Tocharian and Samoyed: on the question of Uralic substrate influence in Tocharian

Warries, A.R.

10 Conclusions

There are many linguistic features of Tocharian that show a correspondence with Uralic and Samoyed. I have discussed them in this dissertation in order to test and evaluate the hypothesis that Tocharian was influenced by Uralic or, more specifically, early Samoyed. In this conclusion, I will give a summary of the results of the investigations carried out in each chapter and answer the various research questions posed in the introduction.

10.1 Overview

Previous research on Tocharian-Uralic/Samoyed contact hypotheses has stressed on the Tocharian stop system (e.g., Kallio 2001, Bednarczuk 2015, Peyrot 2019a) and, with a specific connection to the Samoyed branch, the vowel system (Peyrot 2019a). The stop systems indeed correspond quite closely if we consider pre-Proto-Tocharian before phonemic palatalization. The pre-Proto-Tocharian phonological system at that point in the relative chronology also provides the closest match for the pre-Proto-Samoyed vowel system. However, the specificity correspondence between the pre-Proto-Tocharian vowel system and the pre-Proto-Samoyed vowel system may have been overestimated (Peyrot 2019a; differently, but still perhaps too optimistically Warries 2022), so that it cannot easily be decided how much value should be assigned to it in relation to a specifically Samoyed connection. The pre-Proto-Tocharian vowel system is also very similar to the Proto-Uralic vowel system. On the phonological comparison in more detail, see chapter 4.

The Tocharian accent or stress system probably passed through a stage of general word-initial accentuation, which corresponds to the stress placement of Uralic, although it is difficult to determine this with certainty due to later changes. Initial stress is typologically common, so that it does not constitute a highly specific match. For more details on the Tocharian accent, see chapter 5.

The agglutinative case system of Tocharian has frequently been connected with substrate interference (e.g., Krause 1951, K.H. Schmidt 1990, Bednarczuk 2015, Peyrot 2019a). Uncertainties around the age of the Tocharian secondary cases and differences in meaning make the comparison with the Samoyed case system insecure. This is discussed in chapter 6.

Tocharian agrees with Samoyed, and other eastern Uralic languages more generally, in the wide use of contextual orientation in the participles, a fact that had not been previously pointed out as a possible contact feature. The participial system, with its distinction between present, past, negative (privative) and necessitative participles (present gerundive) is also structurally reminiscent of the Samoyed participial system. See chapter 7 for more information.

The Tocharian use of pronoun suffixes to mark pronominal objects is similar to the Uralic objective conjugation, but the constructions in the two language groups differ in their origins and in the specifics of the types of syntactic structures that they could be used for. This may be understood as the result of the building blocks that were available in pre-Proto-Tocharian to make such a structure in the first place, so that it can plausibly still be linked with Uralic influence. The age of the objective conjugation in Uralic itself is a disputed problem, but I believe that it goes back to Proto-Uralic and probably finds its origins in a pre-Proto-Uralic stage. This issue is treated in chapter 8.

Finally, a handful of potential loanwords might support the connection with Samoyed specifically, but the evidence does not reach a critical mass (Janhunen 1973, Napol'skikh 2001, Kallio 2004, Warries fthc.). Proposed Tocharian loanwords in other Uralic languages or Proto-Uralic itself are too uncertain to use as evidence for prehistoric contact. An overview can be found in chapter 9.

10.2 Phonology

Tocharian phonology underwent major changes from Proto-Indo-European. The loss of distinctions related to voicing or aspiration, as well as vowel length, constitute significant innovations that resulted in a Uralic-looking phonological system in Tocharian. No other known language in the area shares this combination of features, with the possible exception of the still poorly understood pre-Proto-Yukaghir (see the discussions in 4.2.2 and 4.3.2). This makes early Uralic/Samoyed the most reliable source of substrate influence in this regard, according to our current knowledge.

In my discussion of Tocharian historical phonology, I treated some problematic points, like the timing of palatalization and umlaut, the merger of PIE **o* and **ē*, the reflexes of PIE **eh₂* and **ō*, and the developments of PIE **d*. I concluded that the rise of the palatalized vs. non-palatalized opposition in consonants can be connected with the merger of PIE **i* and the **ə* that arose from syllabic resonants and anaptyxis in zero-grades as **i*, and the merger of PIE **o* and **ē* as **e* (2.4). The merger of **i* and **u* as **ə* has to be dated later, since the **u* caused *u*-umlaut after palatalization had been completed, indicating that it must have remained distinct for a longer period of time. The merger of **o* and **ē* as **e* can be further supported by the fact that these vowels were affected in the same way by *u*-umlaut and *a*-umlaut, which can both be dated to the (pre-)Proto-Tocharian period as well (2.5.1 and 2.5.5).

