

6 Tocharian

Michaël Peyrot

6.1 Introduction

The Tocharian languages A and B are attested in manuscripts from the northern Tarim Basin, present-day Northwest China. Tocharian B is attested from about the fifth to the tenth centuries of the Common Era. Originally from Kuča, it spread east to Yānqí and Turfan, probably in the late sixth and in the seventh century. In Tocharian B itself, the language is referred to as the language of *kuśi* ‘Kuča’. Tocharian A is attested a little later, from about the seventh to the tenth centuries. It is originally from Yānqí, spread with Tocharian B east to Turfan, but not west to Kuča, and is referred to as the language of *ārśi* ‘Yānqí’. Both languages are written in the Indian Brāhmī script, and the vast majority of the manuscripts are of Buddhist content.

Traces of a third Tocharian language have been claimed to be preserved in the Middle Indic Gāndhārī dialect of Niya in the southern Tarim Basin (Burrow 1935). This hypothesis has not received wide support and must still be considered very uncertain (see further below in Section 6.3).¹

6.2 Evidence for the Tocharian Branch

The existence of the Tocharian branch of Indo-European is beyond any doubt. The two languages A and B are closely related and share numerous significant innovations, so it is unnecessary to give a full list here. Some of the more important, branch-defining developments are:

- loss of the threefold Proto-Indo-European distinction between the conventionally termed voiceless, voiced and voiced aspirated stops, i.e. **k̰*, **g̰*, **g̰h* merged into **k* (on **d*, see below);

This research was financed by the European Research Council (ERC-2017-STG 758855).

¹ I will not discuss in detail a posthumously published proposal by Schmidt (2018: 161–271) to read previously undeciphered manuscript fragments in Formal Kharoṣṭhī as a Tocharian variety from Lóulán. His tentative decipherment is not convincing. Instead, these fragments are probably written in an Iranian language related to Khotanese and Tumšūqese (Dragoni, Schoubben & Peyrot 2020).

- several mergers and shifts in the vowel system, including loss of vowel length, merger of **i*, **e*, **u* into **ə* (the first two regressively palatalising), shifts of **o* to **e* and of **ā* < **eh₂* to **o*, monophthongisation of **ei* to **i* and of **eu* to **u*, etc.;
- rise of distinctive and morphological palatalisation, principally through the transformation of the contrast between **o* : **ē* into **e* : **'e* and **∅* : **e* into **ə* : **'ə*;
- loss of word-final **-s*, **-m*, **-n*, **-t* (**-d*), which has led to heavy restructuring of both the nominal and the verbal inflection;
- rise of agglutinative case inflection in the noun, next to agglutinative number inflection in some noun classes;
- almost complete loss of prefixing morphology;
- rise of an intricate system of verbal derivation to form intransitives and transitives or causatives;
- numerous significant innovations in the lexicon.

Even considering the late attestation of the Tocharian branch, the extent of structural change is surprisingly large, and it can be argued that this is partly due to a substrate effect. The loss of the distinction between the so-called voiceless, voiced and voiced aspirated stops, the rise of agglutinative case inflection, and the functions of these case suffixes, which include the perlicative, denoting movement through, along or over something, point to Uralic influence. A pre-Proto-Tocharian phase of the vowel system can be compared more specifically with an early form of Samoyedic. Pronoun suffixes attached to the finite verb denoting the object may be compared with the objective inflection in Uralic (Peyrot 2019a with references; on the vowel system, see Warries in press).

It is more difficult to assess the Iranian impact on Tocharian. There has been considerable Iranian influence on the lexicon (Isebaert 1980; Tremblay 2005), but only the oldest layer of borrowings from Old Iranian may possibly be added as a branch-defining feature of Tocharian. The reason is that any feature defining the whole branch should have been acquired before the break-up of unitary Tocharian into Tocharian A and B. This is clearly the case with the structural shift attributed to Uralic above. However, many borrowings from Iranian are to be dated after the break-up and therefore do not define the Tocharian branch as such. Examples of this include borrowings from Bactrian, such as Toch.B *akālk* and Toch.A *ākāl* ‘wish’ from Bactrian *αγαλγο* /aγalg/: the *ā* *ā* vocalism of Tocharian A, instead of the *ā* *a* vocalism regular in inherited vocabulary, shows that the word has entered the language later, and the Toch.B and Toch.A forms cannot be reconstructed to a common proto-form. Bactrian influence is therefore to be dated after the split of Proto-Tocharian. The case of borrowings from Old Iranian is different. An example is

Toch.B *perne*, Toch.A *paräm* ‘glory’, which allows a Proto-Tocharian reconstruction **perne*, borrowed from Old Iranian **farnah-* (Av. *xʷarənah-*).

Nevertheless, for the Old Iranian layer, the details are not fully clear either. Tocharian B would have preserved a word like **perne* unchanged, and the amount of change in Tocharian A is limited: **e > a* in the first syllable; apocope of **e* in the final syllable; *ä*-epenthesis in the final cluster *-rn*. Since these changes in Tocharian A cannot be dated exactly, it cannot be excluded that **farnah-* was borrowed into Tocharian B and A independently, at an early stage, before the relevant sound changes in Tocharian A occurred but after the break-up of Proto-Tocharian. A reason to consider this more complicated chronology are the sound changes **rn > rr* and **ln > ll* in both languages. Good examples of the former are not found in Tocharian A, but the latter is certain. Since old geminates are generally simplified in Tocharian A, the rise of new geminates from **rn* and **ln* must be dated after the general simplification of geminates. The preservation of *rn* in ‘glory’ thus suggests an early but post-Proto-Tocharian borrowing according to the following relative chronology:

1. break-up of Proto-Tocharian;
2. degemination in pre-Tocharian A;
3. assimilation of **rn*, **ln* to *rr*, *ll* (the same change occurred independently in pre-Tocharian B);
4. borrowing of **farnah-* as **perne* (the same borrowing occurred independently in pre-Tocharian B);
5. **e > a*, apocope of final **e*, and *ä*-epenthesis to produce Tocharian A *paräm*.

Another indication of this chronology is offered by Toch.B *etswe* ‘mule’, borrowed from Old Iranian **atswa-* ‘horse’ (Av. *aspa-*). Although Toch. B *mättsa-*, Toch.A *nätswā-* ‘starve’ shows that Proto-Tocharian **tsw* has developed to *tsts* in pre-Tocharian B after the break-up of Proto-Tocharian, *etswe* has *tsw* unchanged, suggesting that the borrowing is post-Proto-Tocharian. Old Iranian borrowings can only be taken as a branch-defining feature if the preservation of the cluster *tsw* in Tocharian B, and of the cluster *rn* in both languages, receives an alternative explanation, notably a conditioning of the relevant assimilations, such as a difference in accent.

