**Universiteit Leiden**

**The Netherlands**

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

# Chapter 4: Conclusion

In the present thesis, I have discussed the oldest Iranian loanwords in Tocharian as well as possible loanwords from the so-called BMAC language.

## 4.1 Results

My starting point has been that there is no systematic study of the oldest layers of borrowing in Tocharian, despite the obvious relevance these layers have for our understanding of Tocharian as well as Old Iranian linguistic prehistory. Although a number of studies on Iranian and Tocharian contact exist, none systematically presents and analyses the relevant material in full. It thus appeared necessary to approach the oldest layers of borrowing in Tocharian systematically, discussing all of the Old Iranian and possible BMAC loanwords in Tocharian, including those which were, accordingly to my analysis, wrongly attributed to Old Iranian, and some which are doubtful.

My investigation was structured along three research questions, which I repeat here:

1. Are the words belonging to the Old Iranian stage from one single language, as Schmidt thought, or multiple ones, as found elsewhere in the literature? Is there any reason to answer this question clearly?
2. If the Old Iranian layer is one single language, can its features be established? Is it homogeneous? How can we identify the words belonging to this stage and layer?
3. Are there any BMAC loanwords in Tocharian? Did they come directly from a BMAC language, or through, for instance, Iranian?

In the following, I will examine whether these questions have been answered, and how.

1. The first and second questions, concerning the Old Iranian stage of loanwords, which was first described by Schmidt (1985) in a seminal paper, I tried to answer in chapter 2. Schmidt grouped together a number of words which he considered to belong together, namely, words deriving from one stage of one dialect, and most of which I used as a departure point in this research. He proposed a number of sound correspondences such as Old Ira-

nian *a* corresponding to Proto-Tocharian *e*. He also demonstrated that the reflexes of Proto-Indo-European *$\hat{k}$ and *$\hat{g}$ were *$t^s$ and *$d\!\!\!\!-$ in that Iranian language. These sounds go back to Proto-Iranian *ć and *ȷ́, respectively, two sounds which had thus become affricates in Old Steppe Iranian. I re-examined these and other correspondences and tried to establish whether they were regular.

It appeared that the correspondence between Old Iranian *ā and Proto-Tocharian *a and Old Iranian *a and Proto-Tocharian *e, as in TB *kertte* 'sword' corresponding to Proto-Iranian *karta-* 'knife' or *waipecce* 'proper-ty, possession' corresponding to Proto-Iranian *$h\under�{u} a\under}{i}-pa\vartheta\underbar{i}a$-, was essential for the selection of these words. This correspondence is not found in other Iranian loanwords in Tocharian, not even in Proto-Khotanese-Tumshuqese and Pre-Khotanese loanwords, which are very old and likewise belong to the Old Iranian stage (cf. Dragoni 2022). This correspondence, in combination with others, led to the rejection of a number of proposed loanwords from Old Iranian, such as TB *witsako* 'root', which does not fit the phonetic patterns of the Old Iranian layer.

I also attempted to determine the chronological stage and the dialect affil-iation of the Iranian source dialect more exactly. To this end I examined all the phonetic correspondences at hand, but also the morphology of the loan-words from an Iranian perspective, and tried to establish whether the portrait of one single language could be drawn. The answer was positive, as all pho-netic and morphological features appear to form a coherent group, with no need to assume dialect differences or chronological developments within the Old Iranian source. At the same time, it is clear that this language was very archaic in a number of its traits: for instance, it had in all positions preserved *$t^s$ and *$d\!\!\!\!-$ as reflexes of Proto-Indo-Iranian *ć and *ȷ́.

2. The second question was whether the features of the Old Iranian source dialect can be established, if it was one homogeneous variety. As explained above, the correspondences between the Proto-Iranian reconstructed forms and the shape of the loanwords in Tocharian had to be established first. Once this was done, it was easier to establish the features of the Old Iranian source dialect itself.

