Artyom Badeev
National Research University Higher School of Economics, Moscow / Institute of Linguistics of the Russian Academy of Sciences, Moscow; badeev@iling-ran.ru

Tarim-type demonstrative systems of Khotanese and Tocharian: towards the reconstruction of language contact in the Tarim Basin[1]

This article discusses the formation patterns of Khotanese (< Eastern Iranian) demonstrative pronouns that are unparalleled in other Iranian languages. It covers the deictic meanings, etymology, and evolutionary patterns of Khotanese demonstrative stems. The deviant features of Khotanese find their parallels in Tocharian. These include the functional distribution of demonstrative pronouns in the system as anaphoric, proximal and distal series; the formation of all three series from two suppletive stems; the marking of deictic contrast by means of suffixal indices; as well as the morphological makeup of individual demonstrative pronouns. By comparing Khotanese innovations with those of Tocharian, these features can be identified as markers of the Tarim-type demonstrative system. It is argued that centuries-long language contact in the Tarim Basin may have led Khotanese and Tocharian to develop a shared typological profile, absent in the neighboring languages.

Keywords: Khotanese language; Tocharian languages; demonstrative systems; language contact; Uralic languages; Samoyedic languages.

1. Introduction

Compared to the rest of Iranian languages, Khotanese presents a unique example of a Middle Iranian language which exhibits a dedicated anaphoric demonstrative pronoun. Furthermore, both Khotanese proximal and distal demonstrative pronouns show innovative formation unparalleled in Iranian languages. In contrast, Khotanese demonstrative adverbs preserve the reflexes of Proto-Iranian demonstrative stems. The areal perspective provided in this article consists in comparing Khotanese innovations with those of the neighboring languages. It is concluded that Tocharian shows parallels to the deviant features of Khotanese, both in terms of the structure of the system as a whole and the makeup of individual elements. Over centuries of language contact with Tocharian, the Khotanese demonstrative pronouns were restructured into a completely new type, which is labeled the Tarim-type demonstrative system. The characteristic features of Tarim-type demonstrative pronouns probably have a substrate origin similarly to the other deviant features of Tocharian resulting from prehistoric contact with Proto-Samoyedic.

The bulk of the paper will be organized as follows:

— in section 2, I reconstruct the Proto-Iranian demonstrative stems and compare various reflexes of demonstrative pronouns in Iranian languages, suggesting the typical Iranian pattern of evolution of demonstrative stems. I then provide etymologies of Khotanese demonstrative stems and argue why Khotanese demonstrative pronouns do not conform to the typical Iranian type;

[1] This research has been supported by the Russian Science Foundation, project No. 25-78-10222 “Between morphology and syntax: challenging phenomena across frameworks”. I thank the anonymous reviewers for their valuable comments. All remaining errors are mine alone.

— in section 3, I aim to demonstrate the similarities between the demonstrative pronouns of Khotanese and Tocharian. This provides more evidence of language contact between Khotanese and the Tocharian languages;

— in section 4, I discuss possible diachronic scenarios for common innovations of Tarim-type demonstrative systems, emphasizing the hypothesis of the Uralic substrate influence. This is followed by the concluding section 5.

2. The history of Khotanese demonstrative stems

2.1 Proto-Iranian demonstrative stems and series

Based on the data of Avestan and Old Persian, one can reconstruct at least four series of demonstrative pronouns in Proto-Iranian (see Table 1). These four series are deduced on the basis of the deictic and textual uses in Old Iranian languages and backed up by the data from Middle and New Iranian languages. Three series are distinguished on the proximity scale: proximal, medial and distal. Apart from this, one series is reserved for non-exophoric use, primarily functioning as anaphoric and third person pronouns. Since the paradigms are often suppletive, the number of stems in Proto- and Old Iranian is considerably larger than that of the series.

Proximal	Medial	Distal	Anaphoric
**a-*, **a̯i-* : **i̯i-*, **ima-*, **anā*	**aiša-*, **aita-*	**au̯a-*, **hāu-*	**ha-*, **ta-*; possibly, **a-*

Table 1. Four series of Proto-Iranian demonstrative pronouns (Badeev 2024a: 97).

The Old Iranian demonstrative systems are characterized by several patterns of suppletion. First, there is a suppletion pattern in Proto-Indo-European where a suppletive demonstrative stem is deployed in the nom.sg.m. and nom.sg.f. Besides that, two suppletive patterns in the formation of demonstrative pronouns go back to the Proto-Indo-Iranian level: first, when a suppletive stem is deployed in the indirect cases (other than the nominative-accusative); second, when a suppletive stem is deployed in the instrumental singular masculine. All three patterns are found in the Avestan proximal demonstrative pronoun *aē-* : *ī-*/*ima-*/*a-*/*ana* (< PIr. **a-*, **a̯i-* : **i̯i-*, **ima-*, **anā*), see Table 2. In this instance, the suppletive stem of the nom.sg.m.-f. consists of the ablaut variants *aē-* : *ī-*; in nom.-acc. except for nom.sg.m.-f., *ima-* is found; all the indirect cases are formed on the basis of the *a-* stem, except for instr.sg.m. *ana*.