6.3 The Internal Structure of Tocharian

As Tocharian A cannot be derived from Tocharian B or vice versa, a common ancestor called Proto-Tocharian needs to be reconstructed. For instance, Toch. B *yente* ‘wind’ cannot have yielded Toch.A *want* ‘wind’, and the reverse is also impossible: a preform **wente* is to be posited, with innovations in both languages leading to the attested forms. There is no need to discuss the internal subgrouping of Tocharian, since only one tree is possible. The dating of Proto-Tocharian, the only node in this tree, will be discussed below. Even though

Burrow's hypothesis of a third Tocharian language is too uncertain to be taken into account for inferences on the prehistory of Tocharian, it presents an illustrative case for the methodology of internal subgrouping.

The Gāndhārī words in the documents from Niya for which Burrow (1935) suggests a Tocharian etymology are few, and among these only two are relevant here: *kitsa 'itsa*, a title, and *aṃklatsa*, a type of camel. *kitsa 'itsa* has a very Tocharian-looking structure and has been convincingly connected to Toch.B *ktsaitse* 'old', Toch.A *ktsets* 'perfect' by Burrow, who suggests 'elder' for the Gāndhārī title. Toch.B *ktsaitse* derives from PToch. **kəṛaitʰse* with degemination after a diphthong,² and Toch.A *ktsets* has undergone apocope of final *-e* and monophthongisation of **ai* to *e*; both languages have syncopated the **ə* in the first syllable. Niya *kitsa 'itsa* could derive from Tocharian B as well as pre-Tocharian A or a third branch and is therefore useless for subgrouping. It could reflect an older form of the type **kəṛaitʰse* with *i* for *ə* and the regular Gāndhārī final *-a* for the regular Tocharian final *-e*. The geminate could be simplified or left unwritten. Equally, it could go back to a form of the type Toch.B *ktsaitse*, with *i*-epenthesis in the first syllable. Since Tocharian A is attested from the seventh century onwards, much later than Niya Gāndhārī, which is from the third–fourth centuries, it could also derive from an early form of Tocharian A in which monophthongisation of **ai* to *e* had not yet taken place.

The key form for Burrow's understanding of the internal subgrouping is *aṃklatsa* (1935: 673). According to him, *aṃklatsa* denotes a relatively cheap camel, which may therefore have been untrained. He connects the word to Toch.B *aknātsa*, Toch.A *āknats* 'fool', which is formed with the negative prefix **en-* from the verb **kna-* 'to know': in both languages, the vowel of the prefix has been affected by *a*-umlaut, and its nasal has been lost before the cluster *kn-*. To explain the different cluster *ṃkl* in Niya Gāndhārī, he assumes that it goes back to an earlier form with **nkn* that was dissimilated to *nkl*, written *ṃkl*. Since the first *n* of the cluster is lost in both Tocharian A and B, he concludes that the Tocharian variety he assumes in the Gāndhārī of Niya is of a different branch, and this is the reason why it is often termed "Tocharian C".

Burrow's Tocharian etymology of Niya Gāndhārī *kitsa 'itsa* is attractive, but his explanation of *aṃklatsa* is not convincing in view of the semantic and formal problems. At any rate, this questionable etymology can never alone bear the weight of proving a third branch of Tocharian, the famous "Tocharian C".³

² See Peyrot 2008: 45 on Toch.B *-auñe* < **euñne* with the same degemination. Adams cites the word as *ktsaitstse* (2013: 263), but this form is not attested.

³ Even in the unlikely event that the etymology should be correct nevertheless, it does not necessarily prove the existence of a third branch of Tocharian. Rather than being a shared innovation of Tocharian A and B, the change **nkn* > *kn* may be a parallel development, since there are cases where the nasal is lost in Tocharian A but preserved in Tocharian B (Hilmarsson 1991: 193–8).

Rather, in the light of research by Niels Schoubben, who proposes new and convincing alternative explanations for some other items that Burrow explained as Tocharian (Schoubben 2021), scepticism about Burrow's hypothesis is definitely due.

No absolute date can be given for Proto-Tocharian, by definition the latest phase of unity before the break-up in pre-Tocharian A and pre-Tocharian B. The languages are closely related, but differences are considerable in the lexicon, and most scholars estimate Proto-Tocharian around 500 BCE: some take it to be a little bit earlier, between 1000 BCE and 500 BCE; others a little bit later, between 500 BCE and the beginning of the Common Era (see the useful overview of different estimates in Mallory 2015: 7–8).

It is commonly agreed that the advent of Buddhism was after the break-up, as such basic terms as *dharma* 'law' (Toch.B *pelaikne*, Toch.A *märkampal*) and *karman* 'act, fate' (Toch.B *yāmor*, Toch.A *lyalypu*) are different (Lane 1966). But since Buddhism arrived late in the region, perhaps in the first or second century CE, this gives only an unsurprising *ante quem* date.

Contacts with the Iranian languages Bactrian and Sogdian took place after the split, probably in the early first millennium CE. Contacts with Old Iranian are more interesting: since it can be debated whether they occurred before or after the break-up, they may have to be dated close to that break-up. In the scenario sketched above, they would have occurred soon after it.⁴ However, the Old Iranian loanwords are themselves difficult to date in absolute terms. The archaic appearance of words such as Toch.B *etswe* 'mule' ⇐ OIr. **atswa-* 'horse' (Av. *aspa-*) or Toch.B *waipecce* 'possessions' ⇐ OIr. **hwai-paθya-* (Av. *xʷaēpaiθīia-* 'own') suggests a date in the middle of the first millennium BCE or earlier, but a more precise dating is difficult. I have suggested that these loanwords may be associated with the presence of Andronovo related groups in Northern Xinjiang in the thirteenth–ninth centuries BCE (Peyrot 2018: 280), which would accordingly push the date of Proto-Tocharian towards the beginning of the first millennium BCE. The assumed contacts with Uralic, which may date to around 2500 BCE, in any case took place long before the split, in a pre-Proto-Tocharian phase.

Archaeological evidence on the Tocharians themselves is at present not clear enough (Mallory 2015: 29 and *passim*). It is uncertain whether the Cháwúhūgōu cultural group near Qarašähär (Debaine-Francfort 1989: 183–9), whose different phases together cover almost the entire first millennium BCE, can be identified with early speakers of Tocharian A, or whether the Hälädün cultural group of the early first millennium BCE in and near Kuča

⁴ It is tempting to consider the possibility that the apparently impressive technological advances brought by the Iranians speaking this Old Iranian language were the impetus for the split of Proto-Tocharian. At present no evidence for or against this scenario seems to be available.

(Debaine-Francfort 1988: 23) can be identified with early speakers of Tocharian B. Accordingly, archaeological evidence for the date of Proto-Tocharian or the place where it was spoken is presently indirect at best.

6.4 The Relationship of Tocharian to the Other Branches

It is now commonly held that Tocharian has no closer affinity to any other branch of Indo-European.⁵ Proposals for closer affinity have been made but have found little acceptance and concern superficial similarities, such as the spread of the *n*-stems in the nominal inflection, which would be shared with Germanic (Adams 1988: 5), or the endings in *-r* of the middle, suggesting a link with Italo-Celtic (e.g. Lane 1970: 78, who attributes the correspondence to post-Proto-Indo-European contact), and so on. References to and discussion of these and other suggestions can be found in Hackstein 2005 and Malzahn 2016: 281.

