word, the BMAC civilization was known for its axes (cf. Parpola 2022: 26) so it is a good candidate. On the other hand, Tocharian seems to have borrowed from a variety that differed in some way from the one the Indo-Iranians borrowed from. This can be seen from the phonetic and semantic differences mentioned above.

I am aware of the fact that this is an entirely new discussion topic, which is subject to caution, and needs much more research. More research would also include archaeological data in order to support or disprove, for example, the idea that Eastern BMAC, from which Proto-Tocharian borrowed its words, was less advanced technologically than Western BMAC, from which Indo-Iranian borrowed its words.

All in all, this new hypothesis has the advantage to explain and justify the fact that both Indo-Iranians and Proto-Tocharian speakers borrowed words from what seems to be the same language, but not at the same level of technological advancement, and certainly not at the same place.

Animal names were adapted to the local fauna: while 'dangerous (or obnoxious) martens' became foxes in the Iranian world (PIr. *raṷpaća-*) they designated jackals in the Indian and Tocharian worlds (Indo-Aryan *laṷpaća-*, PT *laipəśe*). Perhaps an interesting way to study these dialect differences would consist in integrating zoological studies to the lexical and etymological studies that could be done.

I have added this speculative perspective in order to try to make sense of some of the frequent differences in BMAC borrowings in Tocharian and Sanskrit or Iranian. Naturally, as I explained throughout the chapter, they could also be due to different adaptations of the same phonemes. The semantic differences, however, could be indeed due to dialectal or linguistic difference, which could in their turn, suggest the existence of an Eastern vs. a Western BMAC variety, dialect or language.