The development of **ō* remains difficult to establish due to a lack of reliable examples. However, it is certain that it became PT **o* in the second syllable in a few examples, so that no general development to PT **a* can be posited (2.4.5). Whether the PT **o* from marginal **ō* and the PT **o* that resulted from umlaut need to be separated from the reflex of PIE **eh₂* is uncertain, due to the small group of examples showing contradictory

reflexes. I suggest that *eh_2 to PT *$å$ as opposed to *o from umlaut and *$ō$ could be merged simply as PT *o if the two examples that show a reflex *a in Tocharian A are assumed to be the result of dissimilation from a preceding labial consonant *w-* or *pr-* (2.5.6.4). However, this remains tentative. A phoneme PT *$å$ can in any case be reconstructed as the reflex of umlauted *a in the *o*-presents [class 4] and in TB *onolme* 'living being' (2.5.4). This *$å$ does regularly yield TB *o* and TA *a*, and thus should not be equated with at least the result of *u*-umlaut or *o*-umlaut of *e, as this normally becomes TAB *o* in the first syllable.

The consonant with the most varied and obscure development in Tocharian is PIE *d, with its various reflexes as PT *t^s, *$ś$, *t and ∅. I concluded that the different reflexes can be interpreted in various ways depending on the assumed phonetic interpretation of PIE *d, and that the Tocharian data is insufficiently clear to allow us to infer the original state of *d in the stage of Proto-Indo-European from which Tocharian descends. A relatively good candidate for an initial stage in the development is a [ð], as this is a natural intermediate step on the way to full loss (the reflex ∅), and lenition from [d] to [ð] can be blocked in an understandable way by a preceding nasal, so that the (devoiced) reflex *t in the original cluster *nd is accounted for. However, the development from [ð] to TAB t^s remains difficult to understand, and I know no direct parallels for such a change. A glottalic interpretation of *d could accord with the reflex ∅ by assuming debuccalization to a [ʔ], and perhaps an originally glottalized dental stop could have become TAB t^s and *ś* by assuming assibilation from a potential (but also unverifiable) intermediate development to a [t^h] (for details, see 2.3.4.3–2.3.4.6).

For the relative chronology of Tocharian, the following conclusion is particularly relevant: at the time that palatalization took place, a vowel system without length distinctions can be reconstructed. This matches very well with the lack of vowel length distinctions in the early Samoyed phonological system, which also cannot be reconstructed with pervasive palatalization oppositions of the type that we see in Tocharian. This means that the phonological systems of pre-Proto-Tocharian and pre-Proto-Samoyed were structurally very similar at some point in the distant past. The pre-Proto-Samoyed system did not differ significantly from the Proto-Uralic system as far as the relevant features are concerned, so that this does not constitute a highly specific match between early Tocharian and Samoyed specifically.

In my discussion of Samoyed historical phonology, I concluded that the Proto-Samoyed reflex *o rather than *u in stems of the type PS *CoC ~ FU *$CuCi$ can be most easily understood as an archaism in Samoyed, in accordance with Janhunen's (1981: 231) original formulation of this correspondence (3.6.3). The split development of PU *a-a into PS *$å$-$å$ and *a-$ə̂$ should probably also be understood as some kind of archaism, since there does not seem to be any Samoyed-internal way to generate the two different reflexes, at least not with our current understanding of Samoyed historical phonology.

Zhivlov's (2014) suggested reconstruction of two different *a*-vowels in the second syllable in Proto-Uralic thus needs to be seriously considered and further researched (3.7.1).

A difficult point in the relative chronology of Samoyed is the split of PU **ę* into **ę* and **į*. The general distribution can be understood on the basis of the following vowel, with PS **ę* resulting from the original sequence PU **ę-i* and PS **į* from PU **ę-a*. However, there also appears to have been an interaction with coda nasals, which is clearly reflected in the developments of **lęmpi* 'swamp' and **lęnti* 'plain, valley', and perhaps in **joŋsi* 'bow' if this developed via **jęŋsi*. The absence of such a development of **δ´ęmi* 'bird cherry' to PS **jęm* rather than ***jįm* might indicate that the change from **ę* to **į* took place before apocope of **-i*, but paradigmatic alternations could always have restored the vowel, analogically restoring hypothetical **jįm* : gen.sg. **jęmân* to **jęm* : *jęmân*. This makes the dating of this vowel development relative to apocope uncertain (3.6.2).