Not accepting any of the adduced old comparisons, Hackstein (2005) proposes instead several close matches between Tocharian and other branches in grammaticalisation processes. According to him, the observed grammaticalisation processes are independent and parallel instead of shared, and indicate post-Proto-Indo-European contact. The matches that he proposes are with Latin, Slavic, Gothic, Greek and Armenian. Although the cases discussed are interesting, the large number of languages in the comparison makes it unlikely that the parallelisms are due to early contact. In addition, it is open to debate whether the parallelisms, if correctly identified, are indeed so salient that they cannot have come about completely independently. For instance, the univerbation of interrogative and demonstrative in Toch.B *kuse* ‘who’ < **k^wi* + *so*, in Alb. *kush* ‘who’ < **k^wis* + *so*, and in OCS *kъto* ‘who’ with *-to* from PIE **tod* (Hackstein 2005: 177) has not proceeded in exactly the same way; it probably compensates, at least in part, for the loss of inflection and word weight; and it appears to be a natural process. Toch.B *ṣ* and *spä* ‘and’, which Hackstein derives from **h₁eti* and **h₁eti-h₁epi* respectively, in fact represent one and the same etymon **ṣpə* with simplification of *ṣp* to *ṣ* in classical and late Tocharian B (Peyrot 2008: 68) so that *ṣ* cannot be directly compared with Latin *et* or Gothic *iþ* (pace Hackstein 2005: 176).

A different case is presented by matches with Anatolian, of which several have been proposed that appear to be fairly solid: see for instance Pinault 2006a: 93. These must be archaisms, not showing any closer affinity between Anatolian and Tocharian, and are potentially relevant to establish the position of Tocharian in the tree of Indo-European, discussed in the following section.

⁵ The prolonged contact with Iranian and the shorter but dramatic impact of Indic are obviously to be discarded as secondary.

6.5 The Position of Tocharian

Tocharian is often claimed to have been the second branch to split off the Indo-European proto-language: after Anatolian, but before all other attested branches. This hypothesis may be called the “Indo-Tocharian” hypothesis, based on the model of Indo-Anatolian (Peyrot 2019b; see Figure 6.1). “Indo-Anatolian”, equivalent to “Indo-Hittite”, is used here in a technical sense for the highest node in the Indo-European tree, before Anatolian split off as the first branch, a scenario for which the evidence is steadily growing (cf. Kloekhorst & Pronk 2019).⁶ Strikingly, the arguments that have been advanced in support of the “Indo-Tocharian” hypothesis vary considerably: many authors making the same claim do not accept each other’s evidence for their claim. The most comprehensive systematic review is that by Ringe (1991), who finds hardly any evidence for the position of Tocharian in the family tree at all. Other relevant contributions include Lane 1970, Schmidt 1992, Winter 1997, Pinault 2013 and Malzahn 2016.

Below, a selection of arguments will be discussed. In general, it appears that aberrancies of Tocharian are due to innovation, and careful reconstruction tends to bring Tocharian closer to non-Anatolian Indo-European. The Indo-Tocharian

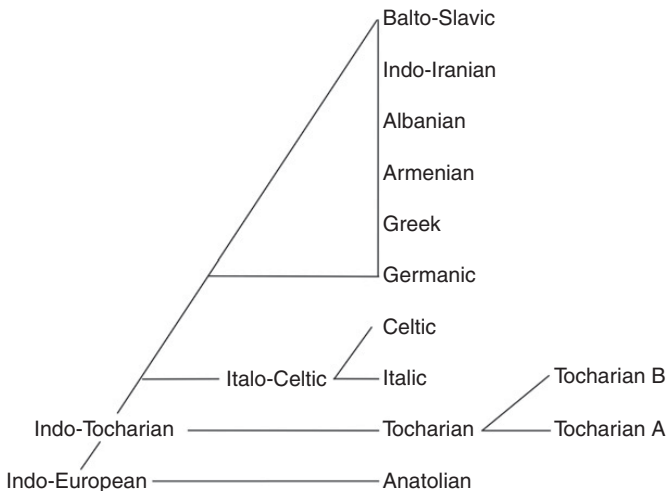


Figure 6.1 The position of Tocharian

⁶ Jasanoff (2017: 233–4) explicitly subscribes to this scenario but rejects the term “Indo-Hittite” because it “acquired tendentious overtones” (p. 233).

hypothesis still seems attractive, but evidence is slim and the difference between Indo-Anatolian and Indo-Tocharian appears to be much larger than that between Indo-Tocharian and the other Indo-European languages. If Indo-Anatolian can be dated to the middle of the fifth millennium BCE, Indo-Tocharian must be much closer to the middle of the fourth millennium. As pointed out to me by Tijmen Pronk, the split-off of the Tocharian branch (Anthony 2007: 305, 307–11; Anthony & Ringe 2015: 208, 211) may be associated with the apparent abandonment of the Caspian steppe in 3500–3400 BCE, probably due to abrupt aridification (Shishlina 2008: 220).

6.5.1 Methodology

In view of the many different arguments that have been proposed for the Indo-Tocharian hypothesis, a brief note on the methodology seems in order.

It is generally agreed that the assumption of an early Tocharian split-off must be based on shared innovations of the other non-Anatolian Indo-European languages. In particular, the branch that split off *after* Tocharian should have shared in such innovations. As the most likely candidate for the branch to have split off third appears to be Italo-Celtic, the supposed shared innovation should ideally be attested in this branch. Conversely, arguments for Indo-Anatolian should be based on shared innovations of non-Anatolian Indo-European, ideally attested also in Tocharian (Peyrot 2019b).

Though clear in theory, in practice finding and defining shared innovations is difficult. There appear to be the following requirements to shared innovations useful for phylogenetic subgrouping:

- *identifiability*: the linguistic element adduced as a shared innovation in the lower node should be clearly identifiable in the higher as well as in the lower node;
- *unidirectionality*: the observed difference with regard to the selected linguistic element should be interpretable as a unidirectional change;
- *salience*: the observed change should be so salient that it is unlikely to have occurred independently in the supposed lower-node branches, in which case it would be a parallel, not a shared innovation.

The requirement of unidirectionality is widely accepted, and discussion tends to focus on the question as to whether a given difference can be interpreted as a unidirectional change, rather than the need of this requirement as such. A case in point is semantic change: phylogenetic arguments based on semantic change are often contested on the grounds that a given semantic difference is not necessarily due to unidirectional change.

The requirement of identifiability, often implicit, may be helpful in discussions about debated phylogenetic arguments based on the loss or addition of features, or on lexical replacement. Arguments based on loss or addition are

notoriously difficult, as for instance with the comparative and superlative suffixes, which are unattested in Anatolian and Tocharian: have they been lost in both branches, or were they added after the Tocharian split-off? Such arguments cannot be applied if the supposedly added feature cannot be identified with any prestige leading to it or if the lost feature has left no trace at all. Arguments based on lexical replacement are weak because the identifiable element would be the meaning, expressed with different etyma in two branches. Meaning is difficult to use as an identifiable element, because several etyma may have similar, overlapping or even identical meanings, and it is therefore difficult to prove that a certain meaning came to be expressed with a different etymon.