Case	Singular			Plural			Dual		
	Masc.	Neut.	Fem.	Masc.	Neut.	Fem.	Masc.	Neut.	Fem.
nom.	*aēm*	*imaṯ*	*īm*	*imą, ime*	*ima, imå*	*imå*	*ima*	—	—
acc.	*imәm*	*imaṯ*	*imąm*	*imą, ime*	*ima, imå*	*imå*	—	—	—
gen.	*ahe, aiŋhe*		*aiŋhå, aiŋhåsca*	*aēšąm*		*åŋhąm*	*aiiå*		—
dat.	*ahmāi*		*aiŋhāi*	*aēibiiō*		*ābiiō*	*ābiia*		*ābiiā*
abl.	*ahmāṯ*		*aiŋhāt*	—		—	—		—
instr.	*ana*		*aiia*	*aēibiš*		—	—		—
loc.	*ahmi, ahmiia*		*aiŋhe*	*aēšu, aēšuua*		*āhuua*	—		—

Table 2. Avestan proximal demonstrative *aē-* : *ī-*/*ima-*/*a-*/*ana* (Badeev 2024: 99).

Considering further etymologies, **a-* comes from PIE **e-* or **h₁e-*. In turn, **a̯i̯-* : **i̯i̯-* follows the Proto-Indo-European alternation between **h₁e-* and **i-*, see the paradigm reconstructed in Beekes 2011: 229. The morpheme **ima-* is formed in Proto-Indo-Iranian on the basis of PIE **i-*stem with the addition of the **-ma-* element (Dunkel 2014: 370). Last but not least, **anā* < PIE **e-nó-h₁* is formed by adding the deictic particle **-no-* to the demonstrative pronoun **e-*. The Old Indic proximal instr.sg. *enā* is formed in a similar way (Kupfer 2002: 49). It is cognate with OCS *onъ* 'that', Lith. *añs*, *anás* 'that', although the details are unclear. The reconstruction of the Old Church Slavonic and Lithuanian pronouns as PIE **áno-* given in Dunkel 2014: 55 is disputed because of the lack of exact cognates in other Indo-European languages, see Peyrot 2017. In Iranian languages it is only found in the instr.sg.m. It is probably connected to the Old Indic instr.sg.m ending of the thematic declension (as well as of the demonstrative pronouns) *-ena*, which displaced the old instr.sg.m *-ā*. The Khotanese instr.sg. ending *-äna* (*-ina*) has the same origin. It was argued that this development in Indo-Iranian languages is due to the expansion of the ending of the *n*-declension to the thematic paradigm, cf. Khot. nom.sg. *tcei'mä* : instr.sg. *tcei'mäna* 'eye' (cf. Av. *čašman-*); nom.sg. *balysä* : instr.sg. *balysäna* 'Buddha' (Emmerick 1968a: 258).

The other Proto-Iranian demonstrative pronouns follow the suppletion and alternation patterns mentioned above. Thus, **au̯a-* and **hāu-* are both distal deixis forms with **hāu-* occurring only in nom.sg.m. and nom.sg.f., and **au̯a-* occurring everywhere else. Both contain the Proto-Indo-European deictic particle **u*. Reflexes of PIE **u* are also involved in the formation of demonstrative pronouns in Proto-Slavic and Old Indic languages, cf. OInd *asaú*, OCS *ovъ*. Considering the above data, one can suggest that **au̯a-* is formed on the basis of PIE **a-* and **u* stems, while **hāu-* involves PIE **só-* in addition to **u.* Apart from **hāu-*, PIE **só-* is involved in the formation of two more Proto-Iranian demonstrative pronouns: medial **aiša-*/**aita-* and anaphoric **ha-*/**ta-*. The latter is basically the reflex of PIE **só-*/**tó-*. The former, however, is the innovative Proto-Indo-Iranian formation reflecting the combination of PIE **a̯i̯-* and **só-*/**tó-*. The most complete paradigm of **ha-*/**ta-* is found in Avestan. However, already in Old Avestan the paradigm is defective: only nominative, accusative and instr.sg.m. are formed from these stems. The other cases are formed by the reflexes of the **a-* stem, with a similar distribution as in the paradigm of the proximal demonstrative pronoun *aēš* : *ī-*/*ima-*/*a-*/*ana*. Thus, the anaphoric demonstrative pronoun in Proto-Iranian probably incorporated the **a-* stem in its paradigm. For a further discussion of the morphological and semantic properties of Avestan, Old Persian and Proto-Iranian demonstrative pronouns see Badeev 2024a.

