



BRILL



brill.com/ieul

The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence

Michaël Peyrot

Leiden University

m.peyrot@hum.leidenuniv.nl

Abstract

Tocharian agglutinative case inflexion as well as its single series of voiceless stops, the two most striking typological deviations from Proto-Indo-European, can be explained through influence from Uralic. A number of other typological features of Tocharian may likewise be interpreted as due to contact with a Uralic language. The supposed contacts are likely to be associated with the Afanas'evu Culture of South Siberia. This Indo-European culture probably represents an intermediate phase in the movement of speakers of early Tocharian from the Proto-Indo-European homeland in the Eastern European steppe to the Tarim Basin in Northwest China. At the same time, the Proto-Samoyedic homeland must have been in or close to the Afanas'evu area. A close match between the Pre-Proto-Tocharian and Pre-Proto-Samoyedic vowel systems is a strong indication that the Uralic contact language was an early form of Samoyedic.

Keywords

Tocharian – typology – substrate – Uralic – Samoyedic – Yeniseian – Ket – Afanas'evu

1 Introduction

The Tocharian languages, once spoken on the Silk Road from Kuča to Turfan in the Tarim Basin in present-day Northwest China, were without any trouble identified as Indo-European from the beginning of their study (Sieg & Siegling 1908). Yet they show several strikingly non-Indo-European typological traits, such as a single obstruent series of voiceless stops and agglutinative case inflex-

ion. Although there is strictly speaking no Indo-European type, as all daughter languages have diverged to different degrees from the proto-language, the typological position of Tocharian is odd (Schulze 1927:177). In this paper, I will argue that the Tocharian language type has to be seen in a South Siberian context. Indeed, many of the defining traits of Tocharian may be attributed to contact with an early form of Samoyedic, probably in the form of substrate influence.

1.1 *Tocharian typological oddities*

In a number of crucial points, Tocharian has undergone a typological shift compared to the Indo-European proto-language. The most important of these typological deviations are the following:

- Only voiceless stops, resulting from a merger of the Proto-Indo-European triple series, for instance **k*, **g*, **g^h*, into a single series, for instance *k*.
- A restructured vowel system without distinctive length. Among the many vowel changes leading to the Tocharian vowel system there is a remarkable shift PIE **o* > Toch.B *e*, Toch.A *a*.
- Agglutinative case marking with the non-Indo-European cases causal, comitative, perlativ, and without the Indo-European dative case.
- Tocharian has a relatively archaic, Indo-European-looking verb, with, nevertheless, a remarkably highly developed system of derived causatives, transitives and intransitives.
- The absence of preverbs and almost complete absence of any prefixing morphology.¹

Some of these developments could and have been explained through language-internal developments, even such heavy restructurings as in the vowel system. However, in view of the enormous consequences for the lexicon of the merger of three stop series into one, which must have led to massive homonymy, this will always be difficult to account for by internal change only. Therefore, the option of an explanation based on external influence is to be investigated seriously.

Apart from difficulties with a language-internal explanation, something that is difficult to objectify, there are a number of other obvious requirements for an explanation based on external influence:

- There need to be parallels between the source language, which exerts the influence, and the target language, which undergoes the influence.

1 Proto-Indo-European probably had no preverbs. Preverbs are frequent in Indo-European languages, but must have arisen secondarily from adverbs. Whether this has also happened in an early stage of Tocharian is unknown. In any case, the absence of preverbs conforms to the Uralic type discussed in this paper.

- The parallels observed need to be salient, that is, they are unexpected in the target language (for instance, related languages are different), and they are unlikely to result from trivial, commonplace tendencies.
- In order to exclude a chance similarity, the parallels observed need to be either sufficiently exact, or they should occur in a larger set of parallels all attributable to one source language.
- There needs to be a historical scenario accounting for the assumed influence: there must be a time and place in which the languages may effectively have been in contact.

As I will try to show, all these requirements are met in the case of very early forms of Proto-Tocharian and Proto-Samoyedic, that is, Pre-Proto-Tocharian and Pre-Proto-Samoyedic. At the same time, a considerable degree of uncertainty remains due to the large time depth involved. In this sense, I do not claim to have reached definitive conclusions on any of the points discussed, apart from the fact that external influence in Tocharian can be successfully studied. The main aim is to outline new perspectives for a field of research that has thus far remained largely unexplored.²

1.2 *The Tocharian Migration Hypothesis*

As I will try to show, the typological position of Tocharian has to be seen against the background of the prehistory of the language. The Tocharian branch is often argued to have split off the Indo-European proto-language at an early stage, but it is attested only from the 5th century CE onwards. Evidence from linguistics, archaeology and genetics that the Indo-European homeland is to be located in the steppe north of the Black Sea is increasing. Early Proto-Indo-European can probably be dated to ca. 4500–3500 BCE, and a later phase of Proto-Indo-European, associated with the Yamnaya culture, can be dated to ca. 3500–2500 BCE (Mallory 1989; Anthony 2007; Allentoft et al. 2015; Haak et al. 2015; Damgaard et al. 2018). The relatively long period for Proto-Indo-European must be associated with the successive splits of branches leaving the homeland, the split of Anatolian being probably as early as the 5th millennium BCE and that of Balto-Slavic and Indo-Iranian rather late, in the 3rd millennium BCE (e.g. Anthony 2013). However, the details of the internal chronology of Proto-Indo-European and the successive splits and spreads of the separate branches are still to be settled. In the case of Tocharian, too, it is unclear how exactly it came to the northern Tarim Basin in present-day Northwest China.

² Systematic, but quite preliminary surveys are those by Krause (1951) and Bednarczuk (2015). A further noticeable contribution is Schulze (1927). Other references will be given below.

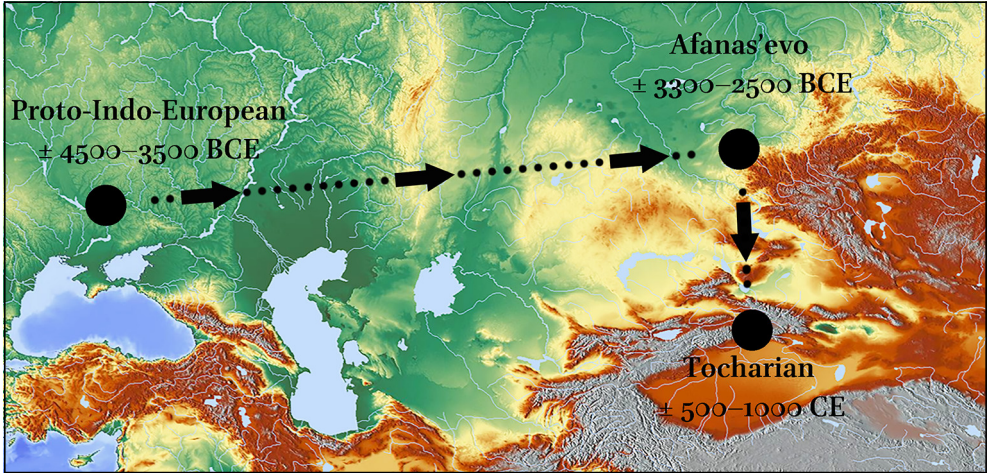


FIGURE 1 The “Tocharian Migration Hypothesis”
SCHEMATIC; BASED ON A MAP FROM MAPS-FOR-FREE.COM

The most coherent scenario holds that the Afanas'ev Culture in the Altai region, dating to ca. 3300–2500 BCE,³ represents an early stage in Tocharian prehistory. Archaeologically and genetically, the Afanas'ev Culture is very close to the late Indo-European Yamnaya Culture further west. From the Altai, Afanas'ev groups would then have to have moved south into the Tarim Basin. It has been suggested, most prominently by Mallory & Mair (2000), that they are there perhaps to be identified with the Xiǎohé Horizon, whose oldest sites and so-called Tarim Mummies date to the 19th century BCE. We may call this scenario the “Tocharian Migration Hypothesis.”

Many leading scholars are of the opinion that the most likely linguistic identification of the Afanas'ev Culture is early Tocharian, e.g. Mallory (1989) and Anthony (2007, 2013). However, especially the second part of the Tocharian Migration Hypothesis, the early southward movement (as assumed by Mallory & Mair 2000), is still full of uncertainties. Obviously, as long as no solid connection can be made from the Afanas'ev Culture to the attested Tocharian languages, we have to remain very cautious.

Most importantly, it is conceivable that the Afanas'ev Culture was indeed an extension of Indo-European culture, while these people are not the ancestors of the Tocharians. Instead they may have spoken an Indo-European dialect that became extinct without leaving any traces (for a more balanced account,

3 The datings in Svyatko et al. (2017) range from the 33rd to the 26th centuries BCE.

see Mallory 2015; see further Kroonen et al. 2018 and Peyrot 2017a). If the Afanas'ev Culture is not to be identified with early speakers of Tocharian, then obviously alternative scenarios are needed, though none is currently more widely supported. The most likely alternative would be that early Tocharians had not yet reached the Tarim Basin when Iranian spread over the Central Asian steppe, and, when the Iranians extended further and further east, they encountered the early Tocharians, who either went with them or were forced to move even further east, ending up in the Tarim Basin.

In my view, the typological traits that set Tocharian apart from Proto-Indo-European can be linked to South Siberia, and in particular to the region of the Afanas'ev Culture, the northern Altai and the Minusinsk Basin. This has no direct bearing on the earliest arrival of early speakers of Tocharian in the Tarim Basin, and thus it has nothing to say about the possible linguistic identity of the oldest Tarim Mummies. However, it would provide the necessary linguistic link between the Afanas'ev Culture and the Tocharian language.

1.3 *Possible prehistoric neighbours of Tocharian*

In the following, I will consider the following languages and language families as potentially relevant for early Tocharian prehistory:

- Turkic. Originally from the Mongolian steppe, Turkic extended at least as far west as the Altai region around the beginning of the Common Era in view of contacts with Proto-Samoyedic (Janhunen 1996; Schönig 2003). Stages of Turkic before this time cannot be reconstructed on the basis of comparative evidence.
- Proto-Samoyedic. This proto-language was spoken around, probably just before, the beginning of the Common Era in South Siberia (Janhunen 1998: 457). Its prehistory is reconstructible through comparison with Finno-Ugric (see also under “Proto-Uralic” below), but the date and location of prehistoric stages are difficult to establish.
- Proto-Uralic. The date and place of Proto-Uralic are debated. A widely held view is that the primary split of Proto-Uralic was into a Finno-Ugric branch on the one hand and a Samoyedic branch on the other (e.g. Janhunen 1981; Sammallahti 1988). This model is adopted here, but an alternative model has been proposed by Häkkinen (2009; see also below, 4.1). If the traditional model with a Finno-Ugric branch is accepted, the Finno-Ugric proto-language must in view of loanwords from Proto-Indo-Iranian (ca. 2200–1800) have been spoken ca. 2500–2000 CE in the southern Ural region, and Proto-Uralic must be dated earlier. The location of Proto-Uralic is hotly debated; I side with those scholars who argue for a homeland east of the Urals (see also below, 4.1).

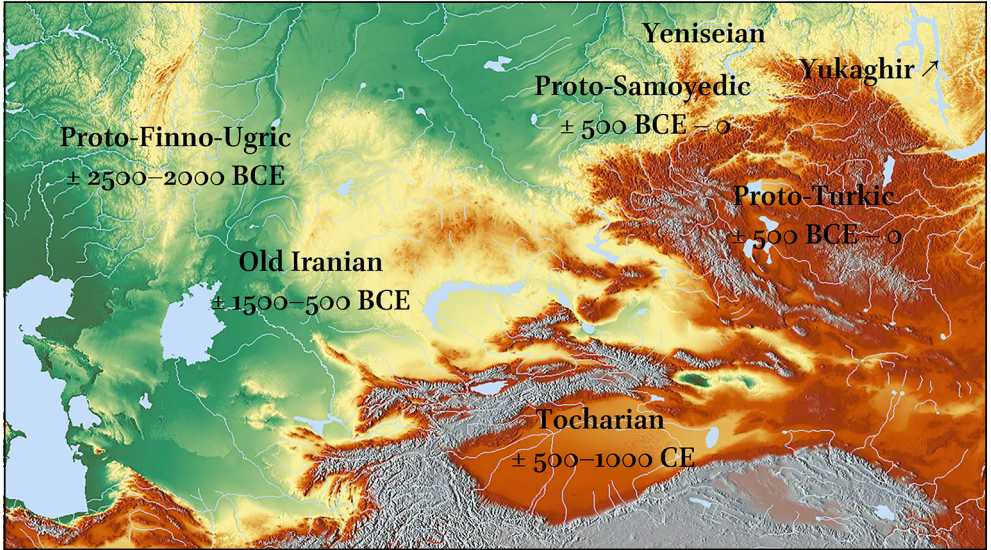


FIGURE 2 Possible prehistoric neighbours of Tocharian
BASED ON A MAP FROM MAPS-FOR-FREE.COM

- Yeniseian. The family was widespread in South and West Siberia, but no secure dates are available (cf. Vajda 2019). In my view, it is likely that Yeniseian predates all other relevant languages in the area.
- Yukaghir. The two closely related, severely endangered varieties of Yukaghir are spoken in Northeast Siberia and no significant prehistory is known. Yukaghir may come from the south in view of parallels with Samoyedic (Aikio 2014a), and might represent an older layer in Siberia than Samoyedic.
- Iranian. Several varieties of Iranian have exerted strong influence on Tocharian. However, most influence concerns loanwords, not structural changes. The earliest presence of Iranians in South Siberia is probably fairly early, around 1500 BCE, but nevertheless later than Afanas'ev. Where contacts between Old Iranian and Tocharian have taken place is unknown.

2 Parallels to the deviant typology of Tocharian

In the following, I consider a number of possible parallels of mostly Uralic, in particular Samoyedic, and Yeniseian to the typological traits of Tocharian that set it apart from Proto-Indo-European. For an evaluation of the value of the different parallels, and a discussion of the consequences for conclusions about the type of language contact that may be supposed, I refer to section 3.

2.1 The stop system

The loss in Tocharian of the Proto-Indo-European obstruent distinctions conventionally noted as voice and aspiration is a very strong indication of foreign influence. Since Proto-Indo-European roots mostly have at least one stop, and often two, the merger of all three stop series into one must have led to massive homonymy and subsequently to heavy restructuring of the lexicon. It is difficult to see how these changes could be motivated language-internally.

TABLE 1 Typological comparison of PIE and PToch. obstruent systems^a

Proto-Indo-European			Proto-Tocharian
*k	*g	*g ^h	*k
*k ^w	*g ^w	*g ^{wh}	*k ^w
*k̑	*g̑	*g̑ ^h	*ts
*t	*d	*d ^h	*t
*p		*b ^h	*p
*s			*s

- a The stops do not always correspond one to one. For instance, PIE *k̑ > PToch. *k, while some PIE *d > PToch. *ts.

It is this innovative typological feature of Tocharian that is the strongest indication of Uralic influence (cf. e.g. Bednarczuk 2015:56). A single stop series as found in Tocharian is reconstructed for Proto-Uralic as well as for Proto-Samoyedic, while other possibly relevant languages all show a system with a contrast between voiced and unvoiced stops, i.e. Proto-Yeniseian, Old Iranian and Yukaghir, or, in Proto-Turkic, a contrast between strong and weak obstruents (see also below).

For Proto-Uralic, Janhunen (1982:23) reconstructs the following obstruents: *k, *c, *t, *p, *δ, *δ';⁴ and *ś, *s. With the development of *s to *t, *ś to *s,⁵ *δ

4 Alternatively, these phonemes may be written *d and *d'. I prefer the more traditional *δ, *δ', which sets these sounds clearer apart from the other stops, with which they have little in common. Kortlandt (2019) interprets *δ as *ȓ and *δ' as *l̑.