The requirement of salience seems so obvious that no further explanation is needed.

6.5.2 Phonology

For our present purposes, phonological evidence appears to be of little relevance in view of the extensive changes in the Tocharian sound system, which are probably due to a Uralic substrate (see Section 6.2). In particular, evidence for the phonetic realisation of the stops in the proto-language has been obscured by this substrate effect. Thus, there is little evidence to establish the position of Tocharian with relation to Kloekhorst's claim that Anatolian preserves an older system of stop distinctions (2016), with classical PIE **t*, **d*, **d^h* from Proto-Indo-Anatolian **t₁*, **[?]t*, **t*.

For Tocharian, the developments of **d* and **b^h* are notable. PIE **d* is by default represented with *ts* and is otherwise often lost, at least before **i*, **u* and **r*, and so differs from **t* and **d^h*, whose default outcome is Tocharian *t*.⁷ Thus, even though the exact phonetics remain difficult to establish, **t* and **d^h* were apparently closer to each other than either of them were to **d*.⁸ At the same time, **b^h* is lost after **m*, for instance in **gomb^ho* > Toch.B *keme* 'tooth', while **p* stays, for instance in **temp-* (Lith. *tempiù* 'stretch') > Toch.B *cəmp-* 'be able'.⁹ This suggests that **b^h* was weaker than **p*: it may have been voiced,

⁷ Also, the palatalised reflex of **d* is *ś*, while that of **d^h* and **t* is *c*.

⁸ Possibly, this distribution also holds for the assibilated variant *-s* < **-ti*, **-d^hi* (Jasanoff 1987), although good evidence for the development of **-di* is thus far lacking.

⁹ With original **mt*, compare also **(d)kántóm* '100' > Toch.B *kante*. Parallel cases with **d^h*, **g^h*, **g^h* and **g^{wh}* are not readily available. Ringe discusses the possibility that the Toch.B subj. stem of *lät-* 'go out' as in the inf. *lantsi* shows /lən-/ < **h₁lucnd^h* (1996: 43). However, he notes that forms with a geminate *nn* like 1sg. *lannu* 'I will go out' rather suggest an original **ləntn-*. Indeed, all forms with a nasal in Toch.B can probably be derived from *lənn-* < **ləntn-*, which arose secondarily through suffixation with **-nəsk-* in the present (Peyrot 2013: 446). Toch.B *lan_utse* 'light' < **h₁leng^hu-* shows that **g^{wh}* was not lost after **n*. It may be supposed that **g^h* and **g^h* were not lost after **n* either. The reason for this exception could be that there was no corresponding velar nasal phoneme and the velar stop had to remain in order to keep the velar nasal allophone.

fricative or both. It is tempting to compare the typologically common loss of voiced stops after nasals, as in English *lamb* /læm/, and posit the value [b] for $*b^h$, but this is certainly not the only option. Combining the evidence from dentals and labials, it appears that the stop system inherited by Tocharian had strong stops for the conventional voiceless stops like $*t$, weak stops for the conventional voiced aspirated stops like $*d^h$, and a series that was different from both for the conventional voiced stops like $*d$.¹⁰ Although Tocharian offers no direct evidence for the reconstruction of glottalic stops in Proto-Indo-European, the fact that $*d$ has a different reflex from $*t$ and $*d^h$ is neatly compatible with it, since under Kortlandt's glottalic theory (e.g. 1985; 2018a) $*d$ [ʔd] on the one hand is set apart from $*t$ and $*d^h$ on the other.

Nevertheless, the value for the phylogenetic position of Tocharian remains undecided. Since there is strong evidence for $*d = *d$ in classical Indo-European, this feature cannot be used. Further, the position of Tocharian cannot be determined with regard to Kloekhorst's claim that classical PIE $*t$ (perhaps phonetically [t]) < Proto-Indo-Anatolian $*t$: and classical PIE $*d^h$ (perhaps phonetically [d]) < Proto-Indo-Anatolian $*t$, since both phonetic stages are compatible with $*t$ being stronger and $*d^h$ being weaker.¹¹

It has been argued that Tocharian shows consonantal reflexes of PIE $*H$ as *k* (e.g. Winter 1965: 206–10; Schmidt 1988; Kortlandt 2018b). Winter adduces Tocharian A “intrusive *k*” as a consonantal reflex of $*HH$, e.g. gen.pl. *lwākis* to nom.-obl.pl. *lwā* ‘animals’ or perl.pl. *puklākā* to nom.-obl.pl. *puklā* ‘years’. However, *k* must be secondary in such examples because it effectively prevents the problematic vowel contractions in the morphologically expected forms $**lwes < *lwā.is$ (next to attested gen.sg. *lwes*!) and $**puklā < *puklā.ā$. Schmidt (cf. also Hartmann 2001) has argued that the *k* in roots ending in *-tk* goes back to $*h_2$, but Melchert's earlier derivation of *-tk-* from *-T-sk-* is definitely to be preferred (1977; cf. also Pinault 2006b). Kortlandt's derivation of Tocharian B *taka-* ‘be’ from $*steh_2-t$ with *-k-* from $*h_2$ is in itself attractive, but since the “*k*-aorist” is also attested in e.g. Gr. *ἔθηκα* and Lat. *fēcī*, this reflex cannot be used to determine the phylogenetic position of Tocharian, even if the

¹⁰ A thorough discussion of these developments can be found in Ringe 1996: 39–66 and Winter 1962. In both accounts, a complicating factor is the Tocharian version of Grassmann's Law, exemplified by e.g. Toch.B *tsik-* ‘form’ < $*d^heig^h$ and *tsək-* ‘burn’ < $*d^heg^{wh}$, allegedly with *ts* < $*d$ after $*d^h$ had been deaspirated to $*d$ before the following $*g^h$ and $*g^{wh}$, respectively. The evidence for Grassmann's Law in Tocharian is circumstantial and probably open to an alternative explanation. It is not taken into account here in view of the solid counterexample of Toch.B *tapre* ‘high’ < $*d^hub^hro-$, to be reconstructed with $*b^h$ instead of $*b$ after Kroonen (2011: 253, 255).

¹¹ It is possible that Tocharian inherited a stop system in which distinctive voice had not yet developed, as argued by Kortlandt (e.g. 1985: 197; 2020: 269), but in my view this is difficult to prove.

evidence as such nicely fits Kloekhorst's reconstruction of $*h_2$ and $*h_3$ as uvular stops for Proto-Indo-Anatolian (2018; cf. also Kortlandt 2002: 218).