2.2 Proto-Iranian demonstrative stems reflected in Middle and New Iranian languages

By the Middle Iranian and New Iranian periods, the demonstrative systems of Iranian languages underwent simplification, and different innovations are found in individual languages. For instance, in the Pamir languages except Yazghulami, a tripartite system prevailed, probably due to convergence, see Wendtland 2009: 180; Edelman 1990: 222–228. Cf. Shugh. (1) *yam* (obl. *mi*, *mam*) < **ima-*; (2) *yid* (obl. *di*, *dam*) < **aita-*; (3) *yu*, *yā* (obl. *wi*, *wam*) < **au̯a-*. Thus, the reflexes of **ima-*, **aita-*, **au̯a-* constitute a system in which (1) proximal, (2) medial and (3) distal deixis are distinguished. Tripartite systems are also found in such Middle Iranian languages as Sogdian, Khotanese and Parthian (Sims-Williams 1994; Wendtland 2009: 181; Durkin-Meisterernst 2014: 282). In the case of Sogdian, the Proto-Iranian suppletive pattern is preserved, with the second series resulting from the contamination of two separate Proto-Iranian demonstrative pronouns **aiša-* and **ta-*. Cf. Sogd. (1) *y-* (obl. *m-*) < **a̯i̯-*/**ima-*; (2) *š-* (obl. *t-*)

< *aiša-/*ta-; (3) *x-* (obl. *w-*) < *hau-/*au̯a-, which denote, respectively, (1) proximal, (2) medial and (3) distal deixis (Sims-Williams 1994). In Parthian, the suppletive inflection of Proto-Iranian is lost, and the tripartite system of deixis prevailed, cf. Parth. (1) *im* < *ima-; (2) *ēd* < *aita-, (3) *hō* < *hau- (Durkin-Meisterernst 2014: 279–290). The Pashto system shows more innovations, since the deictic functions of the Pashto proximal *daɣ-* and medial *(h)aɣ-* do not correspond to Proto-Iranian, cf. Pashto (1) *daɣ-* < *aita-ga-; (2) *(h)aɣ-* < *a-ga-; (3) *(h)uɣ-*, *hoɣ-* < *hāu-ga-* or *au̯a-ga- (Edelman 1990: 220–221; Boyle David 2014: 172–176).

The reflection of Proto-Iranian stems in languages with bipartite systems of demonstrative pronouns is not uniform either. Yazghulami differs from other Pamir languages by the loss of the third series, while the origin of the first series is being debated, cf. Yazgh. proximal (1) *yu(k)* < *a̯iam/*i̯iam or *au̯a-; distal (2) *du(k)* < *aita- (for the discussion see Edelman 1990: 225–227; Wendtland 2009: 180). In Ossetic, Iron and Digor proximal (1) *a-* < *a-, and Iron distal (2) *wy-* < *au̯a- are found. The etymology of Digor distal nom. *je* (obl. *wo-* < *au̯a-) is uncertain, it is probably derived from *hāu- with the genitive case marker, see the discussion in (Edelman 1990: 216–217; Badeev 2024b). The bipartite system in Kurdish languages goes back to the Proto-Iranian as follows: (1) Kurm. *ev*, Sor. *em* < *ima-; (2) Kurm., Sor. *ew* < *au̯a- (Chyet 2003: 180, 181). A distinct set of demonstrative pronouns characterized by the final *-n* is found in Middle and New Persian, as well as in various other Iranian languages, such as Khwarezmian, Gilaki and Mazanderani. Cf. MP *ʿyn* /īn/, perhaps, /ēn/ (NP *īn*) 'this', MP *hʾn* /hān/, *ʾn'* /ān/ (NP *ān*) 'that'; Khw. *ny*, *nyn* /nī, nīn/ 'this', *nʾn* /nān/ 'that'; Gil. *(h)an* 'this', *(h)un* 'that'; Maz. *in(to)* 'this', *un(to)* 'that' (Bogolyubov 2012; Rastorgueva, Edelman 1982: 511–515). Probably, the suffix *-na* is involved in the formation of the above demonstrative pronouns, while the corresponding Old Iranian stems of the proximal and distal demonstrative pronouns are preserved, cf. MP *īn* < *iyam-na* (OP prox.nom.sg. *iyam*); MP *(h)ān* < *hāu-na* (OP dist.nom.sg. *hāu*) (Klingenschmitt 2005: 112). However, if the loss of final *-m* in case endings occurs as early as in Late Old Persian (Fattori 2024: 253), the more probable etymon *iya-na should be reconstructed. Klingenschmitt (2005: 111) also mentions Sogdian forms that are used both with and without the suffix *-na*, cf. Sogd. dist.gen.sg.m. *ʾwy(n)* < *au̯ahi̯a(-na)*, prox.gen.sg.m. *ʾwy(n)* < *imahi̯a(-na)*. Considering the above, Middle Persian also has demonstrative pronouns formed from other Old Persian stems, cf. MP *im* < OP prox.acc.sg.m. *imam*; *ōy* < OP dist.acc.sg.m. *avahyā*; *ēd* < OP nom.-acc.sg.n. **aita* (discourse deictic). However, these forms have no exophoric uses in Middle Persian. Thus, *im* is used only in adverbial constructions such as *im rōz* 'today'; *ōy* is primarily used as the third person pronoun, and *ēd* as a discourse-structuring device (Durkin-Meisterernst 2014: 284), in fact, similarly to the use of OP **aita-* (Badeev 2024: 89). The discussion on the etymology of these pronouns is far from complete. However, it can be noted that the Middle and New Persian bipartite system represents a specific innovation. It is probable that demonstrative pronouns of this type in other Iranian languages may be the result of language contact and significant Persian influence. A summary of the reflexes of Proto-Iranian demonstrative pronouns can be seen in Table 3.