In general, paradigmatic alternations resulting from some of the Samoyed sound changes should be borne in mind, as with the split development of roots of the shapes **CVli* and **CVji* to both **CVj* and **CVə* (Aikio 2012). The regular development is that **CV{l/j}i* became **CVə* when the final **-i* was preserved, and **CVj* in consonant stems or after apocope. The resulting alternations were levelled in most roots, but the original distribution can be seen in the verb PS **åâ- ~ *åj* 'be'. Other changes that are expected to have caused paradigmatic alternations include the shift from **o* to **å* in **CoCa* and **CoCCi* but not **CoCi*. The split vowel reflex of the verb PU **ćoji-* 'sound' and the derived noun **ćojma* 'sound' into PS **soj-* and **såjmå* respectively shows this. Aside from **såjmå*, other derivations and inflectional forms of **soj-* should regularly have undergone the same change, e.g., quasi-PU **ćojća-* (pst.), **ćojnta* (ipfv.ptc.), **ćojŋa-* (aor.) must have developed into **såjså-*, **såj(n)tå*, **såjŋa* by regular sound law, but were evidently restored based on forms like the imperative/connegative **ćoji-k* or the old prs.3sg. **ćoji*, where no development from **o* to **å* is expected. The forces of analogical regularization of paradigms should not be underestimated in the historical phonology of Samoyed (3.4.2).

10.3 Accent

The Tocharian accent, as can be deduced from Tocharian B only, does not conform to the Samoyed or Uralic stress on the first syllable. Rather, Tocharian B had a general second-syllable accent, albeit with many exceptional categories where the accent was fixed on the first syllable for various reasons. From the discussion of the diachronic development of the Tocharian accent in chapter 5, I concluded that earlier Tocharian would have had initial accent, and that the accent patterns found in Tocharian B are partially innovative. The pre-Proto-Tocharian initial accent is different from both the ancestral Proto-Indo-European mobile accent, and from later Tocharian B, but it does

accord with the first-syllable stress of early Uralic and Samoyed. It thus provides an additional phonological feature of pre-Proto-Tocharian that may have been taken over from early Uralic/Samoyed. Since first-syllable accentuation is quite common in languages around the world, it does not provide a particularly specific point of comparison, however, and other languages could also have caused this type of shift in early Tocharian.

10.4 Morphology

The morphological comparisons between Tocharian and Uralic/Samoyed are difficult to evaluate. The agglutinative case system of Tocharian offers a general match with Uralic/Samoyed, but it is not quite specific. The meanings of the Tocharian locative and perlative correspond at their core to the Samoyed locative and prolative, but they also have a wider functional range that is not paralleled. Especially the static use of the Tocharian perlative, and illative use of the Tocharian locative find no direct correspondence in Samoyed. It is unknown how much of this discrepancy may be due to later change in Tocharian, or whether it can be attributed to the result of the original meaning of the elements that were used to form the case suffixes.

Furthermore, the Tocharian agglutinative case system was clearly productive even after the Proto-Tocharian period, with diverging cases and case endings appearing in Tocharian A and B. This causes uncertainty over the true age of the Tocharian cases: they may (partially?) be old enough to result from language contact in the distant past, but some elements of the system could also be relatively too young for that. The Tocharian perlative case does receive its most appropriate counterpart in the Samoyed prolative, which is not reconstructed for Proto-Uralic. This may connect pre-Proto-Tocharian to pre-Proto-Samoyed specifically, rather than early Uralic more generally, but it is unknown how old exactly the Samoyed prolative case itself is. In principle it could be younger than the Tocharian perlative, so that the connection must remain tentative (6.4).

The Tocharian participial system also deviates from that of the Indo-European languages, as it contains a number of contextually oriented participles. This is most clearly seen in the preterite participle, which descends from an active Indo-European participle formation, but displays a flexible contextual range in both Tocharian languages. The privative (a negative participle) and the gerundives (necessitative or future participles) behave in a similar manner. The two present participle formations of Tocharian stand out in this context. One is an active participle that continues a Proto-Indo-European active voice participle. It commonly functions as an agent noun formation in Tocharian. The other descends from the Proto-Indo-European middle voice participle, but in Tocharian it has for the most part lost its connection with the middle voice. The common use of this participle as an adverbial form in Tocharian makes it

difficult to establish its behaviour as a true participle, but some examples of contextual orientation can be found (7.2.2).