5 In a paper given at the *Seminar po sravnitel'no-istoričeskoj fonetike samodijskix jazykov*, 25–26 May 2018 in Moscow (Institute of Linguistics, Russian Academy of Sciences), Mikhail

to **r* and **δ'* to **j*, the Proto-Samoyedic obstruent system had become: **k*, **c*, **t*, **p*, **s* (a secondary **ś* arose later). The Tocharian obstruent system is much closer to both these reconstructed obstruent systems than to the Proto-Indo-European system that is commonly assumed.⁶

TABLE 2 Typological comparison of PIE, PToch., PU and PSam. obstruent systems

Proto-Indo-European			Proto-Tocharian	Proto-Uralic	Proto-Samoyedic
<i>*k</i>	<i>*g</i>	<i>*g^h</i>	<i>*k</i>	<i>*k</i>	<i>*k</i>
<i>*k^w</i>	<i>*g^w</i>	<i>*g^{wh}</i>	<i>*k^w</i>		
<i>*k̑</i>	<i>*g̑</i>	<i>*g^h</i>	<i>*ts</i>	<i>*c</i>	<i>*c</i>
<i>*t</i>	<i>*d</i>	<i>*d^h</i>	<i>*t</i>	<i>*t</i>	<i>*t</i>
<i>*p</i>		<i>*b^h</i>	<i>*p</i>	<i>*p</i>	<i>*p</i>
				<i>*δ, *δ'</i>	
				<i>*ś</i>	(<i>*ś</i>)
<i>*s</i>			<i>*s</i>	<i>*s</i>	<i>*s</i>

Two problems need to be highlighted. First, for Tocharian we have to set up a labiovelar stop **k^w* that was certainly not there in either Proto-Uralic or Proto-Samoyedic. However, this may not be so much of a mismatch since many PIE labiovelars in fact became a plain velar in Tocharian, and many Tocharian labiovelars can be shown to be secondary (cf. Kim 1999; Hackstein 2017:1325). Nevertheless, a minority of the PIE labiovelars have survived as a labiovelar. Second, it is uncertain whether Tocharian **ts* can be compared with Proto-Uralic and Proto-Samoyedic **c*. According to Sammallahti (1988:482; cf. Janhunen 1982:24), PU **c* was retroflex. Proto-Samoyedic **c* “is preserved only in part of the Selkup dialects, where its quality varies between a dental affricate

Zhivlov (Mixail Živlov) has convincingly argued that there are several traces of the original palatal pronunciation of PSam. **s*, of which I cite here: 1) the palatal reflex of PSam. **ns* in Selkup; 2) the palatal reflex *d'* of PSam. **ns* and **ms* in Tundra Enets; 3) the weak grade *d'* of *s* in Nganasan consonant gradation; and 4) the shift of PSam. **ε* to Nganasan *i* or *ɨ* after *ń* and *s*, which only makes sense if *s* was palatal, like *ń*. In my view, this does not yet mean that there was a contrast between **s* and **ś* in Proto-Samoyedic, since this would mean that the merger of PU **s* and **t* had not yet taken place, for which there is thus far no evidence.

6 Here as above, the Tocharian obstruents are given without their palatalised counterparts. If the comparison made here is correct, this obviously means that the contacts have to be dated to the Pre-Proto-Tocharian period, before palatalisation had run its course. On this, see below.

and a retroflex stop, while in the rest of the Samoyedic idioms it has invariably merged with the dental stop" $*t$ (Janhunen 1998:462). Another problem with Tocharian $*ts$ is that it goes back in part to PIE $*d$. It is also possible, therefore, to compare Tocharian $*ts$ with PU $*\delta$ or $*\delta'$. This would exclude any advanced stage of Pre-Proto-Samoyedic as the source of influence, since there is no trace in Tocharian of the Samoyedic developments of PU $*\delta$ to $*r$ or PU $*\delta'$ to $*j$.

In spite of the difficulties with Tocharian $*k^w$ and $*ts$ and Samoyedic and Proto-Uralic $*c$, the structural resemblance between the Tocharian and Uralic systems is striking.

Finally, it should be noted that possible alternative contact languages in South Siberia offer clearly worse matches. This is the case for Yukaghir, which has a voice contrast, for Proto-Yeniseian, for which such a contrast can be reconstructed (Starostin 1982:145), and for Proto-Turkic, which had an opposition between strong obstruents (unvoiced or aspirated stops) and weak obstruents (voiced and in some cases fricative; Erdal 2004:62).

2.2 *The vowel system*

As I will argue, the development of the Tocharian vowel system can be understood very well in light of a South Siberian vowel system today represented by the Yeniseian language Ket. This South Siberian vowel system is different from both the Proto-Tocharian and the Proto-Uralic and Proto-Samoyedic vowel systems. However, a successful comparison is possible when intermediate phases are taken into account: a Pre-Proto-Tocharian phase between Proto-Indo-European and Proto-Tocharian; and a Pre-Proto-Samoyedic phase between Proto-Uralic and Proto-Samoyedic. For a Pre-Proto-Tocharian phase, a vowel system identical to that of Ket can be reconstructed. For Proto-Samoyedic, several different reconstructions of the vowel system have been proposed. Depending on which reconstruction turns out to be correct, a Pre-Proto-Samoyedic vowel system can be reconstructed that is close to the Ket system or perhaps even identical to it.

It will not come as a surprise that the comparison of the vowel systems of two intermediate proto-languages requires several steps of argument. I will first try to show that in the course of its development from Proto-Indo-European the Tocharian vowel system must have gone through a stage that happens to be identical to the system of modern Ket. In order to see whether this Ket system can be compared in a meaningful way, I will investigate whether it can be reconstructed for an earlier period. I will then argue that a very similar or even identical system may be assumed for a certain stage of Pre-Proto-Samoyedic. Finally, Yukaghir will be drawn into the comparison as well.

2.2.1 The development of the Tocharian vowel system

At first sight, the late Proto-Indo-European and Proto-Tocharian vowel systems are not strikingly different:

TABLE 3 Typological comparison of the PIE and PToch. vowel systems

Late Proto-Indo-European	Proto-Tocharian
<i>i</i>	<i>u</i> <i>i</i> <i>ə</i> <i>u</i>
<i>e, ē</i>	<i>o, ō</i> <i>e</i> <i>o</i>
<i>a, ā < *h₂e, *eh₂</i>	<i>a</i>

However, if the developments that led to the rise of the Tocharian system are considered, it becomes clear that Tocharian has undergone heavy changes in the vowel system as well (cf. Peyrot 2013:395). Even though a language-internal development of the vowels is conceivable, external influence, as indicated in any case by the developments in the stop system, discussed above (§ 2.1), would certainly be worth considering in this domain as well.

The basic vowel changes from Proto-Indo-European to Proto-Tocharian are the following (Ringe 1996; Hackstein 2017):⁷

TABLE 4 Main vowel changes from PIE to PToch

PIE	PToch.	PIE	PToch.
<i>*h₂e > *a</i>	<i>> *a</i>	<i>*eh₂ > *ā</i>	<i>> *o</i>
<i>*o</i>	<i>> *e</i>	<i>*ō</i>	<i>> *a</i>
<i>*e</i>	<i>> *'ə</i>	<i>*ē</i>	<i>> *'e</i>
<i>*i</i>	<i>> *'ə</i>	<i>*ei</i>	<i>> *'i</i>
<i>*u</i>	<i>> *ə</i>	<i>*eu</i>	<i>> *'u</i>

7 In the table, ' denotes palatalisation, e.g. *'ə = shwa with preceding palatalisation. The Tocharian reflex of PIE *ō is difficult to establish. Word-finally, it turns into *u in certain contexts (cf. also Kim 2018:101–102), and the number of examples showing the development to *a is limited. The outcome of *ō has no bearing on the argument made here.

To understand how these vowel shifts are connected, the most important development is the merger of PIE **i*, **e*, **u* into PToch. **ə*. As a consequence of these changes, **o* was probably shifted to become a more central vowel, here provisionally written “*ë*.”⁸ The restructuring of the short vowel system thus likely proceeded according to the following steps (cf. also Meier & Peyrot 2017:18–19):

TABLE 5 Shifts in the Pre-Proto-Tocharian short vowel system

Pre-Proto-Toch.1		Pre-Proto-Toch.2		Pre-Proto-Toch.3		Pre-Proto-Toch.4	
<i>i</i>	<i>u</i>	<i>i</i> > <i>ə</i>	<i>ə</i> < <i>u</i>	<i>ə</i>		<i>ə</i>	
<i>e</i>	<i>o</i>	<i>e</i> > <i>ə</i>	<i>o</i>		<i>o</i>	<i>ë</i> < <i>o</i>	
	<i>a</i>		<i>a</i>	<i>a</i>		<i>a</i>	

This short vowel system with only central vowels was then subsequently enlarged with vowels resulting from the shortening of long vowels and the monophthongisation of diphthongs. Finally, old short **o*, which had probably become a central vowel, “*ë*,” in Pre-Proto-Tocharian 4, merged with short *e* from old long **ē*:

TABLE 6 Merger of the Pre-Proto-Tocharian long and short vowel systems

Pre-Proto-Toch.5			Proto-Tocharian		
<i>ei</i> > <i>i</i>	<i>ə</i>	<i>u</i> < <i>eu</i>	<i>i</i>	<i>ə</i>	<i>u</i>
<i>ē</i> > <i>e</i>	<i>ë</i> < <i>o</i>	<i>o</i> < <i>ā</i>	<i>e</i> (< <i>*ē</i> , <i>*o</i>)		<i>o</i>
	<i>a</i>			<i>a</i>	

This reconstruction of the Proto-Tocharian vowel system represents a minimal set of vowels that is widely agreed upon (e.g. Jasanoff 1978:33).⁹

An additional closed **ē* is posited by Ringe (1996:80–86; cf. Hackstein 2017:1315) for the correspondence between word-final Toch.B *-i* and Toch.A *-e*. There can be no doubt that this correspondence reflects PIE **-oi*, as argued by Ringe.

⁸ This notation is taken over from Ringe (1996).

⁹ Several authors use different symbols for the same reconstruction: **e* is also found written “*æ*” or “*ā*”; **o* is often noted with “*ā*”; and **a* with “*ā*.”

However, in Proto-Tocharian this probably still was a diphthong *-ey, with regular monophthongisation to -e in Toch.A and a special development in word-final position to -i in Toch.B. According to Ringe, the monophthongisation of *-ey must be of Proto-Tocharian date because this ending palatalises. This is not correct: palatalising -i in Toch.B matches -i in Toch.A, not -e, and thus reflects PIE *-eies (e.g. Toch.A *kärkălyi* ‘ponds’), or palatalisation is found in many forms of the paradigm according to the distribution of initial palatalisation in the demonstratives (e.g. Toch.B *trici* ~ Toch.A *trice*, nom.pl.m. of ‘third’).

Likewise, Ringe (1996:98–99; cf. Hackstein 2017:1321) reconstructs an additional closed * φ for Toch.B *o* ~ Toch.A *o* correspondences due to *u*-umlaut of **e*. As it is not economical to assume that *u*-umlaut occurred independently in both Tocharian languages, it seems indeed likely that the vowel resulting from this umlaut is to be added to the Proto-Tocharian vowel system. Nevertheless, the final -*u* that caused umlaut was still kept in loanwords from Old Iranian such as Toch.B *tsain* ‘arrow’, borrowed from **d^zainu*–: the plural *tsainwa* < **tsainu*-*a* shows that at the time of borrowing the singular still was **tsainu*, and the -*u* was apocopated later. Therefore, if an additional * φ is to be posited for Proto-Tocharian, this phoneme arose only at a late stage, and it is not relevant for the present discussion.

2.2.2 The Ket and Proto-Yeniseian vowel systems

It is the seven-vowel system of Pre-Proto-Tocharian stage 5 above that is structurally identical to the South Siberian system represented by Ket (see table 7, next page). According to Vajda (2004:5), Ket *i* and *ə* are further back than IPA central [i] and [ə], but not as far back as the unrounded back vowels [u] and [ɤ] of IPA. The allophonic variation in the mid vowels *e*, *ə*, *o* is correlated with tone: they are pronounced as high-mid [e, ə, o] with high-even tone, and as low-mid [ɛ, ʌ, ɔ] elsewhere (Vadja l.c.).¹⁰

Obviously, this parallel with Ket can only be meaningful for Tocharian linguistic prehistory if the same vowel system can be reconstructed for earlier stages. Indeed, Vajda assumes an original Pre-Proto-Yeniseian five-vowel system with *i*, *a*, *ʌ*, *o*, *u* that was in Common Yeniseian enlarged with **e* and **i* (2010:78–79).

10 No exact phonetic values for Pre-Proto-Tocharian can be given, but it is likely that *ə* was a central vowel because it goes back to both **i* and **u*. It is impossible to say what the exact value of **ē* was. The Ket vowels *i* and *ə*, as noted, are not central, but rather back.

TABLE 7 Typological comparison between Pre-Proto-Tocharian and Ket vowel systems

Pre-Proto-Tocharian			Ket		
<i>i</i>	<i>ə</i>	<i>u</i>	<i>i</i>	<i>ɨ</i>	<i>u</i>
(* <i>ē</i> >) <i>e</i>	ë (< * <i>o</i>)	<i>o</i> (< * <i>ā</i>)	<i>e</i> [e, ε]	ə [ə, ʌ]	<i>o</i> [o, ɔ]
	<i>a</i>			<i>a</i>	

However, Starostin (1982:186–189) reconstructed two additional vowels for Proto-Yeniseian: a low front vowel **ä* and a low back vowel **ɔ*.¹¹ He sets up **ä* for the correspondence between Ket *a* and Kott *e*, and **ɔ* for the correspondence between Ket *o* and Kott *a*. For the latter correspondence, Vajda notes that an original **a* is rounded to Ket *o* adjacent to an original uvular corresponding to Proto-Na-Dené **ɕ*, which had probably become a voiced fricative in Proto-Yeniseian (2010:43).¹² Indeed, among Starostin's etymologies with **ɔ* in his 1995 dictionary the majority have the relevant vowels adjacent to uvulars. Also, especially in the first syllable of polysyllabic words original **o* often passes to Kott *a*, probably under influence of the accent and a following *a*. This is clear from *atax* 'tent', which is borrowed from Khakas *otax* (Castrén 1858:ix; Werner 1997b:36). This development may explain cases such as Ket ³*o:ŋ* ~ Kott *apan* 'healthy' (Starostin 1995:199; Werner 2002:2.49), and it may be an alternative to Vajda's explanation from the adjacent uvular in for instance Kott *pagan* ~ Yugh *бóχον* 'mittens' and Kott *hapar* ~ Ket *qóvat* 'back' (2010:43; Werner 2002:1.146, 2.120¹³).

For Starostin's Proto-Yeniseian **ä*, based on the correspondence Ket *a* ~ Kott *e*, there are a few examples in which Kott *e* may derive from original **a* before

- 11 I thank Edward Vajda for answering many questions on Yeniseian in general, and discussing the matter of Proto-Yeniseian **ä* and **ɔ* with me in particular. In addition to the explanations for the relevant correspondences in his published work, he has made several suggestions for individual etymologies to me. Though in this way the evidence in favour of **ä* and **ɔ* has been reduced, it has not yet been eliminated completely. Some of the suggestions that follow are in line with his ideas, but not all, and it is me who is to blame in case they will turn out to be wrong.
- 12 He extended this rule to correspondences with Proto-Na-Dené **g*^w (2010:81, 86) for Ket *ko'd* 'rump' ~ Kott *kar* 'vagina', but has recently rejected this etymology, and now reconstructs Kott *kar* with *k-* from **tl-* (2018:291).
- 13 Several notational systems for Ket and the other Yeniseian languages are in use. In order to maintain consistency, I cite forms after Werner (2002).

i, as in *aršei*, gen. of *arša* 'knee' (Werner 1997b:29). This may be the explanation for Kott *e* in Ket ²*haʔ* ~ Kott *fei* 'cedar' (Werner 2002:1.310), Ket ²*qaʔt* ~ Kott *hei*, *hêi* 'upper clothes' (Werner 2002:2.79) and Ket ²*qaʔj* ~ Kott *xei*, *qei* 'mountain' (Werner 2002:2.78–79). In a fair number of instances of the Ket *a* ~ Kott *e* correspondence, Kott has *a* in the plural, for instance Kott *xe:p* 'boat', pl. *xapaŋ*, *xem* 'arrow', pl. *xamaŋ* (Werner 1997b:33). The *e* of the singular must be original here, with a change to *a* in the plural. Possibly, the same or a similar assimilation operated in Ket to produce *a* corresponding to Kott *e*. Note, for instance, that Ket *lam-* 'flat', *lam-* 'small', which Vajda (2010:91) connects with ¹*e·m* 'flat' and ¹*i·m* 'small' (Werner 2002:1.272, 1.393; both with loss of **t* before a front vowel), could show secondary *a* in a compounded variant.¹⁴ This may, with apocope in Ket, account for Ket ¹*qa·k* ~ Kott *xe:gä*, *qe:gä* 'five' (Werner 2002:2.80). In other cases, the vocalism of Ket is the result of contraction, so that there seems to be no need for **ä* at all, e.g. Ket ³*ta:l'* ~ Kott *t^hégär*, *t^hé²är* 'otter' (Werner 2002:2.251; Starostin 1995:283). Finally, it must be noted that uvulars are also frequent in Starostin's etymologies with **ä*, though it is unclear whether a sound change like **qe* > Ket *qa* is warranted in view of Vajda's rule that uvulars shift to velars before front vowels (2010:88).