This sample of languages gives an idea of Proto-Iranian demonstrative pronouns, pronominal stems and their typical reflexes in various Iranian languages. Thus, the typical composition of the system of demonstrative pronouns in Iranian languages includes the following Proto-Iranian stems: proximal (1) *a̯i- : *i̯i-, *ima-; medial (2) *aiša-, *aita- (and perhaps *ta-); distal (3) *au̯a-, *hāu-. This tripartite contrast may be simplified to bipartite with one of the three elements being lost. In this section, I intentionally avoided mentioning Khotanese. In the next subsections I describe its system of demonstrative pronouns and show why it does not correspond to the typical Iranian profile.

Proto-Iranian	*a-, *a̯i- : *i̯i-, *ima-, *anā	*aiša-, *aita-	*au̯a-, *hāu-	*ha-, *ta-
Avestan	a-/aē- : ī-/ima-/ana	aēša-/aēta-	auua-/hāu	ha-/ta-
Old Persian	a-/iya-/ima-/anā	aita-	ava-/hau̯	
Middle Persian	īn, im	ēd	(h)ān, ōy	
Parthian	im	ēd	hō	
Khwarezmian	ny, nyn	nyš	nʾn, nʾw(y)r	
Sogdian	y- (obl. m-)	š-	x- (obl. w-)	t-, oblique for š-
New Persian	īn		ān	
Kurmanji	ev		ew	
Shughni	yam (obl. mi, mam)	yid (obl. di, dam)	yu, yā (obl. wi, wam)	
Yazghulami	yu(k)	du(k)		
Pashto	(h)aɣ-	daɣ-	(h)uɣ-, hoɣ-	
Ossetic	a-		wy-/je (obl. wo-)	

Table 3. Reflexes of Proto-Iranian demonstrative pronouns in a sample of Iranian languages.

2.3 Khotanese demonstrative stems compared to the typical Iranian patterns

Three series of demonstrative pronouns are distinguished in Khotanese. Khot. *ṣäta-/ttäta-* and *ṣāra-/ttāra-* are labeled as the "near" and the "remote" in the secondary literature (Sims-Williams 1994: 49; Emmerick 2009: 387). For the contrast between the two, cf. Khot. *ne* ***ttuto*** *tcalco u ne* ***ttāro*** *ne myaño* 'Not this boundary, and not that one, and not the middle one' (Skjærvø 2002: 247). According to the typological terminology adopted in this article, Khot. *ṣäta-/ttäta-* and *ṣāra-/ttāra-* would be referred to as proximal and distal demonstrative pronouns. The demonstrative pronoun Khot. *ṣa-/tta-* is referred to as "neutral" in the literature (see Sims-Williams 1994: 49). However, I argue in favor of the more accurate label "anaphoric". Khot. *ṣa-/tta-* is typically used in anaphoric contexts, where the reference is made to a noun phrase in the preceding discourse, cf. *gyastūñu āysanu ṣkodde* ***ttye*** *bendä balysä niga'lstä* 'He created the divine seat. The Buddha sat on **it**' (Emmerick 1968b: 26). Thus, while Khot. *ṣa-/tta-* is used for anaphoric (and third person pronominal) reference, proximal and distal demonstrative pronouns, namely, Khot. *ṣäta-/ttäta-* and *ṣāra-/ttāra-* express the deictic contrast between near and remote objects. For the complete paradigms of Khotanese demonstrative pronouns, see Tables 4–6.