Like other Indo-European languages, Tocharian shows reflexes of metathesis of $*Hi$ to $*iH$ and $*Hu$ to $*uH$. For instance, metathesis of $*Hu$ to $*uH$ is attested by such forms as Toch.B *puwar* 'fire' < $*puh_2r$ ¹² (as in Greek $\pi\delta\rho$) from earlier $*peh_2-ur$ (as in Hitt. *paḥḥur*) and Toch.B *ləw(a)-* 'rub' (prt.3sg.-3sg.obj. *lyawā-ne* 'he rubbed him') < $*leuh_3-$ from earlier $*leh_3u-$ (as in Hitt. *lāḥu-i* 'pour'). Even though unmetathesised forms are also found, for instance Toch.B *kaw-* 'kill' < $*keh_2u-$, the existence of metathesised forms in Tocharian clearly shows that this sound change is to be dated before Tocharian split off. However, even though Hittite often shows unmetathesised forms next to metathesised forms elsewhere (Kloekhorst & Pronk 2019: 5), the metathesis must have already occurred before Proto-Indo-Anatolian on the evidence of forms such as Hitt. *šuhḫa-* 'pour, sprinkle' < $*suh_2-$ next to *išḫu(wa)-* < $*seh_2u-$ and *lu-u-* 'pour' < $*luh_3-$ next to *lāḥu-* < $*leh_3u-$ (Melchert 2011: 129, 131). At this point, therefore, the mere attestation of laryngeal metathesis cannot be used for inner Indo-European phylogeny.

However, another Indo-European metathesis may be used: that of word-final $*-ur$ to $*-ru$ (Lubotsky 1994: 99–100). This sound change seems to have occurred only after Proto-Indo-Anatolian. Strong evidence for it in Tocharian has been discovered by Del Tomba (2021), who shows that Toch.B plurals in *-wa* to nouns in *-r*, such as *tarkār* 'cloud', pl. *tārkarwa*, presuppose metathesis of $*-ur$ to $*-ru$ in the singular, on which the plural *-r-wa* < $*-ru-h_2$ was built. Although this sound change may be used for the phylogeny of Indo-European, it clearly groups Tocharian together with the non-Anatolian languages.

6.5.3 Morphology

Morphology is the domain that is often ascribed the highest potential to yield evidence for the phylogenetic position of Tocharian. Indeed, morphology meets two essential needs: it is constantly in the process of change, and, at the same time, shifts in function, though commonplace, are subject to

¹² The Tocharian word for 'fire' is variously reconstructed. Hackstein, for instance, reconstructs $*ph_2u\bar{o}r$ (2017: 1314). It is, however, questionable whether $*h_2$ would be lost in this context, and whether the reconstruction of a collective ending $*-ōr$ for this etymon is warranted. A derivation of Toch.B *puwar* from $*puh_2r$ is the most straightforward. Winter (1965: 192) reconstructs the Tocharian A equivalent *por* as $*paur$ from unmetathesised $*peh_2-ur$. This is phonologically possible but most difficult morphologically, since it is not clear what the distribution of these variants in the Proto-Tocharian paradigm might have been. It is therefore preferable to assume a development $*wa > o$ similar to $*we > o$ in *koṃ*, obl.sg. of *ku* 'dog', < $*kwen$ and $*iye > e$ in *karemām* 'laughing' < $*keriyemane$ (Hilmarsson 1989: 135; Hackstein 2017: 1314; Peyrot 2012a: 210).

limitations. Unfortunately, Tocharian morphology is heavily reorganised and its prehistory is often very obscure. Even worse is the fact that the reconstruction of Proto-Indo-European in exactly the relevant points is difficult and often disputed.

Without a doubt, the most prominent argument for phylogeny based on morphology that has been advanced comes from the Tocharian *s*-preterite. In the active of the Tocharian *s*-preterite, an element *s* is only found in the 3sg.: 1sg. *prek-uwa* ‘asked’, 2sg. *prek-asta*, 3sg. *prek-sa*, 1pl. *prek-am*, 2pl. *prek-as**, 3pl. *prek-ar*. This is reminiscent of the Hittite *hi*-preterite, which likewise has *-š* only in the 3sg.: 3sg. *ākkiš* ‘died’, 3pl. *aker* (Pedersen 1941: 146). There are two schools of thought to explain this correspondence. The first, most prominently voiced by Jasanoff (e.g. 2003: 204–5),¹³ takes the restriction of the *-s-* as an archaism of Anatolian and Tocharian, while the rise of the classical *s*-aorist through generalisation of the *-s-* from the 3sg. throughout the whole paradigm is a common innovation of the other Indo-European branches. According to the second one, the *-š* in Hittite is secondary, probably somehow from the *s*-aorist, while in Tocharian the *s*-preterite forms without *-s-* lost it due to the effects of sound law and analogy (Ringe 1990; Kortlandt 1994; Peyrot 2013: 503–7). The matter cannot be treated here in detail. Suffice to say that the assumption of loss of *-s-* accounts best for the inflection of the Tocharian preterite and its patternings with the subjunctive. At any rate, this famous case very clearly shows how different views on the reconstruction of Proto-Indo-European logically lead to different evaluations of arguments for phylogeny.

Another phylogenetic argument is based on the middle endings in *-r* (e.g. Ringe, Warnow & Taylor 2002; Ringe 1991: 98–9). It is widely held that the shorter middle endings 3sg. **-to* and 3pl. **-nto* were secondary endings in Proto-Indo-European, while the corresponding primary endings were originally 3sg. **-to-r*, 3pl. **-nto-r*, which were later replaced by 3sg. **-to-i*, 3pl. **-nto-i*, marked with the productive primary marker **-i* as found in the active endings. This would not be a valid argument for Indo-Tocharian, since the *r*-endings are also found in Italo-Celtic and Phrygian, but it would group Tocharian with the older branches.

However, a number of problems with this argument need to be noted:

- It is questionable as to whether the contrast between Toch.B pres. 3sg. *-tär*, 3pl. *-ntär* and pret. 3sg. *-te*, 3pl. *-nte* continues an original primary–secondary contrast, because the Tocharian preterite active endings do not continue the secondary endings of Proto-Indo-European. In the copula 3sg. *ste*, 3pl. *skente*, the endings *-te*, *-nte* are even used as present endings. Hackstein (1995: 273–5) explains these forms as original resultatives, i.e. “is” < “has

¹³ Two more recent contributions are Melchert (2015) and Jasanoff (2019).

become”, and notes that presentic readings of preterites are found elsewhere. However, it remains problematic as to why no shade of the past meaning has been preserved in 3sg. *ste*, 3pl. *skente*, and why the corresponding suffixed forms, such as 3sg.-1sg.obj. *star-ñ*, have present endings. This distribution is difficult to explain from an original difference in tense.