In order to definitely reduce Starostin's Proto-Yeniseian nine-vowel system with the additional low vowels **ä* and **ɔ* to the seven-vowel system of Ket, the relevant correspondences should be explained systematically. This is not possible here, but clearly some of the reconstructions with **ä* and **ɔ* may receive an alternative explanation. It remains to be seen whether this is possible for all relevant lexical items. Although both Ket and Kott display a bewildering array of alternations in nominal plural formation, there is no reason to think that no regularisation has taken place at all, and this seems to me an important issue to investigate further.

2.2.3 A Pre-Proto-Samoyedic vowel system

In spite of the problems involving the details of the reconstruction of the Proto-Yeniseian system, the similarity to the Pre-Proto-Tocharian system reconstructed above is obvious. The case of Samoyedic is quite different. A first inspection of the Proto-Uralic and Proto-Samoyedic vowel systems does not yield any striking resemblances. For instance, both Proto-Uralic and Proto-Samoyedic had front rounded vowels, which are absent from Proto-Indo-European and Tocharian, and do not have to be assumed for any intermedi-

14 Compare also, again with different conditioning, Ket *béjas'* 'wind' from ¹*be:j* ¹*e·s'* 'wind god' (Werner 2002:1.120).

ate stage. The exact reconstruction of the Proto-Samoyedic vowel system is debated. I will come back to this below and give here first the reconstruction of Janhunen (1977:9) and Sammallahti (1988:485; for an additional weak vowel **a*, see below):

TABLE 8 The Proto-Uralic^a and Proto-Samoyedic (Janhunen 1977) vowel systems

Proto-Uralic				Proto-Samoyedic			
<i>i</i>	<i>ü</i>	<i>ĩ (= ĩ)</i>	<i>u</i>	<i>i</i>	<i>ü</i>	<i>ĩ (= ĩ)</i>	<i>u</i>
<i>e</i>			<i>o</i>	<i>e</i>	<i>ö</i>	<i>ę (= ě)</i>	<i>o</i>
<i>ä</i>		<i>a</i>		<i>ä</i>		<i>a</i>	

a Häkkinen (2009) reconstructs PU **ę* instead of **ĩ*. This alternative reconstruction has no consequences for the structural points addressed here and below.

As with the Proto-Indo-European and Proto-Tocharian systems, the similarity between Proto-Uralic and Proto-Samoyedic is deceptive. Several shifts have taken place, and in an intermediate Pre-Proto-Samoyedic phase the vowel system must have looked quite different.

First of all, **ö* was still exceedingly rare at the latest Proto-Samoyedic stage just before it dissolved (Mikola 1988:222). It is put in brackets by Sammallahti (1988:485) and must have entered the language at a very late stage.

The other front rounded vowel, **ü*, was more frequent, and it is clear that it must be reconstructed for Proto-Samoyedic. However, Proto-Samoyedic **ü* does not correspond to Proto-Uralic **ü*; rather, Proto-Uralic **ü* was systematically changed to Proto-Samoyedic **i*. Apparently this change was subject to no contextual restrictions: according to Sammallahti, it occurred “in all cases” (1988:484; Janhunen 1981:247). If all original **ü* changed to **i*, all **ü* must be secondary, according to Sammallahti (l.c.), “through irregular changes or new vocabulary items.” In part, **ü* arose from rounding of **i* after **p* and **w*, and from **äw*: PSam. **wüt* ‘ten’ < PU **witi*; PSam. **pütä* ‘cord’ < PU **piksi*; PSam. **kürä* ‘band, strip’ < PU **käwdi* (Aikio 2006:19–20). Although the exact conditions are not yet clear—in particular, there are counterexamples in which rounding did not occur—it is obvious that secondary rounding took place. There are two Proto-Samoyedic items in which Proto-Uralic **ü* seems to have been irregularly preserved (Janhunen 1981:254–255): PSam. **küntä* ‘smoke’ < PU **künti*; PSam. **sünsä* ‘breast’ < PU **šünši/ä*. Rather than assuming that unrounding of **ü* to

**i* was blocked here (because of the following tautosyllabic nasal?), one may as well provisionally state that these items present a further context of secondary rounding of **i* to **ü* (Janhunen l.c.).

According to Sammallahti (1988: 484; Janhunen 1981:247), Proto-Samoyedic **ä* also arose secondarily “through irregular changes or new vocabulary items.” Indeed, there are many good examples for a shift of Proto-Uralic **ä* to Proto-Samoyedic **e*, and Janhunen notes that Proto-Samoyedic **ä* occurs mainly in non-Uralic vocabulary (1981:255–256). He cites two irregular cases in which Proto-Samoyedic has **ä* in inherited words: PSam. **äj* ‘mouth’ < PU **aj*; PSam. **wäj* ‘breath’ < PU **waj*. Whatever the exact explanation of PSam. **ä* in these cases, it probably does not continue Proto-Uralic **ä*, but rather **a*, and must be the result of a secondary development.

In the reconstruction of Janhunen (1977; 1981) and Sammallahti (1988), all Proto-Samoyedic **e* thus reflect Proto-Uralic **ä*. In turn, Proto-Uralic **e* had become **i* in Samoyedic. It is this latter development that has been contested by Helimski (2005). Although the matter clearly deserves a more detailed look than is possible here, I will briefly go into this problem further below, basing myself on Janhunen and Sammallahti’s earlier work first.

The last Proto-Samoyedic vowel to be discussed is the weak vowel **ə* (variously transcribed as “*ə*” in Janhunen 1977, “*ø*” in Sammallahti 1988 and “*ø*” in Janhunen 1998). This vowel is frequent in the second syllable, which has a reduced vowel system that is not relevant for our present purpose. It also occurs in the first syllable through a reduction of original **u* (before an **a* in the next syllable, or when **i* in the next syllable was lost, except when the intermediary consonant was **x* or **l*) or original **i* (before tautosyllabic **l*; Sammallahti 1988:484). According to Helimski (1993; Mikola 2004:18–19), traces of the old sources **u* and **i* of **ə* are preserved in Nganasan vowel harmony, so that he reconstructs a back **ə̄* and front **ə̆*. There is no reason to think that the change of **u* and **i* to **ə* (or **ə̄* and **ə̆*) occurred very early in the development of Pre-Proto-Samoyedic; it does not require original Proto-Uralic contrasts not preserved otherwise and may have occurred at a later stage.

Let me briefly summarise the above points. Of the eleven vowels reconstructed for Proto-Samoyedic by Janhunen and Sammallahti, the following arose in the course of Pre-Proto-Samoyedic:

- **ö* is rare and was clearly added at a late stage;
- **ü* arose secondarily, amongst others from PU **i*, while PU **ü* changed to PSam. **i*;
- **ä* arose secondarily, while PU **ä* changed to PSam. **e*;
- **ə* in first syllables, or back **ə̄* and front **ə̆*, arose secondarily from **u* and **i*.

Since these four vowels arose secondarily, the following seven-vowel system

can be assumed for a very early stage of Pre-Proto-Samoyedic. This system is structurally identical to the system of Ket and to that reconstructed for Pre-Proto-Tocharian:¹⁵

TABLE 9 Typological comparison of Pre-Proto-Samoyedic and Pre-Proto-Tocharian vowel systems

Pre-Proto-Samoyedic			Pre-Proto-Tocharian		
<i>i</i>	<i>ĩ (= ĭ)</i>	<i>u</i>	<i>i</i>	<i>ə</i>	<i>u</i>
<i>e</i>	<i>ę (= ě)</i>	<i>o</i>	(* <i>ē</i> >) <i>e</i>	ě (< * <i>o</i>)	<i>o</i> (< * <i>ā</i>)
	<i>a</i>			<i>a</i>	

An important revision of Janhunen's reconstruction of the Proto-Samoyedic vowel system has been proposed by Helinski (2005). He argues that Janhunen's Proto-Samoyedic **i* has a twofold representation in Nganasan: 1) *i*, corresponding to Old Nganasan *ĩ*; and 2) *ĩ*, corresponding to Old Nganasan *e*. The distribution between Modern and Old Nganasan *i* : *ĩ* on the one hand, and *ĩ* : *e* on the other, would correspond to Proto-Uralic **i*, **ü* versus **e*: MoNgan. *i*, ONgan. *i* < PU **i*, **ü* and MoNgan. *ĩ*, ONgan. *e* < PU **e*. Obviously, this would mean that in Proto-Samoyedic **i* < PU **i*, **ü* and **e* < PU **e* had not yet merged, and consequently the Pre-Proto-Samoyedic vowel system given above would be enlarged with a low front vowel **ä* corresponding to Janhunen's **e*:

TABLE 10 Pre-Proto-Samoyedic enlarged with Helinski's **e* < PU **e*

Pre-Proto-Samoyedic		
<i>i</i> < PU * <i>i</i> , * <i>ü</i>	<i>ĩ (= ĭ)</i>	<i>u</i>
<i>e</i> < PU * <i>e</i>	<i>ę (= ě)</i>	<i>o</i>
<i>ä</i> < PU * <i>ä</i> (Janhunen's * <i>e</i>)	<i>a</i>	

15 I note here again that the phonetic value of Pre-Proto-Tocharian **ə* and **ě* cannot be established in any detail. The Pre-Proto-Samoyedic vowels **ĩ* and **ę* are usually classified as back vowels, like their Ket structural counterparts *ĩ* and *ə*.

Helimski's reinterpretation is accepted by Aikio (2006:9–10; cf. also Salmi-
nen 2012), but the number of examples is relatively small and, as with any
theory, there are counterexamples.¹⁶

Helimski is forced to change several reconstructions for Proto-Uralic to make
his distribution work. For instance, PU **ki* 'who' needs to be changed to **ke*
because of MoNgan. *sij̄i*, ONgan. *sele*; and PU **mexi-* 'give, sell' to **mixi-* because
of MoNgan. *mis-*, ONgan. *m̄ji'ema*. While the interrogative may have been sub-
ject to irregular change and is difficult to reconstruct in detail, his revised recon-
struction of 'give, sell' is contradicted by Skolt Saami *miökkä-* < PU **mexi-* vs.
viikkä- 'take' < PU **wixi-* (Aikio 2014a:45).

It is striking that almost all Helimski's examples of MoNgan. *i* ~ ONgan. *i* with
a good Proto-Uralic etymology go back to stems ending in **i*. The only exception
is MoNgan. *d̄imi*, ONgan. *jimi* 'glue' < PSam. **jimä* < PU **ḍ'ümä*. On the other
hand, most of his examples of MoNgan. *ĩ* ~ ONgan. *e* go back to stems ending
in **ä*. The exceptions here are MoNgan. *m̄jn-* 'go' < PSam. **min-* < PU **meni-*;
MoNgan. *h̄ijm-* 'be afraid' < PSam. **p̄ijm-* < PU **pele-*; MoNgan. *b̄i''* 'water' <
PSam. **wit* < PU **weti*. Although I have at present no explanation for the excep-
tions just listed, it is conceivable that at least part of the distribution noted by
Helimski is due to a secondary change of PSam. **i* (< PU **i*, **ü*, **e*) to ONgan. *e*,
MoNgan. *ĩ* before a following low vowel.

2.2.4 Yukaghir

The vowel system of Ket, which has also been reconstructed for Pre-Proto-
Tocharian, and which may possibly be reconstructed for Pre-Proto-Samoyedic
as well, has a further parallel in Siberia: it is very close to that reconstructed for
Proto-Yukaghir by Nikolaeva (2006:57):

TABLE 11 Typological comparison of Pre-Proto-Tocharian, Ket and Yukaghir vowel systems

Pre-Proto-Tocharian			Ket			Proto-Yukaghir		
<i>i</i>	<i>ə</i>	<i>u</i>	<i>i</i>	<i>i</i>	<i>u</i>	<i>i</i>	<i>y</i> (= <i>ī</i>)	<i>u</i>
(* <i>ē</i> >)	<i>e</i>	<i>o</i> (< * <i>ā</i>)	<i>e</i> [e, ε]	<i>ə</i> [ə, ʌ]	<i>o</i> [o, ɔ]	<i>e</i>	<i>ö</i>	<i>o</i>
	<i>a</i>			<i>a</i>			<i>a</i>	

16 In a paper given at the *Seminar po sravnitel'no-istoričeskoj fonetike samodijskix jazykov*, 25–26 May 2018 in Moscow (Institute of Linguistics, Russian Academy of Sciences), Juha Janhunen has discussed problems in the reconstruction of the Proto-Samoyedic vowel system, including the theory of Helimski. I thank him here once again for sharing his Power-

Yukaghir does not fit the Ket system as well as the one reconstructed for Pre-Proto-Tocharian does. Most importantly, Nikolaeva suspects that **u* was originally a front rounded vowel **ü*, because it normally behaves as a front vowel in vowel harmony. In addition, we would have to see in **ö*, which also behaves as a front vowel, the equivalent of the back unrounded mid vowel **e* of Proto-Samoyedic, *a* of Ket, and centralised **ë* < **o* of Pre-Proto-Tocharian.

The phonetic characterisation of this vowel as front rounded mid *ö* (IPA ø , Cyrillic ø) is peculiar in view of the lack of a front rounded high vowel *ü*. According to Krejnovič (1968:435; cf. Krejnovič 1958:9), Tundra Yukaghir *ö* is slightly retracted and labialised. Odé has analysed the position of Tundra Yukaghir *ö* in the vowel triangle and concludes that it is “a mid central rounded vowel with variable realizations that can be more near-front and near-back” (2012:42).¹⁷ It is attractive to think that the imbalances of the Yukaghir vowel system and vowel harmony reflect the adaptation of an original system with front rounded **ü* and **ö* to a system very similar to that seen in Yeniseian, Pre-Proto-Samoyedic and Pre-Proto-Tocharian.

2.2.5 Conclusion

To sum up, the development of the Tocharian vowel system can be understood very well in light of the South Siberian system represented by Ket. Although theoretically this could be due to influence from Uralic, Yeniseian or even Yukaghir, contacts with an early stage of Samoyedic seem the most likely in view of the evidence of the stops and other evidence still to follow. In the vowel system there are no parallels between Tocharian on the one hand and Turkic or Iranian on the other.

Further research on the historical development of the Yeniseian and Samoyedic vowel systems may show whether the correspondence with Pre-Proto-Tocharian was exact, or whether the three language groups were only partially adapted to each other on this point. The same is true, to a lesser degree, of Yukaghir.

It must be noted that in language contact situations typological features of genetically unrelated languages may converge without becoming identical. A

Point presentation and discussing the problem of Helimski's “thirteenth vowel” with me. He lists more counterexamples to Helimski's distribution, notably PSam. **timä* ‘tooth’ (Ngan. *čimi*), related to PU **sewi* ‘eat’, without giving, as yet, a final solution.