All three series of Khotanese demonstrative pronouns, respectively, Khot. anaphoric *ṣa-/tta-*, proximal *ṣäta-/ttäta-*, and distal *ṣāra-/ttāra-*, are formed on the basis of two Proto-Iranian stems, **aiša-* and **ta-*. N. Sims-Williams analyzes Khot. *ṣa-/tta-* and Sogd. *š-/t-* demonstrative

Case	Singular		Plural	
	Masculine	Feminine	Masculine	Feminine
nom.	*ṣä*	*ṣa*	*ttä*	*tte, ttä*
acc.	*ttu*	*tto*	*ttä*	*tte, ttä*
gen.-dat.	*ttye*	*ttye*	*ttānu*	*ttānu*
instr.-abl.	*ttäna*	*ttäñe*	*ttyau*	*ttyau*
loc.	*ttäña*	*ttäña*	*ttuvo'*	*ttuvo'*

Table 4. Khotanese anaphoric demonstrative pronoun (Emmerick 2024: 301).

Case	Singular			
	Masculine		Feminine	
	Uncontracted	Contracted	Uncontracted	Contracted
nom.	*ṣätä*	*ṣei(')*, *ṣä'*	—	*ṣā(')*
acc.	*ttutu*	*ttū*	*ttuto*, *ttuvo*	*tvo*
gen.-dat.	*ttät(ä)ye*	*ttīye*	—	*ttīye*
instr.-abl.	*ttätäna*	—	*ttätäñe*	—
loc.	*ttätäña*	—	*ttätäña*	*ttīña*
	Plural			
nom.-acc.	*ttätä*	—	*ttäte*	—
gen.-dat.	*ttätānu*	*ttyānu*	—	*ttyānu*
instr.-abl.	—	*ttiyyo*	*ttätäyyo*	—
loc.	*ttävo'*, *ttäto*	*tvo*	*ttäto*	*tvo*

Table 5. Khotanese proximal demonstrative pronoun (Emmerick 2024: 301).

Case	Singular		Plural	
	Masculine	Feminine	Masculine	Feminine
nom.	*ṣārä*	*ṣāra*	*ttārä*	*ttāre*
acc.	*ttāru*	*ttāro*	*ttārä*	*ttāre*
gen.-dat.	*ttārye*	*ttārye*	*ttārāṇu*	—
instr.-abl.	*ttārna*	—	*ttāryau*	—
loc.	—	*ttārña*	—	—

Table 6. Khotanese distal demonstrative pronoun (Emmerick 2024: 302).

stems as an innovative contamination of Proto-Iranian **a̯iša-* and **ta-* (Sims-Williams 1994). There is a formal distinction between Khot. *t* (earlier [d], later [ʔ]) < PIr. **VtV* and Khot. *tt* [t] < **tV* (Edelman 1986: 166; Hitch 2016: 94). Thus, Khot. *tta-* is the regular reflex of PIr. **ta-*, rather than **aita-*. However, this argument does not hold for Sogdian, where both PIr. **VtV* and **tV* yield Sogd. *t* (Edelman 1986: 165–166). Therefore, Sogd. *t-* can be considered the reflex of both PIr. **ta-* and **aita-*. The main argument of Sims-Williams in favor of deriving Sogd. *t-* from PIr. **ta-* rather than **aita-* is the use of two distinct Aramaeograms[2] for Sogdian suppletive *š-* and *t*-stems (Sims-Williams 1994: 50). For each other suppletive stem (*y-*/*m-*, *x-*/*w-*) only one Aramaeogram is used. Thus, it is assumed that at the time of the earliest written fixations, prior to forming the innovative paradigm, *š-* and *t-* were recognized as two distinct pronouns. In Sogdian, *š-*/*t-* has medial deictic and anaphoric uses, which also corresponds to the semantics of **a̯iša-* and **ta-* in Proto-Iranian. In Khotanese no distinctive Aramaeograms are used, but the argument of phonological correspondences still holds.

Two features make the Khotanese system unique in comparison with other Iranian languages.[3] First, the system of three series in Khotanese is based on historical modifications of

[2] Regarding the use of the term Aramaeogram, see the discussion in Kudrinski & Yakubovich 2016: 54.

[3] I intentionally avoid discussing Tumshuqese, a language closely related to Khotanese, in this article. Tumshuqese demonstratives are far less studied, and the data are probably not enough to speak of its demonstrative system with certainty. However, there is evidence that the Tumshuqese demonstrative pronouns differ from Kho-

the same two Proto-Iranian suppletive stems. While many Iranian languages have tripartite systems, these are always based on at least three different Proto-Iranian stems (six in languages that preserve the old suppletion pattern, like Sogdian). Furthermore, the makeup of the system is drastically different: Iranian languages typically possess the typologically common proximal-medial-distal or proximal-distal system, while Khotanese has proximal-distal-anaphoric. To the best of my knowledge, Avestan is the only Iranian language, except for Khotanese, to possess a specific series of anaphoric demonstratives, cf. Av. *ha-*/*ta-*/*a-*. However, Avestan has a four-series contrast with *aēša-*/*aēta-* for medial deixis, like in typical Iranian tripartite systems. On top of this, the *a-* stem is used in the oblique cases of the proximal demonstrative pronoun in Avestan. Thus, only nom., acc. and instr.sg.m. are formed from the stems of the dedicated anaphoric pronoun Av. *ha-*, *ta-*. The details of the etymology of the Khotanese proximal and distal demonstrative pronouns are discussed in subsections 2.3.1 and 2.3.2.