It is useful to discuss the methodology I used in greater detail. Naturally, language contact is a very wide concept, and it is quite usual that language contact, especially when done due to population contact (versus, for exam-ple, elite domination), involves more than two varieties. If one looks at the

French influence over English, it is quite obvious that multiple dialects and varieties of French influenced English, both diachronically and synchronically. This is also the case for Persian and Arabic, and it is even true with dead languages (for example, Latin words were borrowed into French from Classical Latin, as well as vulgar Latin). It is thus entirely conceivable that a group of Iranian languages, dialects or varieties (with the vagueness these notions carry) went into contact with Proto-Tocharian, and that these yielded various Tocharian words, with different phonetic and semantic features. Not including this possibility would certainly cause a grievous bias to the present work. Nonetheless, starting from this assumption would also be damageable, as, if I started to consider each word for its own features as borrowed "on its own", it could soon be established that Tocharian borrowed from as many Old Iranian dialects as words. If one starts with the assumption that there were multiple sources, but in fact there only was one, it would become difficult to reach that conclusion, while, if one starts with the assumption that there was one source, but in fact there were multiple ones, it is easier to reach that conclusion. Epistemologically, it is necessary to start with the easiest solution: the existence of one single source variety.

I thus departed from the more economical assumption that all the Old Iranian words studied here were borrowed from one single source, while keeping in mind the possibility that it was not so. It was necessary to look at every problem, every unexpected outcome, compare it with the available data we have on Iranian languages, in order to see if it did not warrant a different source. In fact, many times I thought this was the case, and throughout my preliminary research, I often posited different Old Iranian stages or dialects in order to explain this or that word. Often, finding a different example (for instance the initial *ye-* in *yentuke* and *yetse*) solved the issue, this was also the case for TB *ekṣinek** 'dove', although the "other example" was found in Ossetic (see p. 173 for more detail). Some examples, like PT *epiyac* 'memory', were problematic, while being of clear Iranian origin (in this specific case, because the origin of the *-c* is not obvious). In that case, the crucial point was to determine that the issue is not related to an internal Iranian problem: the Proto-Iranian form of this word is *\*abi-i̯āta-*, and there is no reason to interpret this "palatalization" as an Iranian phenomenon. With this methodology, I was able to ascertain the identity of the source of all these words, and thus to attribute their origin to a prehistorical contact

situation relatively limited in place and time (that is, one single language →
one single language contact).

To establish the features of the Old Iranian source dialect I have naturally
focused on innovations rather than on archaisms. I have so identified a num-
ber of sound changes, such as the shortening of *$\bar{a}$ in front of *yod* and *waw*;
the loss of *$h$ in intervocalic and word-initial (and possibly word-final) posi-
tion; the simplification of the cluster *$dm$- to *$m$-; the change of *$rd$ to *$ld$,
etc. Morphological features were also proposed, such as the loss of suffixes
in a number of words, while the meaning of the suffixed form was still found
in the base form of the word. I also observed that there were some important
semantic changes.

A very important feature concerned the stress system of the Old Iranian
source language. Indeed, apart from some suffixes and prefixes bearing fixed
stress (identical to the stress seen in the Vedic cognates of these suffixes and
prefixes for instance), Old Steppe Iranian seems to have had fixed initial
stress, as far as can be judged from the evidence I found. This is not a unique
feature, as it is shared by Ossetic, but it is remarkable.

The discussion on the origin of PT *$ek\d{s}ineke$ 'dove', section 2.6.2.h has
led me to propose a very ancient shared sound change between Old Steppe
Iranian and Ossetic, namely *$an\d{i}V > *\d{i}\d{i}nV$. If I am correct, this would
demonstrate the existence of an Old Steppe Iranian – Ossetic node. Further-
more, Old Steppe Iranian and Ossetic share nine to ten relevant phonetic
isoglosses (see p. 188). But Old Steppe Iranian is not the ancestor of Ossetic,
as the Old Steppe Iranian sound law *$rd > *ld$ shows, but also Old Steppe
Iranian *$m\bar{a}n\acute{i}\d{i}a$- 'servant' vs. Proto-Ossetic *$m\bar{a}ni\d{i}a$- 'husband'. Indeed,
neither 'servant' nor 'husband' can reasonably derive from each other, but
here it is clear that they were specializations of the Old Iranian word
*$m\bar{a}n\acute{i}\d{i}a$- 'the one of the house'. If this conclusion is correct, it  could imply
that no known Iranian language directly descends from Old Steppe Iranian,
but that it shares genetic kinship with Ossetic. The systematic comparison of
both languages could possibly lead to the reconstruction of the Old Steppe
Iranian – Ossetic branch, which could be tentatively be named "Scytho-
Steppic".