17 Her investigation was not focused on roundedness. She has been, however, so kind as to send me audiofiles of a female and a male speaker of the words in her appendix on p. 42. As far as I can judge, all instances of *ö* in these recordings are rounded, the least rounded being the third *ö* of *örkōbö* ‘lynx’ by the female speaker, and *mōjér lačil* ‘lightning’ and *mōrd'ë* ‘message, rumour’ by the male speaker.

well known example is the famous Balkan *Sprachbund*. Rumanian and standard Bulgarian have similar vowel systems, yet Rumanian has two central vowels, *ă* [ə] and *â* [i̯], in addition to the basic five *a*, *e*, *i*, *o*, *u*, while standard Bulgarian has just *ǎ* [ə] (Schaller 1975:124–133). The Rumanian system is structurally similar to standard Albanian, which has the standard five vowels plus *ë* [ə] and *y* [y], though, obviously, Rumanian *â* [i̯] and Albanian *y* [y] are phonetically clearly different.

Another point that should be raised is that the seven-vowel system reconstructed for Pre-Proto-Tocharian requires the merger of PIE **i*, **e*, **u* into **ə*, which suggests that contrastive palatalisation had already developed by this time, even though **o* and **ē* had not yet merged. At the same time, the parallels with the Uralic and Samoyedic stop systems discussed above in § 2.1 suggest that palatalisation had not yet run its course.

2.3 *Agglutinative case marking and case functions*

Although other Indo-European languages also occasionally show agglutinative case markers,¹⁸ one of the most striking typological characteristics of Tocharian are the agglutinative so-called “secondary” cases. It is obvious that for such a major shift in language type substrate influence must be considered as a serious option. Indeed, this has been proposed in the literature, but thus far without much precision. Pedersen hesitantly suggested Turkic as the model (1931:247). Krause (1951) considered Tibetan, Altaic, Dravidian, Caucasian and Finno-Ugric influence in the case system; although he deemed the last three more promising for further research (p. 202), he did not make a definite choice. See further Bednarczuk (2015:58–59) and Schmidt (1990).

With the exception of Old Iranian, all candidate contact languages of Tocharian have agglutinative case inflexion, and in general a comparable set of cases, see table 12, next page (Samoyedic: Mikola 1988:236–237; Janhunen 1998:469; Castrén 1854:108; case names after Nikolaeva 2014; Proto-Uralic: Janhunen 1982:30–31; Yukaghir: Maslova 2003; Turkic: Erdal 2004; Ket: Vajda 2004).¹⁹

The key to identifying the model of the Tocharian case system is to be found in the functions of the cases. On the functional level, the Tocharian case system shows the following non-Indo-European peculiarities: it lacks a dative, whose functions are fulfilled by the genitive; and it has a local case termed “perlative”

18 Famous are, for instance, the Lithuanian illative, allative and adessive (Stang 1966:228–232).

19 Several further cases have not been included in the table: the Tocharian B vocative and causal; the Yukaghir predicative (focus marker of subject and object); the Turkic equative and similitive; the Ket vocative, benefactive, adessive, caritative, and translative.

TABLE 12 Typological comparison of case functions

Toch.	Proto-Sam.	Proto-Uralic	Yukaghir	Turkic	Ket		
nom.	nom.	nom.	nom.	nom.	abs.	"towards"	
obl.	acc.	acc.	acc.	acc.			
gen.	gen.	gen.	gen.	gen.	gen.		
allat.	dat.		dat.	dat.	dat.		
		dat. (local)		directive			
loc.	loc.	loc.	loc.	loc.	loc.		"in"
abl.	abl.	abl.	abl.	abl.	abl.		"from"
perlat.	prolat.		prolat.		prosec.		"over, through"
com.			com.	(com.)	ins.		"together with"
ins. (A)			ins.	ins.			"with"

which denotes movement along, through or over something, as well as a comitative case denoting accompaniment.

The perlativ is the strongest indication of Siberian, and most probably Uralic or Pre-Proto-Samoyedic influence. A similar local case is widely found across Uralic and in Samoyedic, and also in Yukaghir and Ket, but not in Turkic.

Another interesting functional phenomenon is the lack of a dative in Tocharian. Here the best match is offered by Uralic, where nominative, accusative and genitive are generally analysed as being the "grammatical cases," while the remaining cases are the "local cases." Depending on the description, there may or may not be a case called "dative," but this case is primarily local. A number of notes must be made on this point, however:

- Dative and allative are not so easily kept apart functionally, and both functions are expressed by one case in for instance Yukaghir and Ket.
- The typical Tocharian use of the genitive for the indirect object of verbs like 'give' (Meunier 2015) is not mirrored in Uralic.
- There are traces of an older dative-locative case in Tocharian that may show that the reconstructed case gap was not yet there, or not fully there, in the early phase we are concerned with (Peyrot 2012).
- Functional merger of genitive and dative, also with verbs like 'give', is widespread in Xinjiang, and is found in e.g. Khotanese and Gāndhārī.

For the comitative I have so far found no match in Samoyedic. There is a comitative in Nganasan, but this is clearly secondary and still in the process of grammaticalisation (Wagner-Nagy 2018:188–189). In Ket there is no special comitative either. The case that Vajda terms “instrumental” is called “Komitiv” by Werner (1997a:115–116) and “*Comitativ* oder *Instruktiv*” by Castrén (1858:26). This case can be used as an instrumental as well as a comitative, and therefore it is not exactly parallel to the Tocharian comitative, because the latter cannot be used as an instrumental, for which Tocharian A uses the instrumental case and Tocharian B the perlicative. However, Kott does have a comitative that is distinct from the instrumental (Werner 1997b:62). Whether the case is old is a different matter: it seems to be etymologically related to the Ket instrumental, so that Ket may have lost the original instrumental, or Kott may have created a new instrumental that shifted the old instrumental-comitative to become a comitative only.

At present, I have no explanation for the fact that Samoyedic has no parallel to the Tocharian comitative case. Obviously, it is possible that in a very early phase of Pre-Proto-Samoyedic it had a comitative that was later lost, or the Tocharian comitative may be a later creation. However, I can see no evidence for either scenario. The Tocharian A and B comitative suffixes are different: Toch.A *-śśäl* vs. Toch.B *-mpa*. The Tocharian A suffix is probably secondary, because it is clearly related to the Toch.B preposition *śale* ‘with’, which also occurs in both languages as the first member in compounds: Toch.A *śla-* ~ Toch.B *śle-*. Nevertheless, the Tocharian B suffix cannot be analysed internally and is more likely to be old, even though it is impossible to say how old it is exactly.

Tocharian, in spite of its comitative, agrees better with the Samoyedic case system than with the more elaborate sets of e.g. Finnish and Hungarian: there is no inessive : adessive or ablative : elative contrast. The Ket system, too, is more elaborate than the Tocharian set.

Agglutinative case marking is also found in Ossetic, an East Iranian language that descends from a steppe dialect, “Scythian,” that is very close and possibly identical to the Old Iranian language that has influenced Tocharian in the lexicon (Peyrot 2018). However, the reorganised Ossetic case system must be due to influence from one or more Caucasian languages in view of the close functional matches with Georgian (Belyaev 2010).²⁰ The rise of agglutinative case in Tocharian and Ossetic must therefore be a parallel, but not shared development.

20 However, Georgian is probably not itself the source because Georgian is spoken south of the Caucasus range and Ossetic was originally spoken only north of it.

Carling points out the parallelism between the Tocharian and Modern Indo-Aryan case systems, in particular that of Romani (2012), and argues that this parallelism is an argument for language-internal development (2005:49–52). Leaving aside the problem of possible substrate influence in Modern Indo-Aryan (e.g. Emeneau 1956:9), I note that there is no need for languages to have case, let alone an elaborate case system, and that there are plenty of languages with the relevant prerequisites, notably postpositions, that do not have agglutinative case inflexion. I do not deny that agglutinative case could arise through internal development, but if close matches are found in neighbouring languages, contact-induced change is evidently a factor to consider. Indeed, in the comparison above, it is a combination of the principle of agglutinative case marking and the functions of the cases that calls for an explanation based on contact-induced change.

2.4 *Differential object marking*

In Tocharian, the loss of Proto-Indo-European word-final **-s* and **-m* has led to the merger of the nominative and accusative in masculine thematic nouns, a frequent class characterised by an element **o* before the ending. For instance, the word for ‘horse’ had a distinction between nominative and accusative in Proto-Indo-European, but the two cases are homonymous in Tocharian:

TABLE 13 The development of the thematic masculine singular in Tocharian

PIE	‘horse’		Toch.B
nom.sg.	<i>*h₁ekuo-s</i>	>	nom.sg. <i>yakwe</i>
acc.sg.	<i>*h₁ekuo-m</i>	>	obl.sg. <i>yakwe</i>

That this homonymy is the result of a phonological rather than a morphological development is shown by Toch.B *kante* ‘100’ < PIE **(d)kmtóm*.

However, nouns belonging to this inflexional class that denote human beings do have a distinct oblique singular, e.g. nom.sg. *en̄kwe* ‘man’, obl.sg. *en̄kweṃ*. Despite its superficial similarity to PIE **-m*, the special ending *-ṃ* for nouns of this class with the feature [+ HUMAN] must be secondary and derives from **-n-m* > **-nə*, originally the accusative singular of *n*-stem nouns.

Although such nouns are normally analysed as belonging to two different classes, it is historically just one class, of which nouns with human referents had a marked accusative and the others did not. In my view, this is an instance

of differential object marking based on an animacy hierarchy (Comrie 1989: 129–136).

In Uralic, differential object marking is not universal, but nevertheless widespread, and it is commonly accepted to be a feature of the proto-language. The conditions vary quite substantially, and many descriptions struggle with the details (see Wickman 1955 *passim*). The most common type is that the accusative is only marked with definite objects. An additional remarkable rule is that the object is never marked with 2sg. imperatives. These rules are often assumed for Proto-Uralic as well (Wickman 1955:146; Janhunen 1982:30–31). Castrén claimed that in Zyrian the accusative is used only of living beings (1844:18), but this observation has not been confirmed by subsequent scholarship (Wickman 1955:60).

The Uralic type of marking only definite objects with the accusative is also found in Turkic. Since the conditioning in Tocharian is quite different, this typological comparison is in my view quite weak, and in this case a language-internal motivation seems more likely than contact-induced change.

2.5 *Nominal dual*

Tocharian has a number of nominal dual endings: Toch.B *-i*, *-ə* (= palatalisation), *-e*, *-ne* (Winter 1962; Kim 2018). There cannot be the slightest doubt that, as a category, the dual is inherited from Proto-Tocharian. Nevertheless, it is striking that one of the endings is clearly secondary: the agglutinative dual suffix Toch.B *-ne*, Toch.A *-ŋ*, *-äŋ*. According to Pronk (2015), the element *-n-* of this suffix is extracted from the *n*-stems, while the *-e* may go back to a reflex of **duo* ‘2’ (he reconstructs **duHo*). Kim (2018), who also discusses other explanations in depth, opts for an explanation that derives *-ne* from a postposed pronominal element **ene*. Yet another explanation takes the suffix to have developed from inflexional elements only, without suffixation of numeral or pronominal elements (see the discussion in Kim 2018:90–91).

As it happens, a dual is reconstructed for Proto-Uralic (Janhunen 1982:29–30), and it has been preserved in Samoyedic.

Although there is no need to attribute the existence of a nominal dual in Tocharian to contact, it is conceivable that the creation of an agglutinative dual suffix was externally motivated, at least in part. The other relevant Siberian language groups, Turkic, Ket and Yukaghir, have no nominal dual.²¹

21 Werner (1997a:100) notes that Ket nouns like *ʼde-sʼ* ‘eye’ can have two plurals, one of which is used to denote a pair, in this case *ʼdesʼ*, and the other to denote a larger number, in this case *desʼtaŋ*. This is rare, and there is no established category of dual number.

However, this comparison remains weak, in my view. Since the dual has three other endings in the Tocharian noun, the dual was well-rooted in Tocharian morphology. In other domains of nominal inflexion too, agglutinative traits arose through language-internal developments. Compare notably the agglutinative plurals, e.g. Tocharian B *palsko* ‘thought’, pl. *pälskonta*, where the plural can be segmented as *pälsko-nta* [thought-PL]. In this case, there is no doubt that these plural suffixes arose through language-internal development: they became reanalysed as plural markers when the same suffix was lost in the singular. The existence of plural suffixes may have supported the creation of the dual suffix, but, in my view, it is also still an option that the dual suffix itself arose through similar resegmentation as in the plural. This would make externally motivated change extremely unlikely.

2.6 Comparison

Unlike most other Indo-European languages, Tocharian does not have synthetic expressions for degrees of comparison (Thomas 1958; Bednarczuk 2015: 60). In this respect, Tocharian is like, for instance, Samoyedic and Ket. However, no single proto-forms for the Indo-European comparative and superlative can be reconstructed, and they are lacking in Anatolian as well, and probably in early Proto-Indo-European too. In Tocharian A, the comparative is syntactically expressed with the standard of comparison in the ablative case. In Tocharian B, the standard of comparison is normally in the perlocative case, e.g.:

ñässa kartse (I:PERL good) ‘better than me’

Neither the Tocharian A nor the Tocharian B syntactic expression has an exact match in Anatolian, where the standard of comparison is in the dative-locative (Hoffner & Melchert 2008:273–275 on Hittite).²² The Tocharian A expression with the ablative does have a parallel in Samoyedic and Ket, where it is also in the ablative case (e.g. Kamass, Joki 1944:135; Tundra Nenets, Nikolaeva 2014:174–175; Ket, Werner 1997a:124).²³

22 It should be noted that even when the synthetic comparative and superlative were created later in (or after) Proto-Indo-European, the standard of comparison might have continued to be marked in the same way. Unlike Hittite, mostly the ablative is used, and the dative or locative is rare (e.g. Delbrück 1888:113, 196 on Vedic; Leumann et al. 1965:107–114 on Latin).

23 In Turkic, a morphological comparative exists. It is formed with the suffix *+rAk* and the standard of comparison takes the case suffix *+dA* (Erdal 2004:150). The suffix *+dA* is a locative, but in older Old Turkic it also functions as the ablative (o.c. 174–175).

It is not clear which of the two expressions found in Tocharian is original. It seems that the Tocharian B use of the perlativ is most likely to be old because it also has an ablativ, and the ablativ is widely found in such constructions, so that the use of the perlativ is clearly more marked. If so, it is not likely that this Tocharian construction can be attributed to language contact, because the parallels are not exact. If the Tocharian A expression with the ablativ is original, the problem is that this construction is so widely found that language contact would be a possibility, but it would be very difficult to prove.

Castrén noted that the prosecutive, the case that functionally corresponds to the Tocharian perlativ, is sometimes used in comparisons in Nenets and Nganasan (1854:188–189). Since the prosecutive is used to express a comparative grade of the adjective, not to mark the standard of comparison as in Tocharian, this is not a typological parallel, e.g. Nenets:

səwa-w^ona (good:PROL) ‘better’

According to Castrén, this use of the prosecutive results from calquing of Russian *po* as in *po bol’she* ‘more’, *po lučše* ‘better’ etc.