2.3.1. Proximal demonstrative stems

Khot. *ṣäta-*/*ttäta-* are the two stems of the proximal demonstrative pronoun. Khotanese grammars explain its form as the result of reduplication of the anaphoric demonstrative stem *ṣa-*/*tta-* (Konow 1941: 48; Emmerick 2024: 301). Since reduplicated demonstrative series are not found in the other Iranian languages, the origin of such reduplication still requires a proper explanation. To begin with, most case forms of the proximal demonstrative pronoun are based on the stem *ttäta-*, with *ä* in the first syllable. The only exception is the stem of accusative singular *ttuta-* in both genders. Therefore, it is possible to suggest two paths of forming the proximal demonstrative pronoun. One involves partial reduplication, i.e. exact reduplication of the first syllable of the stem, as we observe in *ttutu* (anaphoric *ttu*), *ttätäña* (anaphoric *ttäña*), but not in *ttätānu* (anaphoric *ttānu*). With respect to gen.sg. and instr.pl. of the anaphoric demonstrative pronoun, one may assume the contraction *ttye* < **ttäye*, *ttyau* < **ttäyau*; respectively, proximal *ttätäye* and *ttätäyyo*. However, *-ä-* is not found in anaphoric *tte* < **tāh*, *ttānu* < **tānām*, *ttuvo*' < **taišu*. Thus, if we accept the scenario with initial syllable reduplication, *-ä-* in *ttäte*[4], *ttätānu*, *ttävo*' remains unexplained. It is possible that the paradigm of the proximal demonstrative pronoun was further leveled by analogy with the most frequent stem *ttäta-*. Then we have to assume that the stem of acc.sg. still retains its original vowel.

The second possibility involves the formation of the proximal demonstrative pronoun by means of compounding of two deictic words via the so-called "deictic reduplication", see de Vaan (2015: 5–6). Compound demonstratives of this kind are found in several Indo-European languages and are formed by compounding either a demonstrative and a deictic particle; or two demonstrative stems. It is through the latter process that the Proto-Iranian demonstrative pronoun **aiša-*/**aita-* < PIE **e̯i-* + **só-*/**e̯i-* + **tó-* is formed (Dunkel 2014). Thus it is possible to suggest that Khotanese proximal demonstrative pronoun was formed by compounding two stems, that is, PIr. **ta-* + **ta-*. Like in PIr. **aiša-*/**aita-*, the first stem of the Khotanese compound demonstrative stem *ttäta-* is fossilized and indeclinable. Assuming that the proximal demonstrative was formed by the second path, i.e. the first stem in the compound is fossilized and the second one is inflected, the stem *ttäta-* is primary, and only acc.sg. *ttuta-* appears to be innovative.

tanese. The pronoun *mā* (nom.sg. *mā*, acc.sg. *maru*, *marä*; gen.sg. *marā*, *marye*; nom.-acc.pl. *mare*; gen.pl. *maranu*) < PIr. **ima-(θ)ra* is most frequently attested in Tumshuqese documents, see Ogihara & Chao-jung 2017.

[4] It should be noted that *e* and *ä* are phonetically similar vowels. The alternations between them are frequent in Khotanese, but are almost exclusively restricted to the word-final position (Hitch 2016: 127).

The second explanation is supported by comparative evidence from a number of Indo-European languages, including Indo-Iranian. Still, the nom.sg. *ṣätä* /ʐəʔə/ appears to be problematic for both strategies. I would like to argue that this form was modified by analogy with the *tta-* forms. This development occurred after the loss of intervocalic *ṣ* /ʐ/ in Khotanese. The lost *ṣ* was indicated by the subscript hook (transcribed as ') (Hitch 2016: 71). It is fortunate for this reconstruction that Khotanese preserves several contracted parallel forms of the proximal demonstrative pronoun. These contracted forms, Khot. nom.sg.m. *ṣei*', nom.sg.f. *ṣā*('), also loc.pl. *ttävo*' < **tataišu* are written with an apostrophe. Cf. Av. loc.pl. *taēšu*, Khot. *gguvo*' < **gauša-* 'ear' (Emmerick 1968a: 263). Considering the evidence above, the shortened forms of nom.sg.m. *ṣei*' and nom.sg.f. *ṣā(')* speak in favor of reconstructing **ṣäṣä* and **ṣäṣa*.