- The reconstruction of the primary middle endings 3sg. **-to-r*, 3pl. **-nto-r* is problematic itself. Indeed, Lat. *-tur*, *-ntur* point to **-tor*, **-ntor*. However, as Weiss (2009: 413) notes, Osc. 3sg. *-ter*, 3pl. *-nter* point to **-tro*, **-ntro*, and Umb. primary *-ter*, *-nter* vs. secondary *-tur*, *-ntur* suggests Proto-Italic primary **-tro*, **-ntro* vs. secondary **-tor*, **-ntor*. Likewise, the Old Irish deponent endings 3sg. *-thir*, 3pl. *-tir* point to **-tr-*, **-ntr-*, probably **-tro*, **-ntro*. Finally, Toch. *-tär*, *-ntär* cannot be derived from **-tor*, **-ntor* directly (cf. also Pinault 2010b). Ringe (1996: 86) discusses the change of **-or* to Toch. **-ər*, but the 3rd person middle endings are his only evidence, against counterexamples such as Toch.B *mallwer* ‘milk’, with suffix *-wer* < **-uor* as in the verbal abstract, e.g. *šesüwer* ‘eating’. A further counterexample seems to be *yerter* ‘felloe’, which on the evidence of the unpalatalised *-t-* must reflect **-tor*.¹⁴
- The assumed replacement of well-marked middle paradigms ending in *-r* with the active marker *-i* is difficult to understand. What would be the motivation to do so? If endings are clearly marked to be primary, there seems no need to replace them. The greatest difficulty here is not the addition of the primary active marker *-i* – such additions are indeed found frequently in e.g. the perfect endings, such as OCS *vědě*, Lat. *vidī*, or Toch.A *kärse* ‘I knew’ < **kər̥sa-a-i* – but the fact that the transparent middle ending **-r* should have been deleted.

In view of these problems, it is tempting to follow Kortlandt’s reconstruction (1981) of the middle endings as **-to*, **-nto* only, without contrast between primary and secondary endings.¹⁵ Apparently such contrasts were created independently in the different branches. In any case, the problematic specifics of the reconstruction of the middle endings make them difficult to use for phylogeny.

Another argument advanced by Ringe, Warnow & Taylor (2002: 117) is the thematic optative in **-o-ih₁-*, attested in Indo-Iranian, Greek, Balto-Slavic and Germanic, but not in Tocharian. Indeed, this may be a later innovation within Indo-European not shared by Tocharian. In Tocharian, there is only one variant

¹⁴ A possible alternative reconstruction would be **-ewer* with contraction of **ewe* to *e*.

¹⁵ The evidence of Anatolian seems compatible with an original **-to*, **-nto* without contrast between primary and secondary endings: synchronically, they are attested in Hittite as pres.3sg. *-tta*, 3pl. *-anta*. However, a derivation from **-tor*, **-ntor* neatly explains the rise of the present particle *-ri* from resegmentation after the loss of *-r* after **-ó-* (Yoshida 1990). The introduction of the particle *-ti*, to mark the preterite endings, i.e. 3sg. *-ttati*, 3pl. *-antati*, would be motivated in both scenarios.

of the optative suffix, *-i-* (*i* with preceding palatalisation), to be derived from **-ih₁-*.¹⁶ However, “present optatives”, synchronically imperfects, are unattested in Tocharian A, and they must consequently have been regularised secondarily in Tocharian B (Peyrot 2012b). Therefore, it is difficult to prove that e.g. Toch.B *pari** ‘he took’ goes back directly to **b^her-ih₁-t* (for **b^her-o-ih₁-t* elsewhere). In any case, since the thematic optative is not attested in Italo-Celtic, it cannot be used to show that Tocharian split off before that branch.

It has been argued that the combination of the Tocharian present participle in *-mane* with both active and middle finite inflection is an archaism: the verbal adjective **-mh₁no-* would originally have been indifferent for voice, very much like the **-nt-*participle in Anatolian (Kloekhorst & Pronk 2019: 3), and became specialised only later, after Tocharian split off, as the middle counterpart of the active **-nt-*participle (Pinault 2012: 229; Peyrot 2017: 339–40). However, I now think that this argument has to be abandoned in the light of a study by Friis (2021), who shows that traces of specifically middle use are preserved, which suggests that active use of *-mane* in Tocharian is secondary.¹⁷

A case from word formation in the grammatical domain is the interrogative stem in **m-* found in Anatolian and Tocharian (Hackstein 2004: 280–3; Pinault 2010a: 359; Peyrot 2019a: 195–9). A weak point of this argument is that the innovation of the other Indo-European languages would consist only in *loss* of the *m*-interrogative, while a strong point is the central position of this stem, paired only with **k^wi-* (**k^we-*, **k^wo-*), in the linguistic system. Thus, while the identifiability of this feature is low, its salience is nevertheless high.

6.5.4 *Lexicon*

Lexical evidence has been variously evaluated. Important papers adducing lexical evidence in support of an early split off of Tocharian are Schmidt 1992 and Winter 1997. This evidence, and the method as a whole, was criticised by Hackstein (2005: 172) and Malzahn (2016) amongst others. For lexical arguments, a distinction should be made between lexical replacements and semantic change.

¹⁶ The full grade variant ***-ye-* < **-ieh₁-* may have been ousted by the zero grade variant through paradigmatic levelling, but it is also possible that the zero grade variant was generalised from *s*-aorist optatives with **-ih₁-* throughout (if the synchronic optatives of root subjunctives of class 1, such as Toch.B *parśi* ‘may he ask’, are to be derived from *s*-aorist optatives, i.e. in this case **prék-s-ih₁-t*).

¹⁷ Thus, even though I cannot agree with the arguments adduced by Fellner & Grestenberger (2018), I do now concur with their main claim.

Arguments based on lexical replacement are especially difficult because the identifiability requirement is not easily satisfied: it is hard to prove that two words did not carry the same or a similar meaning. An example of such an argument is Anatolian (i.e. Luvian) and Tocharian (i.e. Toch.A) **uel(H)-* ‘die’ vs. **mer-* elsewhere (Ringe, Warnow & Taylor 2002: 99).¹⁸ Although **mer-* indeed acquired the meaning ‘die’ from ‘disappear’ after Indo-Anatolian (Kloekhorst & Pronk 2019: 3), and thus became a new word for the meaning ‘die’, the Luvian and Tocharian A words cannot be shown to represent the original word for ‘die’, let alone that it was ousted by the new **mer-* (see Malzahn 2016: 285–6).

Another example is **h₁eg^{wh}-* ‘drink’, well attested in Tocharian and Anatolian, as against **peh₃-* elsewhere (Ringe, Warnow & Taylor 2002: 99). This may indeed be a case of lexical replacement, i.e. the meaning ‘drink’ came to be expressed by a different word. However, the details are complicated: Hitt. *pāš-i* ‘swallow’ shows that **peh₃-* needs to be reconstructed for Proto-Indo-Anatolian, with possibly only a slightly different meaning; and Lat. *ēbrius* ‘drunk’ and Gr. *νίφω* ‘be sober’ show that **h₁eg^{wh}-* was preserved after Tocharian split off, possibly with a shift to ‘be drunk’ (Peyrot 2019b). Thus, the argument for lexical replacement remains fragile, while the best phylogenetic evidence is formed by the possible semantic developments of ‘drink’ to ‘be drunk’ for **h₁eg^{wh}-*, and ‘swallow’ to ‘drink’ for **peh₃-*. The attestation of the meaning ‘be drunk’ in Latin is favourable for the Indo-Tocharian hypothesis, because it suggests that this semantic change occurred after Tocharian split off, but before Italo-Celtic split off.