2.7 *Object marking on the verb*

Within Indo-European, a striking feature of the Tocharian verb is the option of object marking. Object marking is expressed by pronoun suffixes that are clearly segmentable, and are often treated under the pronominal system (e.g. Sieg, Siegling & Schulze 1931:166–168; Krause & Thomas 1960:162–163), and only rarely under the verbal system (Krause 1952:203–207; Peyrot 2013:32–33). The following arguments can be adduced to argue that these pronoun suffixes express object marking of the verb:

- The pronoun suffixes only occur on the finite verb and cannot occur anywhere else in the clause. A few exceptions are attested in Tocharian A nominal sentences, where they are mostly attached to a gerund (Meunier 2015:107–108; Peyrot 2017b:634).
- The pronoun suffixes form one phonological word with the finite verb, as can be seen from the accent in Tocharian B (Krause 1952:203) and from morphophonological alternations and assimilations in Tocharian A (Sieg, Siegling & Schulze 1931:166, 328–331, 334–335).
- There is little formal resemblance between the pronoun suffixes of the verb and the personal pronouns: the pronoun suffixes form their own independent morphological system. The 1sg. pronoun suffix Tocharian A *-ñi*, Tocharian B *-ñ* is close to the gen.sg. of the 1sg. personal pronoun Tocharian A *ñi*, Tocharian B *ñi*, and the

2sg. pronoun suffix Toch.A *-ci*, Toch.B *-c* is close to the obl.sg. of the 2sg. personal pronoun Toch.A *cu*, Toch.B *ci*. However, the 3sg. pronoun suffix Toch.A *-ṃ* (= *-n*), Toch.B *-ne* has nothing in common with the obl.sg.m. demonstratives Toch.A *cam*, Toch.B *ce_u* ‘him’, etc., and the same is true of the plural pronoun suffix Toch.A *-m*, Toch.B *-me* (one form for all three persons)²⁴ and the 1pl. personal pronoun Toch.A *was*, Toch.B *wes* or the 2pl. personal pronoun Toch.A *yas*, Toch.B *yes*, nor with the obl.pl.m. demonstratives Toch.A *cesäm*, Toch.B *ceṃ*, etc.

TABLE 14 Tocharian pronoun suffixes vs. personal pronouns and demonstratives

Pronoun suffixes		Personal pronouns and demonstratives
1sg. suffix Toch.A <i>-ñi</i> , Toch.B <i>-ñ</i>	CLOSE TO	1sg.gen. pronoun Toch.A <i>ñi</i> , Toch.B <i>ñi</i>
2sg. suffix Toch.A <i>-ci</i> , Toch.B <i>-c</i>	CLOSE TO	2sg.obl.sg. pronoun Toch.A <i>cu</i> , Toch.B <i>ci</i>
3sg. suffix Toch.A <i>-ṃ</i> , Toch.B <i>-ne</i>	NOT CLOSE TO	dem. obl.sg.m. Toch.A <i>cam</i> , Toch.B <i>ce_u</i> ‘him’, etc.
pl. suffix Toch.A <i>-m</i> , Toch.B <i>-me</i>	NOT CLOSE TO	1pl. pron. Toch.A <i>was</i> , Toch.B <i>wes</i> , 2pl. Toch.A <i>yas</i> , Toch.B <i>yes</i> , dem. obl.pl.m. Toch.A <i>cesäm</i> , Toch.B <i>ceṃ</i> ‘them’, etc.

- Finally, a fourth argument that the pronoun suffixes express object marking on the verb is that they may occur together with a coreferential noun (conominal, in the terminology of Haspelmath 2013). This is rare, however (cf. Meunier 2015:139–140).

The Uralic languages are well known for a phenomenon that is often called “subjective” versus “objective” inflexion. The subjective inflexion is used with intransitive verbs and transitive verbs with indefinite objects, while the objective inflexion is used with transitive verbs with definite objects. The phenomenon as such seems to go back to Proto-Uralic, being attested in Mordvin, Ugric and Samoyedic (Comrie 1988:466), but there are many differences between the systems in morphological expression, as well as in structural fea-

24 I note here briefly that in Ket possessive prefixes distinguish person in the singular, but not in the plural (Werner 1997a:117–118). I do not venture to say whether this has any significance, since these nominal prefixes are syntactically very different from the verbal suffixes in Tocharian.

tures of syntactic use and information about the object that is expressed. For instance, in Hungarian in essence only definiteness of the object is expressed, in many Samoyedic languages also number, and in Mordvin number and person (Abondolo 1998:30).

The large number of mismatches between the Uralic languages points to an earlier simpler system that was elaborated independently in different ways. The only feature common to all objective conjugation systems seems to be an element that is confined to the 3sg. of the subject and can be reconstructed as Proto-Uralic **sa* / **sä*, originally a 3sg. personal pronoun. This pronoun is reflected as North Saami, Mordvin *son*, Fi. *hän*, Khanty *teγ^w*, Mansi *taw*, Hu. *ő*, and perhaps as Selkup *tep₂* (Abondolo 1998:25, 29–30).

Even though there is in Tocharian no connection between the pronoun suffix and definiteness, as in Uralic, it is in my view possible that the integration of pronominal elements, which are themselves inherited from Proto-Indo-European, into the verbal complex is due to influence from Uralic (cf. also Bednarczuk 2015:61–62). However, in order to see this parallel between Tocharian and Uralic in the first place, one needs to realise that the Tocharian pronoun suffixes are object markers of the verb, and that this constitutes a marked typological contrast with Proto-Indo-European.

2.8 *Converbs*

Tocharian widely makes use of two converbs: the so-called absolute in Toch.B *-rmem*, Toch.A *-räs* denoting anteriority, typically with an unexpressed subject identical to that of the following main clause, and the so-called present participle in Toch.B *-mane*, Toch.A *-mām*, denoting simultaneity. Such converbs are not unheard of in Indo-European languages, and close parallels exist not only in Turkic (Pinault 2015:95–97; Peyrot 2018), but also in Sanskrit. It is striking, though, that the present participle in Toch.B *-mane*, Toch.A *-mām* is to be compared with a verbal adjective in Proto-Indo-European, grammaticalised in many languages as the present participle middle, that must have been inflected. The loss of inflexion is peculiar in Tocharian historical grammar and may point to foreign influence.

Converbs are widespread not only in Turkic languages, but also in Samoyedic (Castrén 1854:372; Nikolaeva 2014 *passim*).²⁵

25 Obviously, a language like Kamass is of little use in this respect, since it is heavily influenced by Turkic itself (Klumpp 2002). Bednarczuk lists this feature as “absolute constructions” (2015:62), claiming that “verbal nouns are widespread in Uralic, Altaic and Paleo-Siberian languages.” Obviously, a verbal noun is not a converb, but can be made into one.

2.9 *Lexical correspondences*

The focus of this paper is on structural matches between Tocharian and Uralic, not on lexical matches. Although lexical matches are a reliable means to determine the source language of contact-induced change, language contact, even if it is profound, does not necessarily entail lexical borrowing. In the case of Tocharian and Uralic, we should not expect to find many borrowings at any event, because if Tocharian took over typical substrate terms from Siberian languages, such as animal and plant names, these were probably lost again after early speakers of Tocharian moved to the completely different ecological surroundings of the Tarim Basin. And if such terms were preserved, they may not be traceable in Tocharian Buddhist literature because this recounts an Indian literary imagery virtually without any connection to the reality of daily life on the Silk Road.²⁶ Borrowing in the opposite direction might be expected to have occurred too, for instance, technical vocabulary related to the wagon or agriculture. In this case, however, if the relevant linguistic varieties survived at all, such terminology must have become obliterated by later innovations brought by for example Iranians, Turks, Tungus or Mongols.

In the literature, very few Tocharian-Samoyedic etymologies have been proposed, and most of these are in my view not convincing at this point (cf. e.g. Napol'skikh 2001; Blažek & Schwarz 2008:57–58). The following selected examples appear to be relatively good to me:

PSam. **sejt*³*wə* 'seven', borrowed from PToch. **s'əptə* 'seven', reflected in Toch.A *špät* (Janhunen 1983:5–6). For this etymology to work, two metatheses have to be assumed: Pre-Proto-Toch. **ʔ* (or **e*, at a very early stage) to **ej*, and **pt* to **tw*. Kallio (2004:132) is critical of this connection. Indeed, the adaptation of **ʔ* or **e* as **ej* is difficult to understand. For the latter metathesis, however, Janhunen (l.c.) adduces a parallel from the Proto-Samoyedic word for 'bed, sleeping place'.

PSam. **wen* 'dog', borrowed from a Pre-Proto-Toch. form of PToch. **kwənə*, i.e. Pre-PToch. **kwənə*, the obl.sg. of **ku* 'dog' (Kallio 2004:133–135). Interestingly, the Tocharian vowel in this word derives from PIE **o*, so that it may have been [ʌ] at the time of borrowing, identical to the **ε* reconstructed for the PSam. word.

26 There are non-Buddhist texts as well, but these are notoriously difficult precisely because of the large number of otherwise unknown content words, such as names of commodities (cf. Ching 2017).

PSam. **menüjâ* (Tundra Nenets *mëñuj*, Tundra Enets *menio*) ‘full moon’ (Helimski 1978:126), borrowed from PToch. **mëñe* ‘moon’ (Blažek apud Napol’skikh 2001:371). In this word, both Tocharian **e* vowels derive from PIE **ē*; this would fit PSam. **e* instead of **ɛ*.

PSam. **wesä* ‘metal’, borrowed into PToch. **wäsa* ‘gold’ (Toch.A *wäs*, Toch.B *yasa*), which reflects an earlier **wesa* (Janhunen 1983:6–7; Driesen 2003:348–350; Kallio 2004:132–133).

Obviously, much more research in this domain is needed. Ideally, this should include the lexicon of individual Samoyedic languages insofar as such items have not been reconstructed for Proto-Samoyedic by Janhunen (1977) because of a limited distribution within Samoyedic. An example of such a word is ‘full moon’ ~ ‘moon’ cited above. Also, one might consider, with due caution, including well established Indo-European vocabulary not surviving into historical Tocharian. However, it would seem better to exclude “Para-Tocharian” material (Napol’skikh 2001), that is, words that do not match well and supposedly derive from a dialect related to Tocharian. Although borrowings of this kind may a priori be expected, such etymologies are unverifiable as long as no coherent set of correspondences in a larger number of words can be established.

Finally, I may note that the relevant phonological stage of Pre-Proto-Samoyedic that would need to be compared is still largely in the dark. On the basis of the correspondences in the vowel system, we may suppose that candidate borrowings took place after the main changes compared to Proto-Uralic, such as **ü* > **i*, but before the rise of secondary **ü*. However, it would be important to know whether the change of PU **ś* and **s* to PSam. **s* and **t* is to be dated before or after possible contacts with Tocharian. If the etymologies for ‘metal’ ~ ‘gold’ and ‘seven’ are correct, they would indicate that the contacts are to be dated after these far-reaching developments.²⁷ Another, less secure correspondence may show that the contacts took place before the change PU **l* > PSam. **j*: PSam. **jäm* ‘sea, big river’ (Janhunen 1977:40), possibly borrowed from PToch. **lämә* ‘lake’ from earlier **lim-* (Toch.B *lyam*; Adams 2013:614). The problem is the vocalism. Toch.A *lyom* ‘marsh, mud’ < PToch. **lem-* would fit better formally, but here the semantics are obviously worse.

27 If Pre-PToch. **s* really corresponds to PSam. **s* here, this would require that the palatal reflexes of PSam. **s* observed by Mikhail Zhivlov (see fn. 5) are to be explained as due to a palatal allophone, and that there was no nonpalatal **s* besides. The correspondences would be difficult to understand if PSam. **s* was palatal and contrasted with a nonpalatal **s*.

2.10 *Lexical typology*

Apart from loanwords, there are possibly also other parallels in the lexicon, for instance in word formation and so-called “nursery words” of the type *mummy* and *daddy*. The evidence on the whole, however, remains weak.

In Tocharian B, the following terms for ‘mummy’ and ‘daddy’ are attested: *ammakki* (voc., the nom. may have been *ammakka**) ‘dear mummy’; *āppa* ‘daddy’ (voc., the nom. may have been *āppo**); *appakke* ‘dear daddy’ (Adams 2013:17, 22, 47).²⁸ In Indo-European, this type is attested, cf. for instance Greek *ἀμμά, ἄππα, ἄπφα*²⁹ (Beekes 2012: 88, 119, 121), but it is rare, especially for ‘daddy’ (Buck 1948:94). On the other hand, fairly close parallels are found in Yeniseian: Ket *ʼa-m* ‘mother’ (voc. *amá, amá* [close by], *amá* [further away]), *ʼo-p* ‘father’ (voc. *obó*; Werner 2002:1.95, 2.50, 1997a:117). For Proto-Samoyedic, Janhunen reconstructs a very similar **emä* ‘mother’ (1977:23; Aikio 2014a:39 spells **ämä*), but **ejsä* ‘father’ (Janhunen 1977:22) is different. Nevertheless, even though there are parallels with Samoyedic and Yeniseian, these are not exact, and external influence in this domain will always be difficult to prove.

An interesting Tocharian term, probably preserving a trace of the world view of the Tocharians before Buddhism, is the Tocharian A word for ‘world’, *ārkišoši*. Etymologically, this is a compound of *ārki* ‘white’ and *šoši* ‘living’, cognate of Tocharian B *śaiṣṣe* ‘world’ (Pedersen 1941:262; Pinault 1994:366). Within Indo-European, there are parallels for ‘white, bright’ as a *Benennungsmotiv* for ‘world’, cf. Slavic words deriving from the etymon of OCS *světo* ‘light’, such as Polish *świat* ‘world’, or Skt. *loka-* ‘open space, world’, which goes back to a root for ‘light’ (cf. Gr. *λευκός* ‘white, clear’ and Skt. *roca-* ‘bright’; Buck 1948:12, 15b), although there is nothing that matches the Tocharian formation in any exact way. Another possible model is formed by very close parallel expressions in Yeniseian, cf. Ket *kúndey* ‘people of this world’, from *²kaʼn ²deʼŋ* ‘bright people’ and *kúnbaŋ* ‘world’ from *²kaʼn ²baʼŋ* ‘bright earth’ (Werner 2002:1.466; Werner 1997a:49; Werner 1998:50). Etymologically, Toch.A *ārki* ‘white’ derives from a root meaning ‘bright, brilliant’ (Adams 2013:53).³⁰

The Tocharian words for ‘sun’, ‘moon’ and ‘earth’, which are compounded with the word for ‘god’, are often cited as possible relics of a pre-Buddhist pantheon: Toch.B *kaṃ-ñäkte*, Toch.A *koṃ-ñkät* ‘sun’; Toch.B *meñ-ñäkte*, Toch.A

28 For none of these is a Tocharian A cognate attested. There is an obl.pl. *āpas* in A 256 a3, but this seems to mean rather ‘ancestors’.

29 This is an “endearing address between brothers and sisters or beloved ones.”

30 Like Toch.B *śaiṣṣe*, Toch.A *ārkišoši* means ‘world’ as well as ‘people’ (similar to Fr. *monde*), but this is probably due to calquing from Skt. *loka-* (SWTF:4.61–62).

mañkät ‘moon’; Toch.B *keṃ-ñäkte*, Toch.A *tkaṃ-ñkät* ‘earth’. There are in Ket several compounds with ³*ku:s* ‘god, spirit’ and ¹*e-s* ‘god, sky’, like *báŋgu-s* ‘earth spirit’ from ²*baʔŋ* ³*ku:s* ‘earth spirit’ (Werner 2002:1.105), *qáŋgu-s* ‘mountain spirit, lord of the animal world’ from ²*qaʔj* ³*ku:s* ‘mountain spirit’ (Werner 2002:2.63), or *béjas* ‘wind’ from ¹*be:j* ¹*e-s* ‘wind god’ (Werner 2002:1.120). However, I have found no parallel formation that is specific enough to be a possible model for the Proto-Tocharian “gods.”

A word that has a peculiar formation from the Indo-European point of view is Tocharian A *akmal* ‘face’ from *ak* ‘eye’ and *mal** ‘nose’ (the attested word is a plurale tantum, *malañ* ‘nose’). There are many compounds and binomials in both Tocharian languages, but most binomials combine two words with a similar meaning to form an expression with the same meaning. The word *akmal* is certainly the most striking example of a compound with a basic meaning formed from two elements with a different meaning. Exact parallels are found in Khanty *ńot-sēm* and Mansi *ńol-sam*, both ‘face’ from ‘nose’ and ‘eye’, while similar compounds such as MOUTH NOSE, NOSE MOUTH and MOUTH EYES, all meaning ‘face’, are likewise found in Finno-Ugric (Schulze 1927; Krause 1951:197–198; Aalto 1964:59; Bednarczuk 2015:61). Although compounds of this type are extremely frequent in Yeniseian, I could find no similar formation for ‘face’ there.