Old Steppe Iranians were not an isolated people, they were probably not
the only Iranian people in the region. They were part of a greater continuum
of Iranian-speaking nomads, who are conveniently called "Scythians". Some
of the Scythian peoples were already known in Antiquity, and they were

mentioned for instance by Greek historians. The languages of Scythian peoples are not documented through texts, but are known to us through undirect sources (cf. Pinault 2008a: 106). The Old Steppe Iranian words discussed in this thesis thus provide further undirect evidence of a Scythian language and its lexicon, and possibly of a greater antiquity than the remainder of Scythian languages. Scythian names recorded by Herodotus (died ca. 425 B.C.E.) possess the form *aspa- for 'horse' (cf. Pinault 2008a: 108), as opposed to Old Steppe Iranian *atˢu̯a-, which is clearly more archaic, as it preserves the Old Iranian sequence *-tˢu̯-, which otherwise became *-sp- in the Scythian languages Herodotus recorded words from.

The establishment of the features of Old Steppe Iranian led to the exclusion of some words, like Tocharian B witsako 'root', mentioned above, but also to a number of possible new loanwords, such as Tocharian B ākteke 'wonderful', epastye 'skilfull', epe 'or', eprete 'resolute, steadfast', Tocharian A nātäk 'lord' and others.

3. I have tried to answer my third research question in chapter 3: "Are there any BMAC loanwords in Tocharian? Did they come directly from a BMAC language, or through, for instance, Iranian?" It is thanks to the study of the Old Steppe Iranian vocabulary in Tocharian that this question could be answered. As explained above, the criteria to determine if a word was of Old Steppe Iranian origin or not led to the rejection of some words from that group. Some of these rejected words belong, in my view, to another language: the BMAC language described by Lubotsky (2001), or a variety closely related to that BMAC language. Pinault (2002; 2003; 2006) also published specifically on BMAC loanwords in Tocharian.

These BMAC loanwords also shared features, but, unlike for Old Steppe Iranian, we have no other BMAC-related languages to which we can confront our loanwords in order to reconstruct proto-forms. That is, the only support we may possibly have to show that a Tocharian word can be of BMAC origin is the existence of parallel borrowings into Indo-Iranian, "borrowing cognates". There is thus no way to verify independently how BMAC phonemes are represented in Tocharian, since we know too little about its phonological system. For instance, in Vedic gardabhá- m. 'donkey, ass'

corresponding to Tocharian B *kercapo* /kercə́po/ 'donkey',[167] we can see that Indo-Iranian *\*a* corresponds to *\*e* in the first syllable of the Tocharian word, but to *\*ə* in the second. It is attractive to posit different phonemes for the BMAC source on the basis of Tocharian, but such a contrast cannot be verified independently.

Together with Michaël Peyrot and Federico Dragoni (Peyrot & Dragoni & Bernard, forthc.), I have suggested that a word previously proposed as being of BMAC origin by Pinault (2006: 184-89) was instead a Pre-Khotanese loanword: TB *eñcuwo*, TA *añcu*\* 'iron'. More precisely, it would have been borrowed from the Pre-Khotanese ancestor of Khotanese *hīśśāna-* 'iron', which we reconstructed as *\*henśu̯ani̯a-*, subsequently analyzed as a -*ññe* adjective by Tocharian speakers, leading to the back-formation of the forms TB *eñcuwo*, TA *añcu*\*.

Because of significant and coherent phonetic differences, I have suggested that the BMAC loanwords in Tocharian do not come from exactly the same BMAC variety as those of Indo-Iranian. In view of the presumably large geographic distance, I provisionally termed these two varieties Western BMAC, the source of borrowings into Indo-Iranian, and Eastern BMAC, the source of borrowings into Tocharian. An example is Western BMAC *\*aw* vs. Eastern BMAC *\*ai*, as can be seen in Proto-Tocharian *\*laipəśe* 'jackal' vs. Proto-Indo-Iranian *\*lau̯paća-* 'fox, jackal' (Proto-Iranian *\*rau̯paća-*). Another example is Western BMAC *\*d* or *\*t* vs. Eastern BMAC *\*ǰ* or *\*č* in the examples Vedic *gardabhá-* 'donkey, ass' vs. Tocharian B *kercapo* 'donkey' and PIIr. *\*iṣtika-* 'brick' vs. Tocharian B *iṣcäke* 'clay', Burushaski *diṣcík* 'brick'. There are also significant semantic differences, such as Western BMAC 'brick' for the latter etymon, versus Eastern BMAC 'clay'. These semantic differences coherently point towards a technologically less advanced society for the Eastern BMAC variety than for the Western BMAC variety. The fact that the source varieties appear to be different further suggests that BMAC loanwords in Tocharian were directly borrowed from a BMAC source, not by mediation of Old Steppe Iranian. From an archaeological point of view, it is difficult to imagine that the people of the Bactriana Margiana Archaeological Complex were identical to a people that was much