The contextual orientation of most Tocharian participles aligns well with the common contextual orientation of Samoyed participles. The presence of negative and future participle formations also conforms to the standard Samoyed participial system (7.3). Since participle orientation of Uralic languages in the west is influenced by the orientation of surrounding Indo-European languages (Shagal 2018), it seems plausible that there could have been an obverse influence from Uralic/Samoyed participles in the east on Tocharian. However, it should be borne in mind that contextual participial systems are quite common in that part of the word, found for example in Ket, and in some Turkic, Tungusic, and Mongolic languages. This non-Indo-European feature of Tocharian can therefore not strictly be linked to Samoyed/Uralic specifically. It does, however, provide a connection with the Siberian linguistic area, and supports the other features that point to Uralic/Samoyed contact.

In verbal object marking, the Tocharian use of pronoun suffixes is similar to the Uralic objective conjugation, but the comparison is certainly not exact. The elements used to form the Tocharian object suffixes are clearly object pronouns that were suffixed to the finite verb form. The object suffix referring to a first-person singular object descends from a first-person singular clitic pronoun, and similarly with the second and third person singular and the plural.

In the Uralic/Samoyed objective conjugation, on the other hand, the objective suffixes derive from pronouns corresponding to the agent, leaving the presence of a topical object implied. In other words, the suffix referring to a first-person agent and a third person singular object shares phonological material with the first-person pronoun only. The origins and original nature of the Uralic objective conjugation are still debated. After considering the evidence, I concluded that the connection with the possessive declension is best supported (8.3.2.6). This connection is reflected most clearly in Samoyed, Mansi and Khanty, but also in traces in Mordvin. The connection to the possessive declension pertains not only to the shape of the 3sg. objective conjugation suffix *-*sA*, as is commonly acknowledged, but extends also to the marking of the number of the object with plural *-*n*-. As far as I can tell, this second connection between the objective conjugation and the possessive declension has received no attention in the discussion surrounding the origins of the objective conjugation, but I think that it clinches the argument against the hypotheses in which direct suffixation of object pronouns is assumed. The formation of the Uralic objective conjugation was thus very different to the creation of the Tocharian system of verbal pronoun suffixes. However, it is possible that the specific use of the pronoun suffixes in Tocharian as verbal suffixes rather than continued use as more general clitics was inspired by the Uralic/Samoyed objective conjugation.

10.5 Lexicon

The number of lexical comparanda between Samoyed and Tocharian remains very low. The most plausible borrowings from Tocharian into Samoyed are PS ±**mänäwjə̂* 'full moon' from pre-PT **mēnē-*, PS **måŋkə̂* 'need, poverty' from pre-PT **monko-* 'lack', and PS **tar-* 'divide, separate' from pre-PT **der-* 'split, separate', of which only the latter is a new proposal (see Warries fthc.). The etymology of PS **wẹ̑n* 'dog' as a borrowing from pre-PT **kwënə-*, the oblique stem of *ku-* 'dog' requires the simplification of **kʷ* to **w*, which has no parallel, but is otherwise plausible. PS **lə̂ŋkə̂* 'slope' is very similar to TB *leṅke* 'slope'. It has been suggested that TB *leṅke* might instead be derived from Iranian, so that the similarity between the Samoyed and Tocharian words is perhaps rather coincidental. However, the Iranian word itself does not have a good etymology, while it can be derived within Tocharian. PS **laŋkå-* 'hang', which is only reflected in Nenets, can be compared to pre-PT **lenk-* 'hang', but the very limited distribution of this word in Samoyed cannot inspire great confidence. The connection of PS **ńuə̂* 'child' with pre-PT *newo-* 'new' works quite well phonologically, but is semantically imperfect.

The connection between PS **säjt³wə̂* 'seven' and pre-PT **septəm* 'seven' is inexact, and the Samoyed word is rather to be derived from PU **ćäjćimä* 'seven'. PS **wäsa* 'metal' and pre-PT **wesa* 'gold' are extremely similar formally, and the difference in meaning may be acceptable if we assume a semantic widening or narrowing in one of the two languages. However, it is unattractive to completely separate PS **wäsa* from words for 'metal' in the other Uralic languages, such as SaaN *veaiki* 'copper', Fi. *vaski* 'copper, bronze', Hu. *vas* 'iron'. The internal correspondences are not entirely regular, pointing to a dispersal of this word among the Uralic branches after Proto-Uralic had split up, but a reconstruction **wäćkä* largely unifies the different forms. Crucially, the **-k-* of **wäćkä* cannot be derived from Tocharian. It is not present in Samoyed **wäsa*, but it would have been regularly lost in the cluster **-ćk-* in any case, so that it is impossible to argue that it was never there. A derivation of the Tocharian word from Samoyed would be phonologically possible, but only after **-ćk-* had been simplified. The full simplification to a single PS **-s-* seems to have been among the later changes in the relative chronology (see 3.9), so that a reverse direction for this loan etymology might be chronologically difficult. Such an interpretation is made further implausible due to the lack of any parallel cases.