Finally, I note a possibly parallel *Benennungsmotiv* in the word for ‘man’ in Tocharian and Samoyedic. The etymology of the Tocharian word, Toch.B *eñkwe*, Toch.A *oñk* is quite clear: as “the mortal one,” it derives from **neku-* ‘dead, corpse’ (Beekes 2010:1003–1004). Possibly, the Proto-Samoyedic word **kaəsa* ‘man’ is derived from **kaə-* ‘die’ as well (Janhunen 1977:61). In this case, however, the metaphor is ready at hand, and we find the same in e.g. Skt. *mártya-* ‘man’ and Av. *mašüa-* ‘man’ (Buck 1948:81).

3 Evaluation and interpretation of the parallels

The parallels to the deviant typological traits of Tocharian that have been discussed in the preceding section are of uneven value.

I consider the evidence from the stop system (§ 2.1), the vowel system (§ 2.2) and the agglutinative case system (§ 2.3) as the strongest indications of language contact. The Tocharian stop system with only voiceless stops is the best evidence for Uralic influence. The vowel system shows neat parallels with Yeniseian and Pre-Proto-Samoyedic. Taken together, this suggests that the Uralic variety with which Tocharian was in contact was a form of Pre-Proto-Samoyedic. Agglutinative case systems are widely found in Siberia and Eastern

Central Asia, but the case functions, in particular the Tocharian perlicative, best match Uralic and comparable systems in South Siberia.

Relatively good matches are further found in object marking on the verb (§ 2.7), matched by Uralic in particular, and the use of converbs (§ 2.8), which is, on the contrary, a widespread feature that can hardly be assigned to a particular contact language. However, these two features cannot be considered proof if they are not combined with the primary arguments from phonology and case inflexion.

No compelling evidence could so far be identified in the domains of differential object marking (§ 2.4), the nominal dual (§ 2.5), comparison of adjectives (§ 2.6) and lexical typology (§ 2.10). There are parallels, but they are not exact enough, or not specific enough to be linked to a particular contact language.

Lexical correspondences (§ 2.9) are strikingly few. Language contact between early Tocharian and early Samoyedic is nevertheless strongly suggested by a few good etymologies in this domain, too. The dominant direction of borrowing, as far as the scanty evidence goes, is from Tocharian into Samoyedic, not the other way around.

The heavy impact in phonology and the scarcity of lexical influence point to substrate influence. In substrate influence, or interference induced by language shift, it is often structural features, in particular phonetics, phonology and syntax, that are carried over from the source language into the target language, and lexical impact need not occur or may remain minimal (e.g. Thomason & Kaufman 1988, in particular pp. 129–146). The reason is, naturally, that speakers of the source language usually attempt to master the target language completely, more successfully avoiding interference in the domains of morphology and lexicon, and less successfully avoiding interference in the domains of phonetics, phonology and syntax (e.g. Van Coetsem 2000).

Indeed, while the strong impact observed in the stop and vowel systems is clearly of a structural nature, the agglutinative case system can be analysed as a structural feature too. The agglutinative case suffixes probably go back to original postpositions, which places this development in the domain of syntax. Also the use of converbs and object marking on the verb belong to the syntactic domain. It appears that all compelling and acceptable cases of contact-induced change belong to the structural domains of phonology and syntax, typical of a substrate situation. This may at the same time explain the scarcity of lexical influence, but a caveat here is clearly due because of the problems noted above (§ 2.9).

A further note on the development of the vowel system under the substrate scenario proposed here is needed. The striking parallels between the vowel systems definitely point to contact, but it is not clear how the adaptation of the

late Proto-Indo-European vowel system to that reconstructed for Pre-Proto-Samoyedic could have led to the changes observed. Pre-Proto-Samoyedic had the vowels **i*, **e*, **u* and there would seem to have been no problem in keeping these as such instead of changing them to **ə*, as in fact happened. In my view, we have to assume that most of the drastic changes in the vowels had already started off before influence from Pre-Proto-Samoyedic took place,³¹ and that these were then under the influence of Samoyedic fixed in the form that I have reconstructed above (§ 2.2).

Finally, I briefly note that the structural impact on Tocharian has been heavy, but, nevertheless, there are many strong typological differences between Tocharian and Uralic. Among the most striking are:

- The negative auxiliary verb typical of Uralic (Janhunen 1982:37) lacks even the slightest trace in Tocharian, which makes use of a “normal” adverbial negation Toch.A and Toch.B *mā* (Tocharian A has a special negation *mar* for commands and directives).
- The limited blurring of the noun : verb distinction in Uralic (Janhunen 1982:38) is not found in Tocharian.
- The widespread suffixation of pronominal possessives to the head noun in Uralic is not found in Tocharian.
- The developed causative, transitive and intransitive system of the Tocharian verb is not mirrored in any exact way in Uralic.
- Uralic has no nominal gender, but Tocharian has a rigid gender system with agreement on demonstratives and adjectives.
- The peculiar 1sg.f. pronoun in Tocharian A, without match in Uralic, is in all probability secondary and cannot be reconstructed for Proto-Tocharian (Jasanoff 1989).

Especially in cases in which the Tocharian state of affairs can be understood in light of its Indo-European origins, as with nominal gender, typological differences need no explanation. It is slightly more complicated when Tocharian is clearly innovative compared to the proto-language. In the list above, this is true of the Tocharian A special negation *mar*; the causative, transitive and intransitive system of the verb; and the Tocharian A 1sg.f. pronoun. Such innovative non-parallelisms need to be accounted for. Either they should result from language-internal change, as probably in the case of the three highlighted items, or they may have been induced by another contact language. In any case, such mismatches do not as such contradict the hypothesis of contact-induced change in Tocharian developed here.

³¹ Whether this was internally or externally motivated is difficult to say at this point.

4 The prehistoric context

The typological parallels between Tocharian and Uralic, and in particular Samoyedic, are strong support for the Tocharian Migration Hypothesis briefly outlined in § 1.2 above. Since the parallels involve in part also Yeniseian, it is likely that these contacts have taken place in Southern Central Siberia. This area is not well defined and potentially very large, but even this approximate location is enough to exclude alternative scenarios in which, for instance, Tocharian came into the Tarim Basin directly from the steppe, or through the Pamirs, or was in contact with Uralic languages in the southern Urals. The Tocharian Migration Hypothesis, with the Indo-European Afanas'ëvo Culture as an intermediate station, was formulated purely on the basis of first archaeological, and then also genetic evidence. The prehistoric South Siberian phase of Tocharian outlined here adds the so far completely missing linguistic argument.

4.1 *Time and place of contact*

The exact location of the contacts of Pre-Proto-Tocharian in Southern Central Siberia is difficult to establish, and it is quite likely that the area was large, or shifted through time, so that there is no “exact location” in the strict sense of the word. However, it seems that the relevant proto-languages were close enough geographically to satisfy the requirement in § 1.1 that there should be a possible historical scenario for the contacts. The location of Proto-Samoyedic can be inferred from the distribution of the historically known languages, which, with the now extinct Kamass and Mator, extended as far south as the Sayan Mountains. Also, there are early Turkic loans into Proto-Samoyedic (Janhunen 1998:477), which suggests a homeland relatively far to the south (cf. also Helimski 2004:120). The large area covered by the Afanas'ëvo finds satisfies these basic requirements easily. The case of Yeniseian is a little different, because Ket is spoken further to the north.³² However, the related extinct Kott, spoken along the Mana south of Krasnojarsk, is already closer, and on the basis of hydronyms the prehistoric Yeniseian area is known to have extended much further west, south, and southeast (e.g. Vajda 2019; Maloletko 2002:156).

More problematic is the chronology. Proto-Samoyedic is often considered to be approximately 2000 years old; according to Janhunen, it can be dated to “the last centuries BCE” (1998:457). Such a late date is excluded for any contacts

32 From the northernmost sites of the Afanas'ëvo Culture, e.g. Černovaja near Novosëlovo between Abakan and Krasnojarsk (Vadeckaja et al. 2014:333), it is around 750 kms to Southern Ket in Sulomaj at the Mountain Tunguska (Vajda 2004:9). From the isolated Afanas'ëvo site at Gljaden northwest of Novosëlovo, it is around 650 kms.

from the Tocharian side, but this need not be an insuperable obstacle, since the contacts must have taken place, in view of the linguistic evidence presented above, well before Proto-Samoyedic dissolved, at a relatively early Pre-Proto-Samoyedic stage.

The question of the dating of Proto-Uralic and the timeline of Pre-Proto-Samoyedic is closely connected to the structure of the Uralic family tree. With the traditional split into Finno-Ugric and Samoyedic, Proto-Uralic must be older than Proto-Indo-Iranian, since the latter is (at least partly) contemporaneous with Proto-Finno-Ugric. At the same time, the timeline of Pre-Proto-Samoyedic would be very long, stretching from Proto-Uralic up to Proto-Samoyedic, and Pre-Proto-Samoyedic may have been spoken in South Siberia already at an early date. The dating of Proto-Uralic in this traditional model is hotly debated, but even with some of the most recent dates, around 3000 BCE (Janhunen 2009:68), there is no problem in dating Pre-Proto-Samoyedic stages to, say, 1000 BCE or 2000 BCE. The end of the Afanas'ev period, around 2500 BCE, would with this chronology also lie within the long stretch of Pre-Proto-Samoyedic.

Häkkinen's alternative model (2009) of a primary split between West-Central Uralic and East Uralic, the latter comprising Ugric and Samoyedic, has serious consequences for the prehistory of Samoyedic. On the one hand, the timeline of Pre-Proto-Samoyedic becomes shorter, since it starts only after the split of East Uralic into Ugric and Samoyedic, and early Samoyedic could then have arrived in Central Siberia only some time after this split. On the other hand, if Proto-Indo-Iranian loanwords into "Finno-Ugric" are still accepted, these now automatically become Proto-Uralic, i.e. they were borrowed before the split into West-Central and East Uralic. This in turn would lead to much later datings, with Proto-Uralic around 2500–2000 BCE, followed by East Uralic (2000–1500 BCE?) and Pre-Proto-Samoyedic starting only after that. In sum, in combination with the more recent datings, Häkkinen's alternative family tree is difficult to reconcile with the Tocharian Migration Hypothesis in combination with the Pre-Proto-Samoyedic substrate hypothesis developed here: in Häkkinen's framework, Pre-Proto-Samoyedic cannot have been in South Siberia early enough.

Häkkinen's main arguments for his alternative model of the family tree are: 1) the common innovations of Finno-Ugric proposed by Janhunen and Sammallahti could also be archaisms, the innovation having happened rather in Samoyedic; and 2) there are shared innovations between Ugric and Samoyedic. However, the developments proposed by Janhunen (1981) and Sammallahti (1988) cannot simply be reversed; e.g. for PFU **uxi* and PSam. **o* only PU **oxi* can be reconstructed because PFU **uxi* may also correspond to PSam.

u*, pointing to PU **uxi*. In addition, the contrast between PU **j* and **u* is also apparently preserved better in Samoyedic (Peyrot fthc.). For a recent treatment of **x* and vowel sequences, another relevant point criticised by Häkkinen, I refer to Aikio (2012). As common East Uralic innovations Häkkinen adduces, among others, the shifts of **s* to **ʃ* or **ʃ* (L*) and **ś* to **s*, as well as the split of **j* (originally **e*, according to him) into **j* and **e*, noting that the conditions of this split are unknown. Indeed, similar developments have taken place in Ugric and Samoyedic, but they are more likely due to areal features, as suggested by Aikio (2014a:35), and possibly, more specifically, to Yeniseian substrates (see also fn. 37). That the innovations listed by Häkkinen are parallel, not common, is strongly suggested by the fact that for the split of **j* into PSam. **j* and **e* clear conditions have been formulated (e.g. Sammallahti 1988:484).

It seems to me, therefore, that the common innovations of Finno-Ugric, even though they are not many, warrant the assumption of this subbranch, and that the alleged common innovations of Ugric and Samoyedic are rather parallel developments. We may thus date Proto-Uralic before Proto-Indo-Iranian, and Pre-Proto-Samoyedic may have been spoken in South Siberia early enough for it to have influenced Pre-Proto-Tocharian in accordance with the chronology of the Tocharian Migration Hypothesis.

The dating of Yeniseian is difficult because there is very little evidence to go on. Vajda argues that the preservation of Yeniseian hydronyms by later Turkic- and Uralic-speaking populations shows that the historically known Yeniseian languages had already diverged by 2000 years ago (2018:280; 2019). On the other hand, he thinks that the close similarities between these languages suggest that they split less than 4000 years ago. This dating is in line with the glottochronological estimate to the 9th century BCE of Blažek & Schwarz (2017:142–143). I agree completely with his line of argument. In my view, an additional reason for thinking that the Yeniseian languages precede other linguistic groups in the area is that it is unlikely that, as hunter-gatherers, they should have spread over the enormous area covered by Yeniseian hydronyms when populations with more advanced economies were already living there. It is possible that Proto-Yeniseian—in the narrow sense of the proto-language of the historically known Yeniseian languages at the latest stage before the break-up—is to be dated, with Vajda, between 2000 BCE and the beginning of the Common Era, while the hydronyms may go back, in part, to earlier, Pre-Proto-Yeniseian varieties (cf. also Vajda 2019). At the same time, 2000 BCE is not a hard date, and it is certainly also conceivable that the age of the Yeniseian family is still underestimated.

If the Siberian traits of Tocharian arose in the Afanas'ev period, ca. 3300–2500 BCE, this would make Pre-Proto-Samoyedic and Yeniseian (Proto-Yeniseian or Pre-Proto-Yeniseian) older than most datings in the literature. Indeed,

from the point of view of Tocharian, it still seems the best scenario that early speakers of Tocharian moved south after the Afanas'ev period and arrived in the Tarim Basin at a very early point in time. However, the Siberian features of Tocharian discussed here in my view only show that Tocharian was in South Siberia, not that Tocharian speakers left and moved south at an early date. I will in the following assume that the contacts are to be dated to the Afanas'ev period, but I note here explicitly that this is at this point no more than a working hypothesis that is inspired by archaeological, not by linguistic arguments.

4.2 *Relative chronology*

A crucial question in the context of this study is whether anything can be said about the relative chronology of the shared linguistic traits between Tocharian, Samoyedic and Yeniseian.

As argued above (§ 3), the parallels between Tocharian and Samoyedic point to substrate influence of Pre-Proto-Samoyedic on Pre-Proto-Tocharian: Samoyedic groups switched to Tocharian but introduced a large number of structural features from their native language.

For Samoyedic, in turn, the assumption of a Yeniseian substrate would provide a neat mechanism for the most important sound changes that set this branch apart from Finno-Ugric:³³

- The unrounding of **ü* to **i*: Yeniseian had no *ü*.
- The split of **j* into **j* and **ç*: Yeniseian had a high back unrounded vowel *i* (parallel to PSam. **j*) as well as a mid back unrounded vowel *ɨ* (parallel to PSam. **ç*).³⁴
- The change of **δ* to **r*: I tentatively compare this change with the intervocalic allophone *r* of Ket *d* (Vajda 2004:7) and the change of word-final **d* to *r* in Kott (Starostin 1982:148). Note that **δ* apparently did not occur initially in Proto-Uralic (Sammallahti 1988:482), so that the intervocalic and word-final positions cover almost all occurrences.
- The change of word-initial **l* to **j*: Yeniseian does not have regular initial *l*. Starostin does not reconstruct initial **l* for Proto-Yeniseian at all (1982:149). Vajda does reconstruct initial **l*, but notes that it was probably a lateral fricative **ɬ*, which is in Ket pronounced with a stop element word-initially: [ɬ] (2010:91). The phonetic properties of Yeniseian **l* = **ɬ*, **ɬ* may be a rea-

33 I leave out the changes **e > *i* and **ä > *e*, which are disputed for Proto-Samoyedic (see § 2.2.3). The relevant vowels are also disputed for Proto-Yeniseian (see § 2.2.2), so that this remains a task for future research.