---

[167] The TB word *kercapo* 'donkey' was previously believed to be related to Vedic *gardabhá-* 'donkey, ass'. I accepted this connection, and explained the relation between these two words with the assumption of a common BMAC source.

further to the north, or to the east (see map p. 249), speaking the exact same language, and this seems to be confirmed by the phonetic correspondences mentioned above.

As a conclusion, it appeared that the BMAC loanwords borrowed into Tocharian are sometimes similar to those borrowed into Iranian from the BMAC language, although less numerous, which is expected given the difference in the size of corpora. I also suggested that BMAC - Tocharian contacts were very early, as the words are less technical than what can be deduced from contact with Iranian and Indic languages. Thus, chronologically, Tocharian speakers would first have come into contact with speakers of the so-called BMAC language, and only later with speakers of Old Steppe Iranian. Furthermore, they show differences both in terms of semantics and phonetics with the BMAC loanwords found in Indo-Iranian languages.

In my investigation of Old Steppe Iranian and BMAC loanwords, I sometimes needed to discuss words that eventually turned out to be of different origin. This is notably the case with borrowings from Khotanese. Apart from TB *eñcuwo*, TA *añcu*\*, mentioned above, I have discussed TB *kamartīke* 'ruler' and TA *kākmärt* 'sovereignty'; TA *kāre* 'sword'; and TB *kāswo* 'skin disease'. All three must derive from Khotanese and are now also included in Dragoni (2022).

## 4.2 The Tocharian way

My inquiry into the oldest layers of loanwords in the Tocharian vocabulary naturally leads to the question of when and where this contact between speakers of Tocharian and both speakers of Old Steppe Iranian and speakers of the BMAC language occurred. Even though the aims of this study were primarily of a linguistic nature, I venture to make a few notes on this issue.

As is well known, the Tocharians were an Indo-European people, and thus did not originate from Western China, or from Central Asia, but rather from the Ukrainian Steppes, where the Yamnaya Culture is strongly associated with speakers of Proto-Indo-European. Thus, we know where the Tocharians started, and that they finally arrived in the Tarim Basin, but many details of their trajectory remain to be clarified.

In line with the focus of my linguistic investigation, I will here concentrate on the question whether the contacts with Old Steppe Iranian and with the so-called BMAC language took place in the Tarim Basin or elsewhere. For the contacts with Old Steppe Iranian ("the archaic stratum"), Peyrot (2018: 272, 280) assumes that the ancestors of the Tocharians had already arrived in the Tarim Basin when these took place. He tentatively locates Old Steppe Iranian north or east of the Tocharian area.
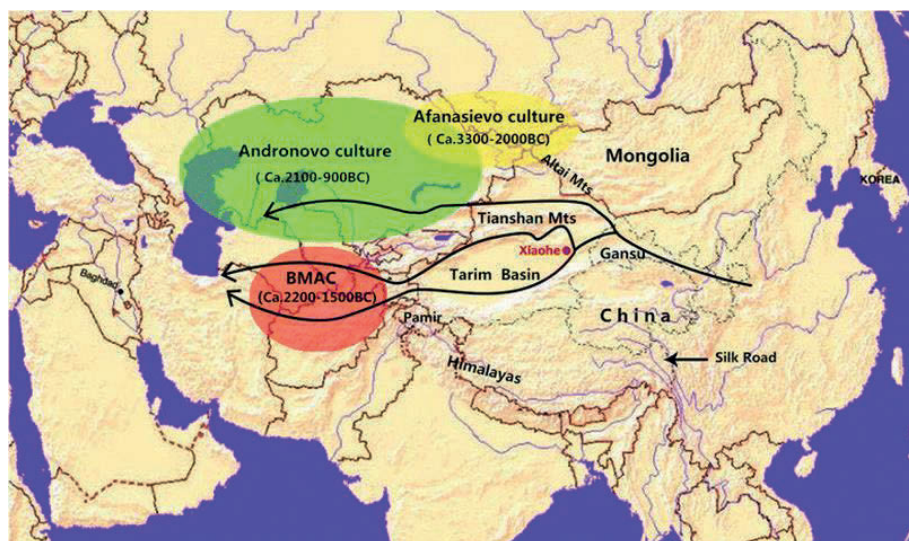