As a lexical argument based on semantics, Ringe, Warnow & Taylor (2002: 99) adduce **meǵh₂*, of which the Anatolian (e.g. Hitt. *mekk-*, *mekki-*) and Tocharian (e.g. Toch.B *māka*) reflexes mean ‘much, many’, as against ‘great’ elsewhere. The distribution is especially neat in this case, since the etymon is also attested in Italo-Celtic (OIr. *maige*, Lat. *magnus*, etc.) and Germanic (Goth. *mikils*). Here, the main problem is the requirement of unidirectionality: the meanings are contingent and a change from ‘great’ to ‘much’ is by no means unlikely.

References

- Adams, Douglas Q. 1988. *Tocharian historical phonology and morphology*. New Haven, CT: American Oriental Society.
- Adams, Douglas Q. 2013. *A dictionary of Tocharian B*. Amsterdam: Rodopi.
- Anthony, David W. 2007. *The horse, the wheel, and language*. Princeton, NJ: Princeton University Press.

¹⁸ Their argument about **ai-* ‘give’ cannot be upheld with the reconstruction of Hitt. *pāi* ‘gives’ as **h₁p-oi-e* by Kloekhorst (2006).

- Anthony, David W. & Don Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1. 199–219.
- Burrow, Thomas. 1935. Tocharian elements in the Kharoṣṭhī documents from Chinese Turkestan. *Journal of the Royal Asiatic Society* 1935. 667–75.
- Debaine-Francfort, Corinne. 1988. Archéologie du Xinjiang des origines aux Han. Première partie. *Paléorient* 14. 5–29.
- Debaine-Francfort, Corinne. 1989. Archéologie du Xinjiang des origines aux Han. Deuxième partie. *Paléorient* 15. 183–213.
- Del Tomba, Alessandro. 2021. Metathesis of PIE **-ur* in Tocharian. *Münchener Studien zur Sprachwissenschaft* 74. 51–85.
- Dragoni, Federico, Niels Schoubben & Michaël Peyrot. 2020. The Formal Kharoṣṭhī script from the Northern Tarim Basin in Northwest China may write an Iranian language. *Acta Orientalia Hungarica* 73. 335–73.
- Fellner, Hannes & Laura Grestenberger. 2018. Die Reflexe der **-nt-* und **-mh₁no-* Partizipien im Hethitischen und Tocharischen. In Elisabeth Rieken (ed.), *100 Jahre Entzifferung des Hethitischen*, 63–82. Wiesbaden: Reichert.
- Friis, Louise S. 2021. Tocharian B agent nouns in *-ntsa* and their origin. *Indo-European Linguistics* 9. 1–25.
- Hackstein, Olav. 1995. *Untersuchungen zu den sigmatischen Präsensstambildungen des Tocharischen*. Göttingen: Vandenhoeck & Ruprecht.
- Hackstein, Olav. 2004. From discourse to syntax. The case of compound interrogatives in Indo-European and beyond. In Karlene Jones-Bley et al. (eds.), *Proceedings of the fifteenth annual UCLA Indo-European Conference*, 257–98. Washington, DC: Institute for the Study of Man.
- Hackstein, Olav. 2005. Archaismus oder historischer Sprachkontakt. Zur Frage westindogermanisch-tocharischer Konvergenzen. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 169–84. Wiesbaden: Reichert.
- Hackstein, Olav. 2017. The phonology of Tocharian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1304–35. Berlin: De Gruyter Mouton.
- Hartmann, Markus. 2001. Wiederum zu den tocharischen Verben auf *-tk-*. *Münchener Studien zur Sprachwissenschaft* 61. 95–117.
- Hilmarsson, Jörundur. 1989. *The dual forms of nouns and pronouns in Tocharian*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Hilmarsson, Jörundur G. 1991. *The nasal prefixes in Tocharian. A study in word formation*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Isebaert, Lambert. 1980. *De Indo-Iraanse bestanddelen in de Tocharische woordenschat*. PhD dissertation, University of Leuven.
- Jasanoff, Jay H. 1987. Some irregular imperatives in Tocharian. In Calvert Watkins (ed.), *Studies in memory of Warren Cowgill (1929–1985)*, 92–112. Berlin: De Gruyter.
- Jasanoff, Jay H. 2003. *Hittite and the Indo-European verb*. Oxford: Oxford University Press.
- Jasanoff, Jay H. 2017. The impact of Hittite and Tocharian: Rethinking Indo-European in the 20th century and beyond. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 1, 220–38. Berlin: De Gruyter Mouton.

- Jasanoff, Jay H. 2019. The sigmatic forms of the Hittite verb. *Indo-European Linguistics* 7. 13–71.
- Kloekhorst, Alwin. 2006. Hittite *pai-/pi-* ‘to give’. *Indogermanische Forschungen* 111. 110–19.
- Kloekhorst, Alwin. 2016. The Anatolian stop system and the Indo-Hittite hypothesis. *Indogermanische Forschungen* 121. 213–47.
- Kloekhorst, Alwin. 2018. Anatolian evidence suggests that the Indo-European laryngeals **h₂* and **h₃* were uvular stops. *Indo-European Linguistics* 6. 69–94.
- Kloekhorst, Alwin & Tijmen Pronk. 2019. Introduction: Reconstructing Proto-Indo-Anatolian and Proto-Indo-Uralic. In Alwin Kloekhorst & Tijmen Pronk (eds.), *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*, 1–14. Leiden: Brill.
- Kortlandt, Frederik. 1981. 1st sg. middle **-H₂*. *Indogermanische Forschungen* 86. 123–36.
- Kortlandt, Frederik. 1985. Proto-Indo-European glottalic stops: The comparative evidence. *Folia Linguistica* 6. 183–201.
- Kortlandt, Frederik. 1994. The fate of the sigmatic aorist in Tocharian. In Bernfried Schlerath (ed.), *Tocharisch*, 61–5. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Kortlandt, Frederik. 2002. The Indo-Uralic verb. In Rogier Blokland & Cornelius Hasselblatt (eds.), *Finnic-Ugrians and Indo-Europeans: Linguistic and literary contacts*, 217–27. Maastricht: Shaker.
- Kortlandt, Frederik. 2018a. Proto-Indo-European glottalic stops: The evidence revisited. *Münchener Studien zur Sprachwissenschaft* 71. 147–53.
- Kortlandt, Frederik. 2018b. The Indo-European *k*-aorist. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 137–42. Ann Arbor, MI: Beech Stave.
- Kortlandt, Frederik. 2020. The disintegration of the Indo-European language family. In Harald Bichlmeier, Ondřej Šefčík & Roman Sukač (eds.), *Etymologus: Festschrift for Václav Blažek*, 267–76. Hamburg: Baar.
- Kroonen, Guus J. 2011. False exceptions to Winter’s Law: On the effects of Kluge’s Law on the Proto-Germanic consonantism. In Tijmen Pronk & Rick Derksen (eds.), *Accent matters: Papers on Balto-Slavic accentology*, 251–61. Amsterdam: Rodopi.
- Lane, George S. 1966. On the interrelationships of the Tocharian dialects. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 213–33. Berkeley: University of California Press.
- Lane, George S. 1970. Tocharian: Indo-European and non-Indo-European relationships. In George Cardona et al. (eds.), *Indo-European and Indo-Europeans*, 73–88. Philadelphia: University of Pennsylvania.
- Lubotsky, Alexander. 1994. Avestan *θβōrəstar-* and the Indo-European root *√turk-*. *Die Sprache* 36. 92–102.
- Mallory, James P. 2015. *The problem of Tocharian origins*. Philadelphia: University of Pennsylvania.
- Malzahn, Melanie. 2016. The second one to branch off? The Tocharian lexicon revisited. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 281–92. Wiesbaden: Reichert.