Some forms are called "double reduplications" in Emmerick (2009). These are nom.sg.m. *ṣeiṣei*; nom.sg.f. *ṣāṣā*, *ṣāṣa*. They originate as full reduplication of the contracted forms *ṣei*' and *ṣā*('), respectively. A similar pattern of full reduplication is found in Sanskrit and is considered to be a Sanskrit innovation, cf. Skt. *sá-sa*, *ayám-ayam* (de Vaan 2015: 6–8). In Sanskrit, these forms denote delimitation or specification. The influence of Sanskrit as a literary language on Khotanese was undoubtedly considerable, not only through adoption of the script, but also through many lexical borrowings into Khotanese from Sanskrit directly or through Gāndhārī transmission, see the analysis of Indic loanwords in Khotanese by D. Loukota (2023). Plenty of Khotanese texts are translations or adaptations made directly from Sanskrit originals, see Emmerick (1992: 17). It is therefore possible that the full reduplication of demonstrative pronouns in Khotanese is an Indic influence, however, further investigation is needed.

2.3.2. Distal demonstrative stems

Khot. *ṣāra-*/*ttāra-* are the distal demonstrative stems. On its origin, Emmerick (2024: 207) claimed that its suffixal element is the PIr. contrastive suffix **-tara-*, thus Khot. *ṣāra-* < PIr. **aiša-tara-*. The original basic function of **-tara-* as a contrastive suffix is seen in its use to indicate deixis with pronominal stems. In this function it is unproductive. However, this etymology is doubtful from both formal and semantic sides. The PIr. /t/ in intervocalic position was weakened to a glottal stop /ʔ/ in Khotanese, but did not completely disappear, at least in Canonical Old Khotanese (Hitch 2016, 94). Moreover, it would be improbable on semantic grounds for PIr. **-tara-* to be added to a pronominal stem: in Proto-Iranian its use was probably limited to the formation of degrees of comparison of nouns and adjectives[5], and in Khotanese it is still productive in this function, cf. Khot. *hastara-* 'better' (Bailey 1979: 475).

I suggest a more appropriate etymon, which is the Proto-Iranian suffix of demonstrative adverbs **-θra*, thus Khot. *ṣāra-* < **aiša-θra*, *ttāra-* < **ta-θra*, cf. the reflexes of PIr. **-θra* in Pashto *dar* 'here (to the speaker)', Khwar *ʾydrθ* 'there' from the PIr. stem **aita-*; Av. *auuaθra* 'there', Sogd. *ʾwrδ* 'there', Munji *ūra*, *wūra* 'there' from the PIr. stem **aua-*, Av. *iθra* 'here' from the stem **i-*, etc. (Rastorgueva & Edelman 2000: 132, 278; Abaev 1989: 96). The proposed etymons **aiša-θra*, **ta-θra* fit better from the formal perspective. A similar reflex of the PIr. cluster **-θr-* can be found in Khot. *pūra-* 'son' < **puθra-*, *tcīra* 'face' < **čiθra*, thus, in addition to the loss of **-θ-*, the length of the preceding vowel is affected. The next question that can be raised here is how the indeclinable, originally demonstrative adverb **aiša-θra* could become the distal deictic pronoun. To answer this question, one should consider the system of pronominal declension in Khotanese and Proto-Iranian.

[5] Cf., however, Oss *dær* enclitic 'too; also' derived from PIr. **-tara-* (Abaev 1958: 356).

The Proto-Iranian system of pronominal declension historically had its own case endings, distinct from nominal and adjectival ones. These are, for instance, dat.sg. *-*hmāi̯*, abl. *-*hmāt*, loc. *-*hmi*; nom.pl. *-*ai̯*, gen.pl. *-*ai̯šām*; for the Avestan reconstruction see Martínez & de Vaan 2014: 69–70. However, these special case endings of Proto-Iranian pronouns left no trace in the declension of Khotanese demonstrative pronouns. The case endings of Khotanese demonstrative pronouns reflect those of adjectives (masculine, as well as feminine). The adjectival case endings reflect the *a-* and *ā-* stems in Khotanese, except for some peculiarities, analyzed in Emmerick 1968a: 281–288.

The development of adjectival declension in Khotanese is rather regular, since the Proto-Iranian adjectival case endings were not different from those of the nouns. The disappearance of the old pronominal inflected forms of demonstrative pronouns, despite the fact that Khotanese preserves five cases of Proto-Iranian, should be considered innovative. The reflexes of old pronominal inflection can be found in various Eastern Iranian languages, cf. Sogd. nom.pl. *wēšand*, *mēšand* < PIr. gen.pl. **au̯aišām*, **imaišām*; cf. Oss. dat.sg. *amæn*, abl.sg. *amæj*, loc.sg. *ami* < PIr. **ahmāi̯*, **ahmāt*, **ahmi* (with secondary endings) (Edelman 1990: 216–218). The innovative nature of Khotanese pronominal case endings is most probably related to the restructuring of the demonstrative system to a new type.