Map from Li & al. (2015): the Andronovo culture is commonly associated with Proto-Indo-Iranian, Afanasievo with Pre-Proto-Tocharian speakers, the Tarim Basin

is where Tocharian manuscripts were found, and where Tocharians and Khotanese (among others) lived. The Old Steppe Iranian people were perhaps living in the part that is to the east of the Andronovo culture circle, and the Afanasievo circle.

Recently it was demonstrated in a genetic study (Zhang & al. 2021) that the famous Bronze Age Tarim Basin mummies, dating back to 2100 - 1700 BCE, were not Indo-European and therefore cannot be identified as Tocharians, as had previously been assumed by many scholars. Indeed, these mummies rather belong to a genetically very different autochthonous population (Zhang & al. 2021: 260). Tocharian presence in the Tarim Basin thus does not need to be as old as these mummies, and there is no longer any need to assume that the Tocharians had already arrived in the Tarim Basin when the contacts with Old Steppe Iranian took place.

In my view, an important argument for dating and locating the contacts is the order in which the contacts with the BMAC language, Old Steppe Iranian and Proto-Khotanese-Tumshuqese, the earliest layer of contacts with Khotanese, took place. Indeed, I believe that a relative order of these contacts can be set up on linguistic grounds.

As I argue, the contacts with Old Steppe Iranian must be dated before the split of Proto-Tocharian in Tocharian A and Tocharian B. This is strongly suggested by the many examples of perfectly regular sound correspondences between Tocharian A and B loanwords from Old Steppe Iranian; by the law of syncope; and in particular, by the connection of this law to the stress system of Old Steppe Iranian; as well as by the reconstruction of Proto-Tocharian verbs based on Old Steppe Iranian nouns. Contacts with Proto-Khotanese-Tumshuqese were either around the split of Proto-Tocharian into the daughter languages or simultaneous with it (Dragoni 2022: 257f.). The later Pre-Khotanese loanwords are to be dated after the split (Dragoni 2022).

On the basis of the BMAC vocabulary, and on the basis of the fact that some BMAC items in Tocharian are reconstructible for Proto-Tocharian, I suggested that BMAC - Tocharian contact preceded the split of Proto-Tocharian in Tocharian A and Tocharian B too. Furthermore, the vocabulary of BMAC borrowings in Tocharian indicates a culture that was not extremely developed, possibly a hunter-gatherer society. We have, notably, a relatively large number of animal names, the word for 'hunter' and the word for 'root'.

It seems the most plausible that Tocharian first came into contact with the so-called BMAC language and borrowed some words for relatively primitive concepts, and then came into contact with Old Steppe Iranian and finally with Proto-Khotanese-Tumshuqese, both of which were apparently technologically more advanced, and were the source of vocabulary referring to more developed concepts.

If we assumed, alternatively, that Tocharian came in contact with Old Steppe Iranian first, and borrowed the words for 'army', 'castle', 'sword', 'lord', 'slave' and even basic words like 'group', why would they later need to borrow basic animal names, and words such as 'root' and 'hunter' from the language of another, much less technologically advanced group? In other words, the speakers of Old Steppe Iranian certainly had words for 'donkey', 'lion', 'root', so why would Tocharians have borrowed these words later from a seemingly less prestigious source language?

Thus, I assume that Tocharian came into contact with the BMAC language first, then with Old Steppe Iranian, and then with Proto-Khotanese-Tumshuqese peoples. It is attractive to identify early speakers of Khotanese with the Aqtala culture (cf. Peyrot 2018: 275f.). The speakers of Old Steppe Iranian are tentatively located to the north or east of the Tocharian area by Peyrot (2018: 280). In any case, it is in my view very unlikely that Old Steppe Iranians were in the Tarim Basin at the time the Tocharians were there or before. On the basis of the semantics of the borrowed vocabulary, I hypothesize that the Old Steppe Iranians probably conquered or subdued the Tocharians in some way. If the Old Steppe Iranians were such an important culture, how should they have disappeared from the Tarim basin without leaving any trace?