- Melchert, H. Craig. 1977. Tocharian verb stems in *-tk-*. *Zeitschrift für vergleichende Sprachforschung* 91. 93–130.
- Melchert, H. Craig. 2011. The PIE verb for ‘to pour’ and medial **h₃* in Anatolian. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 22nd annual UCLA Indo-European Conference*, 127–32. Bremen: Hempen.
- Melchert, H. Craig. 2015. The Tocharian *s*-preterite. In Melanie Malzahn et al. (eds.), *Tocharian texts in context*, 127–35. Bremen: Hempen.
- Pedersen, Holger. 1941. *Tocharisch vom Gesichtspunkt der indoeuropäischen Sprachvergleichung*. København: Munksgaard.
- Peyrot, Michaël. 2008. *Variation and change in Tocharian B*. Amsterdam & New York: Rodopi.
- Peyrot, Michaël. 2012a. The Tocharian A match of the Tocharian B obl.sg. *-ai*. *Tocharian and Indo-European Studies* 13. 181–220.
- Peyrot, Michaël. 2012b. Tocharian ‘eat’ and the strong imperfect in Tocharian A. In Olav Hackstein & Ronald I. Kim (eds.), *Linguistic developments along the Silk Road: Archaism and innovation in Tocharian*, 85–119. Vienna: Österreichische Akademie der Wissenschaften.
- Peyrot, Michaël. 2013. *The Tocharian subjunctive: A study in syntax and verbal stem formation*. Leiden: Brill.
- Peyrot, Michaël. 2017. On the part of speech and the syntax of the Tocharian present participle. In Claire Le Feuvre, Daniel Petit & Georges-Jean Pinault (eds.), *Verbal adjectives and participles in Indo-European languages*, 327–41. Bremen: Hempen.
- Peyrot, Michaël. 2018. Tocharian B *etswe* ‘mule’ and Eastern East Iranian. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 270–83. Ann Arbor, MI: Beech Stave.
- Peyrot, Michaël. 2019a. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European Linguistics* 7. 72–121.
- Peyrot, Michaël. 2019b. Indo-Uralic, Indo-Anatolian, Indo-Tocharian. In Alwin Kloekhorst & Tijmen Pronk (eds.), *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*, 186–202. Leiden: Brill.
- Pinault, Georges-Jean. 2006a. Retour sur le numéral ‘un’ en tokharien. *Indogermanische Forschungen* 111. 71–97.
- Pinault, Georges-Jean. 2006b. Sur l’évolution phonétique *tsk* > *tk* en tokharien commun. *Münchener Studien zur Sprachwissenschaft* 62. 103–56.
- Pinault, Georges-Jean. 2010a. Le pronom d’ipséité en tokharien. In Injoo Choi-Jonin, Marc Duval & Olivier Soutet (eds.), *Typologie et comparatisme: Hommages offerts à Alain Lemaréchal*, 351–65. Leuven: Peeters.
- Pinault, Georges-Jean. 2010b. On the *r*-endings of the Tocharian middle. In Ronald I. Kim et al. (eds.), *Ex Anatolia lux: Anatolian and Indo-European studies in honor of H. Craig Melchert on the occasion of his sixty-fifth birthday*, 285–95. Ann Arbor, MI: Beech Stave.
- Pinault, Georges-Jean. 2012. Sound laws and the suffix of the PIE “middle” participle. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphonemics*, 227–51. Munich: LINCOM.
- Pinault, Georges-Jean. 2013. Anatolien et tokharien: Des langues décisives pour la reconstruction indo-européenne. *Histoire Épistémologie Langage* 35. 13–44.

- Ringe, Don. 1990. The Tocharian active *s*-preterite, a classical sigmatic aorist. *Münchener Studien zur Sprachwissenschaft* 51. 183–242.
- Ringe, Don. 1991. Evidence for the position of Tocharian in the Indo-European family? *Die Sprache* 34. 59–123.
- Ringe, Don. 1996. *On the chronology of sound changes in Tocharian*. Vol. 1. *From Proto-Indo-European to Proto-Tocharian*. New Haven, CT: American Oriental Society.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Schmidt, Klaus T. 1988. Stellungsbedingte Konsonantierung von a_2 im Tocharischen? In Alfred Bammesberger (ed.), *Die Laryngalthorie und die Rekonstruktion des indogermanischen Laut- und Formensystems*, 471–80. Heidelberg: Winter.
- Schmidt, Klaus T. 1992. Archaismen des Tocharischen und ihre Bedeutung für Fragen der Rekonstruktion und der Ausgliederung. In Robert S. P. Beekes, Alexander M. Lubotsky & Jos Weitenberg (eds.), *Rekonstruktion und relative Chronologie*, 101–14. Innsbruck: Institut für Sprachwissenschaft.
- Schmidt, Klaus T. 2018. *Nachgelassene Schriften*. Vol. 1. *Ein westtocharisches Ordinationsritual*; Vol. 2. *Eine dritte tocharische Sprache: Lolanisch*. Ed. by Stefan Zimmer. Bremen: Hempen.
- Schoubben, Niels. 2021. Accent sign matters. The Niya Prakrit grapheme <ḱ> and its connection to Bactrian <ḥk>. *Journal Asiatique* 309. 47–59.
- Shishlina, Natalia. 2008. *Reconstruction of the Bronze Age of the Caspian Steppes: Life styles and life ways of pastoral nomads*. Oxford: Archaeopress.
- Tremblay, Xavier. 2005. Irano-Tocharica et Tocharo-Iranica. *Bulletin of the School of Oriental and African Studies* 68. 421–49.
- Warries, Abel R. In press. Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems. *Indo-European Linguistics*.
- Weiss, Michael. 2020. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor, MI: Beech Stave.
- Winter, Werner. 1962. Die Vertretung indogermanischer Dentale im Tocharischen. *Indogermanische Forschungen* 67. 16–35.
- Winter, Werner. 1965. Tocharian evidence for Proto-Indo-European laryngeals. In Werner Winter (ed.), *Evidence for laryngeals*, 190–211. The Hague: Mouton.
- Winter, Werner. 1997. Lexical archaisms in the Tocharian languages. In Hans H. Hock (ed.), *Historical, Indo-European, and lexicographical studies: A festschrift for Ladislav Zgusta on the occasion of his 70th birthday*, 183–93. Berlin: Mouton de Gruyter.
- Yoshida, Kazuhiko. 1990. *The Hittite mediopassive endings in -ri*. Berlin: de Gruyter.