All in all, while some connections look very attractive, the lexicon on its own does not provide a truly convincing indication that early Samoyed and early Tocharian were in contact with one another, as the small number of comparanda might be considered coincidental. It also remains possible that the contact between Tocharian and Uralic involved a para-Samoyed language that was generally identical in terms of phonology and morphology to pre-Proto-Samoyed but not the ancestor of Proto-Samoyed itself. If

this para-Samoyed was in contact with both Tocharian and pre-Proto-Samoyed, it is possible that just a few words were passed along from pre-Proto-Tocharian to pre-Proto-Samoyed via this intermediary, before para-Samoyed disappeared.

10.6 Evaluation

The results of this investigation are altogether mixed. On the one hand, a number of features that have been adduced as Uralicisms in Tocharian hold up well to closer scrutiny, especially in the phonology, and it is possible to identify more, such as in the participial system. On the other hand, there are some shared features that are widely distributed in the area, and could be attributed to influence from other languages as well, such the agglutinative case system.

Cumulatively, Uralic/Samoyed does provide the most coherent source for innovative Tocharian structural features. For instance, while Turkic also has agglutinative cases and while Ket has contextually oriented participles, these languages also have very different phonological systems when compared to Tocharian. The problem is that some other languages like pre-Proto-Yukaghir cannot be reliably reconstructed far enough back in time to test any Tocharian-Yukaghir contact hypothesis. Thus, if we try to understand Tocharian innovative features as the result of language contact, the hypothesis that the contact language was Uralic/Samoyed has the most explanatory power, and is to be preferred.

The connection with specifically the Samoyed branch of Uralic rests on a very narrow evidentiary basis according to the majority of parameters investigated here. The pre-Proto-Tocharian vowel system may be more in line with a vowel system that can be reconstructed for pre-Proto-Samoyed than with that of Proto-Uralic, but the differences are not very pronounced. In most other relevant aspects, Samoyed is like Proto-Uralic, so that no proper differentiation is possible. The prolative case and the necessitative participle provides a better connection with Samoyed specifically, and the few loanwords, if correctly identified, also do seem to provide a tether to Samoyed. However, in theory they could also stem from a para-Tocharian language that was spoken farther to the north or remained in Siberia when the ancestors of the Tocharians that ended up in the Tarim Basin had already migrated south.

A large number of the innovative typological features of Tocharian can thus be understood as the result of contact with early Uralic/Samoyed. This anchors pre-Proto-Tocharian to the forest area of south Siberia. Especially if the Samoyed connection is taken seriously, pre-Proto-Tocharian should accordingly be moved to the eastern part of south Siberia. Geographically, this matches the Afanasievo-Tocharian hypothesis very well. Chronologically, there are some uncertainties, however. If Proto-Uralic is dated around 2500 BCE, that means that there was no differentiated pre-Proto-Samoyed at the

time that the early Tocharians moved east with the Afanasievo Culture around 3100 BCE. Even if the early Tocharians only moved into the Dzungar basin around 2800, Tocharian-Samoyed contact specifically might have been impossible at such an early date.

However, these time estimates may be subject to change. If Samoyed is more conclusively shown on the basis of phonological and morphological arguments to be the first branch of Uralic to have split off, as per the traditional phylogeny, the dating of Proto-Uralic may shift farther back in time again to accommodate a period of shared Finno-Ugric innovations. A small core of Samoyed specific features that have been adduced seem difficult to entirely ignore in this regard, such as the presence of **o* in **CoCi*, associated with the lack of a merger of PS **tuj* 'fire' and **toj-* 'come', as opposed to **tuli* 'fire' and **tuli-* 'come' in the Finno-Ugric languages. The addition of the suffix **-la* to PU **ńoma* 'hare' could be another Finno-Ugric innovation, although this word is not attested in all branches. This topic requires further research before we can draw our conclusions.

Another possible solution for the chronological uncertainties would be to assume contact between pre-Proto-Tocharian and pre-Proto-Uralic instead. However, without adequate knowledge of pre-Proto-Uralic phonology and morphology, this can hardly be tested concretely at our current state of knowledge. The discussion surrounding the exact positions of the Proto-Uralic and pre-Proto-Uralic homelands would also interact with such a contact scenario.

A Tocharian-Uralic/Samoyed contact hypothesis can thus support the Afanasievo-Tocharian hypothesis and the associated placement of pre-Proto-Tocharian in space, but not necessarily in time. Further research is needed in that regard, and alternative archaeological scenarios with a later migration to the east should be explored if they can at some point be identified in the archaeological and genetic record on the basis of future discoveries.