With the above considerations in mind, one can suggest the following scenario for the arrival of Tocharians in the Tarim Basin: Tocharians, arriving to eastern Central Asia through South Siberia, first encountered the so-called BMAC speakers possibly in Dzhungaria. Tocharians learned from these BMAC speakers about a number of animals in the region, about plants, about food-gathering techniques.

One of the conclusions of this thesis (section 3.4.3) was that the BMAC language in contact with Tocharian was not identical to the one Indo-Iranians were in contact with. This can be shown by constant phonetic differences in the adaptation of words, but also by the differences in meanings found in BMAC words in Indo-Iranian and in Tocharian. As a consequence,

Tocharians did not meet the BMAC people in the same place as the Indo-Iranians did, but more to the east, probably when they migrated south from Siberia into Dzhungaria.

In the late Bronze Age, that is, 1500 to 800 BCE, we find archaeological cultures distinguished by roller pottery in the Steppes. These cultures had acquired horse riding techniques around the 16th century BCE (cf. Parpola 2022: 48-49). Parpola (2022: 49) further adds that these horse riding peoples were the "immediate ancestors of the Iranian-speaking mounted horsemen known as Scythians, Sarmatians and Sakas […]". In all likelihood, the Old Steppe Iranians were part of this roller pottery culture, as were the ancestors of the Ossetes, who are most closely related to the Old Steppe Iranians from a linguistic point of view (see for example p. 188). We also know that a drought occured between 1200 and 800 BCE on the Eurasian Steppes where the Old Steppe Iranians likely were (cf. van Geel et al. 2004). This drought perhaps pushed the Old Steppe Iranians (and possibly other Iranian peoples) into the periphery of the steppe, so that they reached Dzhungaria, where they encountered the Tocharians (cf. Peyrot 2022). This would be in agreement with the archaeological data presented by Parpola (2022).[168]

The Old Steppe Iranians very likely conquered or subdued the Tocharians in some way, as the type of vocabulary (military, social, and even the grammatical loanword *epe*) seems to suggest. Although there probably was at least a part of the population which was bilingual Tocharian - Old Steppe Iranian, as is suggested from the ease with which they integrated the loanwords, the influence of the Old Steppe Iranian language seems to have been rather limited in time. Notably, I have observed no heterogeneity in this layer of vocabulary, which I take as an indication that the borrowings occurred in a rather short period. Furthermore, although a word such as TB *epe* 'or' shows the strong cultural domination of Old Steppe Iranians on Tocharian (as grammatical borrowings usually do), the semantic fields of Old Steppe Iranian loanwords are relatively limited. For instance, we have no evidence for the borrowing of words for food or animals, besides *etswe* 'mule', which might well be expected if the contacts extended over a longer period.

---

[168] It is possible that it is during this migration that the Old Steppe Iranians, the ancestors of the Sogdians, of the Ossetes, and of various other Iranian peoples, borrowed the word *paratu- 'axe', possibly from Eastern BMAC people.

After the contacts with the Old Steppe Iranians, the Tocharians entered the Tarim Basin and encountered early speakers of Khotanese. It seems that the split of Proto-Tocharian into the two daughter languages Tocharian A and Tocharian B took place only then, after their arrival in the Tarim Basin. The contacts with Khotanese and its prestage lasted for centuries, way into the historical period, since we find in Tocharian traces of multiple stages of these languages: Proto-Khotanese-Tumshuqese, Pre-Khotanese, Old Khotanese, Late Khotanese and Tumshuqese (Dragoni 2022).

I should stress again that this scenario is principally informed by my linguistic findings, combined with those of Dragoni (2022). Obviously, additional archaeological and genetic evidence is needed to see if it can be further confirmed or rather needs to be revised. However, at this point it is as coherent from the linguistic point of view as I could make it. In my view, a task for future research is to search for possible descendants of the Old Steppe Iranians, which will probably yield more insights on the prehistory of this hitherto unknown ancient Iranian population. It would also be profitable to examine in detail the specific links this language has with other Iranian languages, especially Ossetic.