34 If the PU phoneme was **ç* instead of **j* per Häkkinen (2009), the mechanism for the split into **ç* and **j* presented here would still hold.

son why Proto-Uralic **l* was replaced in initial position in Samoyedic, but there is also another possibility. In some cases, PU initial **l* is unexpectedly preserved, according to Aikio (2014b:86) before PU **i* ~ PSam. **i*, **e*. This provisional rule—there are only very few examples of either development—is reminiscent of Vajda's rule that initial **t* is lost before front vowels in Ket/Yugh (2010:91–92), although it must be stressed that the conditions are not identical, as far as they are known at all.³⁵

- The change of PU **s* to PSam. **t*, and of PU **ś* to PSam. **s*: These changes are difficult to explain through Yeniseian influence. Starostin reconstructs only one sibilant for Proto-Yeniseian, **s*, whose pronunciation varied between *s* and *ś* / *š* (1982:152). This may explain the change of PU **ś* to PSam. **s*, but it is difficult to see why PU **ś* and **s* have not simply merged in Proto-Samoyedic. The change of PU **s* to PSam. **t* is reminiscent of the change of Proto-Yeniseian **s* to **t* in Pumpokol (Starostin 1982:155; Vajda 2010:82), but in this case it is unclear how PSam **s* from PU **ś* could be preserved as such. Perhaps a way out is to assume that the shift of PU **ś* to PSam. **s* is due to influence from Yeniseian, and that this change triggered the shift of original **s* to **t* in a push-chain development³⁶ (possibly through an intermediary *ʃ* that later merged with **t* < PU **t*, cf. Aikio 2014a:35).³⁷

The assumed parallels between Samoyedic and Yeniseian listed above concern an old layer of contact: if they are correctly identified, they explain changes from Proto-Uralic to Proto-Samoyedic.³⁸ There is no doubt that at a later stage Samoyedic and Yeniseian languages were still, or again, in contact. Examples are not difficult to find: the phonologisation of *k* vs. *q* in Selkup; the loss of (secondary) front rounded vowels in several Samoyedic languages (on Enets, cf. Georg 2008:156–157); parallel changes of *č* to *t*, *w* to *b*; etc. (for this later layer of contact, cf. e.g. Anderson 2003).³⁹

35 Note that Vajda assumes general loss of initial **j* in Yeniseian (2010:75–76).

36 Note, however, that PSam. **s* probably was still palatal or had a palatal allophone, as shown by Mikhail Zhivlov (see fn. 5).

37 Aikio (l.c.) notes that these shifts have parallels in Ugric: “Apparently, the restructuring of the sibilant system through the changes **s* > **θ* and **ś* > **s* is an old areal phenomenon connecting Samoyed and Ugric.” Of course, this is also possible. At the same time, it does not completely exclude Yeniseian influence either. If the Proto-Ugric homeland was south of the Ob-Ugric languages Mansi and Khanty, it was probably quite close to Maloletko's Yeniseian hydronym area number 3 (“Omsko-priirtyšskij,” 2002:156).

38 These observations make it unlikely that Samoyedic precedes Yeniseian in the Minusinsk Basin, as argued by Janhunen (2009:72).

39 Janhunen cautiously suggests that Yeniseian influence may have been an external factor in the rise of phonemic glottal stop in Samoyedic (1986:168), but, as far as I can see, it is difficult to explain the distribution of, for instance, the Nenets glottal stop from the tonal

Although it seems clear that Yeniseian represents an archaic linguistic layer in the area, it would be naive to think that it was not itself influenced by other languages at an early stage. Examples of larger changes possibly due to external influence can be found throughout Vajda 2010. A case in point is the rise of the seven-vowel system discussed in § 2.2.2 above, which according to Vajda (2010:78–79) derives from a five-vowel system with *i*, *a*, *ɛ*, *o*, *u* through phonologisation of allophonic variation of **ɛ* with **e* and **ɛ̄*. The fact that the Proto-Yeniseian seven-vowel system is possibly secondary needs to be stressed, since it seems that the changes leading to the same system in Pre-Proto-Tocharian cannot be explained by influence from Pre-Proto-Samoyedic or Proto-Yeniseian alone (see above). However, the assumption that the Yeniseian vowel system developed under Tocharian influence would lead to a very complicated scenario, since all other evidence rather indicates that Tocharian is influenced by a Samoyedic substrate, and that Samoyedic is influenced by a Yeniseian substrate.

4.3 *Towards a unified prehistoric interpretation*

In a prototypical substrate situation, an incoming language is influenced by a language already spoken in the area. If the above conclusions about the relative chronology of contacts are correct, this suggests that Yeniseian represents the oldest (at this point recoverable) linguistic layer in South Siberia, that Samoyedic came afterwards, and that Tocharian arrived as the last of these three. In terms of population prehistory, this would mean that incoming speakers of Samoyedic mixed with already present speakers of Yeniseian, and that incoming speakers of Tocharian then mixed with already present speakers of Samoyedic and possibly of Yeniseian too.

This scenario is difficult to reconcile with the archaeological and genetic data. According to recent genetic research, “[t]he Early Bronze Age Afanasievo culture in the Altai-Sayan region is genetically indistinguishable from Yamnaya” (Allentoft et al. 2015: 169b). Also from the archaeological point of view, there are close similarities between Afanas’evo and Yamnaya, certainly to be identified with a late phase of Proto-Indo-European. There is no evidence for any heavy influence from a local population or culture. This is at variance with the linguistic substrate scenario sketched above, and rather suggests that the

system of Ket. Likewise, it is difficult to see a substrate effect of Yeniseian tone in the so-called pharyngealised vowels of Tuvan and Tofa (Georg 2008:155), which correspond to preaspiration on the following consonant in e.g. Western Yugur, spoken in Gansu, well outside even the widest Yeniseian area.

people associated with the Afanas'ev Culture were, also linguistically, not very different from those associated with the Yamnaya Culture.

The easiest way out is definitely to say that the 4 Afanas'ev individuals that were tested (Allentoft et al. 2015: supplementary table 9, supplementary materials p. 43) are simply not enough, and the picture may change if more individuals from throughout the Afanas'ev area and period are tested. However, another solution is also possible. It is highly unlikely that all Afanas'ev people gathered and left the area together to move south into the Tarim Basin. Rather, some smaller groups will have split off and moved away. It is therefore entirely possible that the contact situation discussed here concerned only a small portion of the Afanas'ev people.

Another option, which certainly does not exclude the preceding, is that the contacts are to be situated only towards the end of the Afanas'ev period. At that time, some admixture took place between the Afanas'ev and the newly arriving Okunevo populations (Parzinger 2001), since “there is an admixture signal of 10 to 20% Yamnaya and Afanasievo” in the 19 Okunevo individuals from the Minusinsk Basin tested by Damgaard et al. (2018 with supplementary fig. S21), and overlap between Afanas'ev and Okunevo is also recorded archaeologically (Mallory 2015:38, citing Sokolova 2011). In the same article, Damgaard et al. analyse two female individuals, labeled “CentralSteppe_EMBA,” from Afanas'ev-like pits from Sholpan at the southwest tip of Lake Balkhash in Kazakhstan dating from ca. 2200 BCE (Damgaard et al. 2018: fig. 1; supplementary table S4). Interestingly, these are genetically “almost indistinguishable” from the Okunevo individuals that were tested. Even more striking, one has mtDNA haplogroup C4 (Damgaard et al. 2018, Fig. 5B; the other has C4a1a4a), which is remarkably frequent in the oldest Tarim mummies (Li et al. 2010; 2015).

Although it is certainly premature to speak of proof in any strict sense of the word, these data are neatly compatible with the Tocharian Migration Hypothesis. The evidence thus far available inspires several subhypotheses that could be tested in the future, such as:

- The people associated with the Afanas'ev Culture remained for a long time unadmixed with indigenous Siberian populations.
- Admixture took place only when Okunevo-related populations arrived.
- In the admixture with Okunevo the Afanas'ev-related element was male-derived (cf. Damgaard et al. 2018).
- The arrival of the same Okunevo-related people prompted some Afanas'ev-related groups to leave the area.
- Admixture with Okunevo-related people possibly continued even after some Afanas'ev groups left, “on the way” (again, male-derived).

- The route from the Afanas'ev area in and around the northern Altai region to the Tarim Basin led southwest onto the steppe (and then, necessarily, southeast, probably through the Dzungar Basin⁴⁰).

The crucial point for a historical scenario for the linguistic contacts discussed here is, obviously, whether it is possible to identify the Okunevo-related populations linguistically. Likewise, it is extremely important to know whether Pre-Proto-Samoyedic and Proto-Yeniseian can be identified with prehistoric cultures. There is no point in concealing that it would suit my case if the Okunevo-related populations spoke Pre-Proto-Samoyedic. They could have been in contact with Yeniseian speakers just before, in the Minusinsk Basin, in the northern part of the Afanas'ev area. However, these matters cannot be decided on the basis of linguistics alone, but need to be addressed in collaboration with archaeologists and geneticists.⁴¹

4.4 *Proto-Uralic*

The interpretation of the prehistory of the area is frustrated by the lack of a clear scenario for the Uralic homeland, such as we have for Indo-European.⁴² This would make it much easier to situate early Samoyedic in place and time. Stressing again that firm genetic and archaeological evidence is needed, I would like to sketch an alternative scenario that differs from that investigated above, but might also be consistent with the linguistic evidence.

If the Pre-Proto-Tocharian seven-vowel system developed before the contacts with Yeniseian or Uralic, as is perhaps suggested by the mechanisms behind the changes (see §3), there is, strictly speaking, no need to identify the Uralic substrate as an early form of Samoyedic: the identifying feature was precisely the parallelism in the vowel systems. This leaves room for the possibility that the Okunevo Culture is not to be identified with early Samoyedic, but with Proto-Uralic. This is consistent with Janhunen's convincing arguments that the Ural-Altai typological profile of Uralic and the primary split

40 It is doubtful whether the Qiēmùěrqìèkè Culture of northern Xīnjiāng represents this passage through the Dzungar Basin. Mallory sees no good connection to either the Afanas'ev Culture or the Xiāohé Horizon (2015: 45 and *passim*; see most recently Betts et al. 2019).

41 Sokolova derives the Okunevo from the Neolithic Ust'-Belaya Culture, located between the Middle Yenisei and the Baikal region, dating to ca. 6500–4000 BCE (2011:252). In order to identify the Okunevo linguistically, we would also need to know if a connection can be made with Finno-Ugric-associated cultures. Finally, a genetic connection to Uralic groups would be needed.

42 This is not the place to discuss the Uralic homeland problem. The literature is extensive. For a recent contribution, cf. Nichols & Rhodes (2018).

between Samoyedic and Finno-Ugric point to an eastern origin (2001; 2009), and it would be just in time for Finno-Ugric to split off and move west towards the Ural Mountains, where this branch was influenced by Proto-Indo-Iranian (e.g. Kuz'mina 2001).⁴³ The Yeniseian impact on Samoyedic could then have occurred when the Samoyeds stayed in the area or moved north. In this scenario, it is possible that the vowel system of Proto-Yeniseian⁴⁴ developed under the influence of early Tocharian.

5 Conclusion

The parallels between Tocharian, Uralic and Yeniseian that I have presented in this paper show that Tocharian must have gone through a “Siberian” phase in its development. The most important feature of Tocharian showing Uralic impact is the reduction of the three Proto-Indo-European stop series to one series of voiceless stops. While agglutinative case inflexion is widely found in Siberia, case functions such as the Tocharian perlativ ‘through, along, over’ indicate Uralic or South Siberian influence also in this domain. Close parallels in the vowel systems of early stages of Tocharian and Samoyedic point specifically to this branch of Uralic, and parallels with Yeniseian in the same domain further confirm that the contacts are to be located in Southern Central Siberia. A number of other features of Tocharian, such as the use of converbs and object marking on the verb, are perhaps also attributable to Uralic influence, but they are of secondary importance compared to the main arguments from the stops, the vowels and the agglutinative case inflexion.

The fact that Tocharian linguistic prehistory is to be placed in part in Siberia provides important, and so far completely missing, support for the Tocharian Migration Hypothesis, in which it is claimed that the Afanas'ëvo Culture of South Siberia can be identified as an early station in the trajectory of early speakers of Tocharian towards the Tarim Basin.

It seems that the succession of the Afanas'ëvo Culture by the Okunevo Culture has played a decisive role in the development of Tocharian, but the linguistic identification of the Okunevo Culture is uncertain. Therefore, a more

43 The speed of this westward movement of Finno-Ugric could be compared in scale with the presumed southward movement of Tocharian. Both could have taken place in the second half of the third millennium BCE. While Finno-Ugric would have had to move further than Tocharian, the latter had a more complicated route.

44 I.e., the Proto-Yeniseian reconstructible on the basis of the historically attested languages; not “hydronym Yeniseian.”

precise interpretation of the prehistoric reality of the contact situation remains speculative at this point and further research combining linguistics with genetics and archaeology is needed.

Both in the case of the assumed Uralic impact on Tocharian and in the case of the Yeniseian impact on Samoyedic, the resulting changes are far-reaching. It is not exaggerated to say that these changes define the respective subbranches within their families. More impressionistically, one could say that these contacts have led to the birth of Tocharian and of Samoyedic.

Acknowledgments

This paper is an adaptation of a lecture with the title “Tocharian as a Central Asian language” held at the conference “Ancient texts and languages of the ethnic groups along the Silk Road” on 5 November 2018 in Göttingen. This research was financed by the European Research Council (ERC-2017-STG 758855). For valuable discussions about this paper and the topics it treats I am grateful to Juha Janhunen (Helsinki), Frits Kortlandt (Leiden), Sasha Lubotsky (Leiden), James Mallory (Belfast) and Edward Vajda (Bellingham). I am further grateful for valuable comments by two anonymous reviewers as well as by the editor Ronald Kim.

References

- Aalto, Pentti. 1964. Word-pairs in Tocharian and other languages. *Linguistics* 5:69–78.
- Abondolo, Daniel. 1998. Introduction. In: Daniel Abondolo (ed.), *The Uralic languages*. London, 1–42.
- Aikio, Ante. 2006. New and old Samoyed etymologies (Part 2). *Finnisch-Ugrische Forschungen* 59:9–34.
- Aikio, Ante. 2012. On Finnic long vowels, Samoyed vowel sequences, and Proto-Uralic *x. In: Tiina Hyytiäinen (ed.), *Per Urales ad Orientem. Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Helsinki, 227–250.
- Aikio, Ante. 2014a. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62:7–76.
- Aikio, Ante. 2014b. Studies in Uralic etymology III: Mari etymologies. *Linguistica Uralica* 50:81–93.
- Allentoft, Morten E. et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.

- Anderson, Gregory D.S. 2003. Yeniseic languages from a Siberian areal perspective. *Sprachtypologie und Universalienforschung* 56:12–39.
- Anthony, David W. 2007. *The horse, the wheel and language*. Princeton.
- Anthony, David W. 2013. Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism. *The Journal of Language Relationship* 9:1–21.
- Bednarczuk, Leszek. 2015. Non-Indo-European features of the Tocharian dialects. In: E. Mańczak-Wohlfeld & B. Podolak (eds.), *Words and dictionaries. A Festschrift for Professor Stanisław Stachowski on the occasion of his 85th birthday*. Kraków, 55–67.
- Belyaev, Oleg. 2010. Evolution of case in Ossetic. *Iran and the Caucasus* 14:287–322.
- Betts, Alison et al. 2019. A new hypothesis for early Bronze Age cultural diversity in Xinjiang, China. *Archaeological Research in Asia* 17:204–213.
- Blažek, Václav & Michal Schwarz. 2008. Tocharians. Who they were, where they came from and where they lived. *Lingua Posnaniensis* 50:47–74.
- Blažek, Václav & Michal Schwarz. 2017. *Early Indo-Europeans in Central Asia and China*. Innsbruck.
- Carling, Gerd. 2005. Proto-Tocharian, Common Tocharian, and Tocharian—on the value of linguistic connections in a reconstructed language. In: Karlene Jones-Bley et al. (eds.), *Proceedings of the Sixteenth Annual UCLA Indo-European Conference: Los Angeles, November 5–6, 2004*. Washington, 47–70.
- Carling, Gerd. 2012. Development of form and function in a case system with layers: Tocharian and Romani compared. *Tocharian and Indo-European Studies* 13:57–76.
- Castrén, M. Alexander. 1844. *Elementa grammatices syrjaenae*. Helsingfors.
- Castrén, M. Alexander. 1854. *Grammatik der samojedischen Sprachen*. St. Petersburg.
- Castrén, M. Alexander. 1858. *Versuch einer jennissei-ostjakischen und kottischen Sprachlehre*. St. Petersburg.
- Ching Chao-jung. 2017. *Tǔhuǒluó yǔ shìsù wénxiàn yǔ gǔdài Qiūcǐ lìshǐ—Tocharian Secular Texts and the History of Ancient Kucha*. Běijīng.
- Comrie, Bernard. 1988. General features of the Uralic languages. In: Denis Sinor (ed.), *The Uralic languages. Description, history and foreign influences*. Leiden, 451–477.
- Comrie, Bernard. 1989. *Linguistic universals and linguistic typology. Syntax and morphology*. 2nd edn. Oxford.
- Damgaard, Peter de Barros et al. 2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360:1422. doi.org/10.1126/science.aar7711
- Delbrück, Berthold. 1888. *Altindische Syntax*. Halle.
- Driessen, C. Michiel. 2003. **h₂é-h₂us-o-*, the Proto-Indo-European term for 'gold'. *Journal of Indo-European Studies* 31:347–362.
- Emeneau, M.B. 1956. India as a linguistic area. *Language* 32:3–16.
- Erdal, Marcel. 2004. *A grammar of Old Turkic*. Leiden.
- Georg, Stefan. 2008. Yeniseic languages and the Siberian linguistic area. In: Alexander

- M. Lubotsky et al. (eds.), *Evidence and counter-evidence. Essays in honour of Frederik Kortlandt*. Volume 2: *General linguistics*. Amsterdam, 151–168.
- Haak, Wolfgang et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207–211. doi.org/10.1038/nature14317
- Hackstein, Olav. 2017. The phonology of Tocharian. In: Jared Klein et al. (eds.), *Handbook of comparative and historical Indo-European linguistics*. Volume 2. Berlin, 1304–1335.
- Häkkinen, Jaakko. 2009. Kantauralin ajoitus ja paikannus: perustelut puntarissa. *Journal de la Société Finno-Ougrienne* 92:9–56.
- Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntactic status of bound person forms. In: Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries. Studies in memory of Anna Siewierska*. Berlin, 197–226.
- Helimski, Eugen [Xelimskij, Evgenij A.]. 1978. Rekonstrukcija prasevernosamodijiskix (PSS) labializovannyx glasnyx nepervyx slogov. In: V.N. Jarceva et al. (eds.), *Konferencija "Problemy rekonstrukcii" (Tezisy dokladov)*. Moskva, 123–126.
- Helimski, Eugen [Xelimskij, Evgenij A.]. 1993. Prasamodijskie *â i *ǝ: praural'skie istočniki i nganasanskije refleksy. In: Marianne Sz. Bakró-Nagy & Enikő Szij (eds.), *Hajdú Péter 70 éves*. Budapest, 125–133.
- Helimski, Eugene. 2004. An outline history of the Samoyed people. In: György Nanovszky (ed.), *The Finno-Ugric world*. Budapest, 119–133.
- Helimski, Eugen. 2005. The 13th Proto-Samoyedic vowel. In: Beáta Wagner-Nagy (ed.), *Mikola-konferencia 2004*. Szeged, 27–39.
- Hoffner, Harry A. & H. Craig Melchert. 2008. *A grammar of the Hittite language*. Part 1: *Reference grammar*. Winona Lake.
- Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77:219–274.
- Janhunen, Juha. 1982. On the structure of Proto-Uralic. *Finnisch-ugrische Forschungen* 44:23–42.
- Janhunen, Juha. 1983. On Early Indo-European-Samoyed contacts. In: Juha Janhunen et al. (eds.), *Symposium saeculare Societatis Fenno-Ugricae*. Helsinki, 115–127.
- Janhunen, Juha. 1986. *Glottal stop in Nenets*. Helsinki.
- Janhunen, Juha. 1996. *Manchuria. An ethnic history*. Helsinki.
- Janhunen, Juha. 1998. Samoyedic. In: Daniel Abondolo (ed.), *The Uralic languages*. London, 457–479.
- Janhunen, Juha. 2001. Indo-Uralic and Ural-Altaic: On the diachronic implications of areal typology. In: Christian Carpelan et al. (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*. Helsinki, 207–220.
- Janhunen, Juha. 2009. Proto-Uralic—what, where, and when? In: Jussi Ylikoski (ed.), *The Quasquicentennial of the Finno-Ugric Society*. Helsinki, 57–78.
- Jasanoff, Jay H. 1978. *Stative and middle in Indo-European*. Innsbruck.

- Jasanoff, Jay H. 1989. Language and gender in the Tarim Basin: the Tocharian 1 sg. pronoun. *Tocharian and Indo-European Studies* 3:125–147.
- Joki, Aulis J. 1944. *Kai Donner's kamassisches Wörterbuch nebst Sprachproben und Hauptzügen der Grammatik*. Helsinki.
- Kallio, Petri. 2004. Tocharian loanwords in Samoyed? In: Irma Hyvärinen et al. (eds.), *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag*. Helsinki, 129–137.
- Kim, Ronald I. 1999. The development of labiovelars in Tocharian: A closer look. *Tocharian and Indo-European Studies* 8:139–187.
- Kim, Ronald I. 2018. *The dual in Tocharian: from typology to 'Auslautgesetz'*. Dettelbach.
- Klumpp, Gerson. 2002. *Konverbkonstruktionen im Kamassischen*. Wiesbaden.
- Kortlandt, Frederik H.H. 2019. On the reconstruction of Proto-Uralic. In: Santeri Junttila & Juha Kuokkala (eds.), *Petri Kallio rocks. Liber semisaecularis 7.2.2019*. Helsinki, 11–14.
- Krause, Wolfgang. 1951. Zur Frage nach dem nichtindogermanischen Substrat des Tocharischen. *Zeitschrift für vergleichende Sprachforschung* 69:185–203.
- Krejnovič, E.A. 1958. *Jukagirskij jazyk*. Moskva.
- Krejnovič, E.A. 1968. Jukagirskij jazyk. In: P.Ja. Skorik (ed.), *Jazyki narodov SSSR*. Volume 5: *Mongol'skie, tunguso-man'čžurskie i paleoaziatskie jazyki*. Leningrad, 435–452.
- Kroonen, Guus J. et al. 2018. Linguistic supplement to Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian. doi.org/10.5281/zenodo.1240524
- Kuz'mina, E.E. 2001. Contacts between Finno-Ugric and Indo-Iranian speakers in the light of archaeological, linguistic and mythological data. In: Christian Carpelan et al. (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*. Helsinki, 289–300.
- Leumann, Manu et al. 1965. *Lateinische Grammatik*. Zweiter Band: *Syntax und Stilistik*. München.
- Li Chunxiang et al. 2010. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biology* 8:15. doi.org/10.1186/1741-7007-8-15
- Li Chunxiang et al. 2015. Analysis of ancient human mitochondrial DNA from the Xiaohe cemetery: insights into prehistoric population movements in the Tarim Basin, China. *BMC Genetics* 16:78. doi.org/10.1186/s12863-015-0237-5
- Mallory, James P. 1989. *In search of the Indo-Europeans. Language, archaeology and myth*. London.
- Mallory, James P. 2015. *The problem of Tocharian origins*. Philadelphia (Sino-Platonic Papers 259).

- Mallory, James P. & Victor H. Mair. 2000. *The Tarim mummies. Ancient China and the mystery of the earliest peoples from the West*. London.
- Maloletko, A.M. 2002. *Drevnie narody Sibiri. Ètničeskij sostav po dannym toponimiki*. Tom. II: *Kety*. 2nd edition. Tomsk.
- Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir*. Berlin.
- Meier, Kristin & Michaël Peyrot. 2017. The word for 'honey' in Chinese, Tocharian and Sino-Vietnamese. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 167:7–22.
- Meunier, Fanny. 2015. Recherches sur le génitif en tokharien. Diss. EPHE, Paris.
- Mikola, Tibor. 1988. Geschichte der samojedischen Sprachen. In: Denis Sinor (ed.), *The Uralic languages. Description, history and foreign influences*. Leiden, 219–263.
- Mikola, Tibor. 2004. *Studien zur Geschichte der samojedischen Sprachen*. Szeged.
- Napol'skikh, Vladimir. 2001. Tocharisch-uralische Berührungen: Sprache und Archäologie. In: Christian Carpelan et al. (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*. Helsinki, 367–383.
- Nichols, Johanna & Richard A. Rhodes. 2018. Vectors of language spread at the central steppe periphery: Finno-Ugric as a catalyst language. In: Rune Iversen & Guus Kroonen (eds.), *Digging for words*. Oxford, 58–68.
- Nikolaeva, Irina. 2006. *A historical dictionary of Yukaghir*. Berlin.
- Nikolaeva, Irina. 2014. *A grammar of Tundra Nenets*. Berlin.
- Odé, C. 2012. Segmental phonetics also has its charms. In: L.A. Verbickaja & N.K. Ivanova (eds.), *Čelovek govorjaščij: issledovanija XXI veka. Kollektivnaja monografija k 80-letiju so dnja roždenija Lii Vasil'evny Bondarko*. Ivanovo, 34–43.
- Parzinger, Hermann. 2001. Südsibirien in der Spätbronze- und Früheisenzeit. In: Ricardo Eichmann & Hermann Parzinger (eds.), *Migration und Kulturtransfer. Der Wandel vorder- und zentralasiatischer Kulturen im Umbruch vom 2. zum 1. vorchristlichen Jahrtausend*. Bonn, 71–83.
- Pedersen, Holger. 1931. *The discovery of language*. Cambridge (MA).
- Pedersen, Holger. 1941. *Tocharisch vom Gesichtspunkt der indoeuropäischen Sprachvergleichung*. København.
- Peyrot, Michaël. 2012. The Tocharian A match of the Tocharian B obl.sg. *-ai*. *Tocharian and Indo-European Studies* 13:181–220.
- Peyrot, Michaël. 2013. *The Tocharian subjunctive. A study in syntax and verbal stem formation*. Leiden.
- Peyrot, Michaël. 2017a. Tocharian: An Indo-European language from China. In: Jorrit M. Kelder et al. (eds.), *Aspects of globalisation. Mobility, exchange and the development of multi-cultural states*. Leiden, 12–17.
- Peyrot, Michaël. 2017b. Slavic *onъ*, Lithuanian *anàs* and Tocharian A *anac*, *anäs*. In: Bjarne S.S. Hansen et al. (eds.), *Usque ad radices. Indo-European studies in honour of Birgit Anette Olsen*. Copenhagen, 633–642.

- Peyrot, Michaël. 2018a. On the part of speech and the syntax of the Tocharian present participle. In: Claire Le Feuvre et al. (eds.), *Verbal adjectives and participles in Indo-European languages / Adjectifs verbaux et participes dans les langues indo-européennes*. Bremen, 327–341.
- Peyrot, Michaël. 2018b. 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*. Ann Arbor, 270–283.
- Peyrot, Michaël. fthc. Indo-Uralic, Indo-Hittite, Indo-Tocharian.
- Pinault, Georges-Jean. 1994. Lumières tokhariennes sur l'indo-européen. In: Jens E. Rasmussen & Benedicte Nielsen (eds.), *In honorem Holger Pedersen*. Wiesbaden, 365–396.
- Pinault, Georges-Jean. 2015. Buddhist stylistics in Central Asia. *Linguarum varietas* 4:89–107.
- Pronk, Tijmen. 2015. On the origin of the dual endings Tocharian A *-ŋ*, B *-ne*. In: Melanie Malzahn et al. (eds.), *Tocharian texts in context*. Bremen, 199–214.
- Salminen, Tapani. 2012. Traces of Proto-Samoyed vowel contrasts in Nenets. In: Tiina Hyttiäinen et al. (eds.), *Per Urales ad Orientem. Iter polyphonicum multilingue*. Helsinki, 339–358.
- Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages, with special reference to Samoyed, Ugric and Permic. In: Denis Sinor (ed.), *The Uralic languages. Description, history and foreign influences*. Leiden, 478–554.
- Schaller, Helmut W. 1975. *Die Balkansprachen. Eine Einführung in die Balkanphilologie*. Heidelberg.
- Schmidt, Karl H. 1990. The postulated Pre-Indo-European substrates in Insular Celtic and Tocharian. In: Thomas L. Markey & John A.C. Greppin (eds.), *When worlds collide. The Indo-Europeans and the Pre-Indo-Europeans*. Ann Arbor, 179–202.
- Schönig, Claus. 2003. Turko-Mongolic relations. In: Juha Janhunen (ed.), *The Mongolic languages*. London, 403–419.
- Schulze, Wilhelm. 1927. Zum Tocharischen. *Ungarische Jahrbücher* 7:168–177.
- Sieg, Emil & Wilhelm Siegling. 1908. Tocharisch, die Sprache der Indoskythen. Vorläufige Bemerkungen über eine bisher unbekannte indogermanische Literatursprache. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1908:915–932.
- Sokolova, L'udmila A. 2011. *Formirovanie okunevskogo kul'turnogo kompleksa*. Saarbrücken.
- Stang, Christian S. 1966. *Vergleichende Grammatik der baltischen Sprachen*. Oslo.
- Starostin, Sergej A. 1982. Praenisejskaja rekonstrukcija i vnešnie svjazi enisejskix jazykov. In: E.A. Alekseenko et al. (eds.), *Ketskij sbornik. Antropologija, ètnografija, mifologija, lingvistika.—Studia Ketica. Physical anthropology, ethnography, mythology, linguistics*. Leningrad, 144–237.

- Starostin, Sergej A. 1995. Sravnitel'nyj slovar' enisejskix jazykov. In: Sergej A. Starostin (ed.), *Ketskij sbornik. Lingvistika.—Studia Ketica. Linguistics*. Moskva, 176–315.
- Svyatko, Svetlana V. et al. 2017. Stable isotope palaeodietary analysis of the Early Bronze Age Afanasyevo Culture in the Altai Mountains, Southern Siberia. *Journal of Archaeological Science: Reports* 14:65–75.
- SWTF = *Sanskrit-Wörterbuch der buddhistischen Texte aus den Turfan-Funden*. Edited by Heinz Bechert et al. Göttingen 1973–2018.
- Thomas, Werner. 1958. Zum Ausdruck der Komparation beim tocharischen Adjektiv. *Zeitschrift für Vergleichende Sprachforschung* 75:129–169.
- Vadeckaja, Èl'ga B. et al. 2014. *Svod pamjatnikov afanas'evskoj kul'tury*. Barnaul.
- Vajda, Edward J. 2004. *Ket*. München.
- Vajda, Edward J. 2010. A Siberian link with Na-Dene languages. In: James Kari & Ben A. Potter (eds.), *The Dene-Yeniseian connection*. Fairbanks, 33–99.
- Vajda, Edward J. 2018. Dene-Yeniseian. Progress and unanswered questions. *Diachronica* 35:277–295.
- Vajda, Edward J. 2019. Yeniseian and Dene toponyms. In: Gary Holton & Thomas Thornton (eds.), *Language and Toponymy in Alaska and Beyond: Papers in Honor of James Kari*. Honolulu, 174–190.
- Van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg.
- Wagner-Nagy, Beáta. 2018. *A grammar of Nganasan*. Leiden.
- Werner, Heinrich. 1997a. *Die ketische Sprache*. Wiesbaden.
- Werner, Heinrich. 1997b. *Abriss der kottischen Grammatik*. Wiesbaden.
- Werner, Heinrich. 1998. *Probleme der Wortbildung in den Jenissej-Sprachen*. München.
- Werner, Heinrich. 2002. *Vergleichendes Wörterbuch der Jenissej-Sprachen*. Wiesbaden.
- [3 vols]
- Wickman, Bo. 1955. *The form of the object in the Uralic languages*. Uppsala.
- Winter, Werner. 1962. Nominal and pronominal dual in Tocharian. *Language* 38:111–134.