Most Khotanese adverbs are fossilized forms of the nominative, accusative or old instrumental cases, ending in *-ä*, *-e*, *-u*, *-a* (as well as the genitive *-ye* of the *ă*-declension). They are formally indistinguishable from the corresponding inflected forms, so homonymy does occur. Cf. *dāru* 'long' (adjective, acc.sg.) and *ma-dāru* 'not long, soon' (adverb, indeclinable); *bihīyu* 'exceeding' (adjective, acc.sg.) and *bihīyu* 'a lot' (adverb, indeclinable); *ṣä* 'this' (pronoun, nom.sg.m.) and *ṣä* 'indeed' (adverb, indeclinable); *ttiye* 'this' (pronoun, gen.sg.) and *ttiye* 'then' (adverb, indeclinable) (Bailey 1979: 157, 294, 322, 406; Herzenberg 1981: 270–271).

To sum up, two processes coincided in Khotanese. First, the declension of demonstrative pronouns was reorganized on the model of adjectives, i.e., they became one single morphological class. Second, adjectives themselves became practically indistinguishable from adverbs, since most adverbs are, in fact, identical to accusative singular masculine adjectives. In other words, while there are still some "pure" adjectives and "pure" adverbs, most property words can be used both adjectivally and adverbially. Thus, the boundary between inflected and non-inflected forms became largely elusive in Khotanese, and the virtual identity of the forms *ttāra-* 'that' and *ttara* 'there' does not look surprising in this context. However, the forms differ in the length of the first vowel with the longer form being etymologically more accurate. If Khotanese demonstrative adverbs *mara*; *ttara*; *vara* 'here; there' derive from Proto-Iranian forms suffixed with *-*θra*, it is probable that the length of their first vowel could be affected by the coexistence of forms with and without *-ra*, since Khot. *ma*; *tta*; *va* are still attested in Old Khotanese texts. However, the irregular reflexes of PIr. *-*θr-* are found not only in Khotanese, cf. Av. *auuarə*, Sogd. *wr* instead of Av. *auuaθra*, Sogd. *ʾwrδ*. It is assumed in Rastorgueva & Edelman 2000: 277 that such forms may reflect the suffix *-*ra*, not *-*θra*; thus, the prototype **au̯a-ra* with the same meaning is to be reconstructed.

Finally, if the processes discussed above contributed to the grammaticalization of the Khotanese demonstrative adverb into a distal demonstrative pronoun Khot. *ṣāra-*/*ttāra-*, the direction of grammaticalization is not typologically uncommon, cf. French demonstrative adverbs *ici* 'here', *là* 'there', > *-ci* 'this', *-là* 'that', parts of the proximal and distal demonstrative pronouns, respectively; Hausa demonstrative adverb *cân* 'there' > *càn* 'that', distal demonstrative pronoun (Kuteva et al. 2019: 430–431).

2.3.3 Other demonstrative stems

While Khotanese demonstrative pronouns are innovative and unique within Iranian, the typically Iranian tripartite contrast still holds in Khotanese demonstrative adverbs *mara*; *ttara*; *vara* 'here; there' < **ima-(θ)ra*; **ta-(θ)ra*, **au̯a-(θ)ra*, which continue the Proto-Iranian stem distribution. The reflexes of **-(θ)ra* are found in various other Iranian languages, see the discussion in the subsection 2.3.2. The forms of demonstrative adverbs Khot. *marata*; *ttarata*; *varata* frequently used in the Book of Zambasta are probably characterized by the archaizing spelling, as in other cases where *-ta* is added for the purpose of archaizing the text, cf. *kṣa'ta'* instead of *kṣa'* (Hitch 2016: 98).

Directional adverbs Khot. *vā*; *ttā*; *hā* 'hither; thither' are most probably derived from Proto-Iranian forms unrelated to Proto-Iranian demonstratives, cf. Khot. *vā* < **upa-* preposition / adverbial prefix 'towards; at, to; on' (AiW 388); *hā* < **fraka* adjective and adverb 'forward, forth'. As the medial forms *ttara* and *ttā* do not occur in Old Khotanese, they are likely to be of later origin, plausibly developed by analogy with the other demonstrative adverbs from the stem of the anaphoric demonstrative pronoun *tta-*. Finally, manner demonstratives, cf. Khot. *ttanda-* 'so much'; *ttrāma-*, *ttandrāma-* 'such'; *ttera-*, *ttedera-* 'to such extent', derived from combining the *tta-* stem of the anaphoric demonstrative pronoun with the ontological category marker (*-nda-* of quantity, *-rāma-* of quality, *-tera-*/*-dera-* < **-tara-* of degree). It is noteworthy that these forms have Late Khotanese variants with initial *d-*, cf. LKh. *daṃda-*, *didaṃda-*, *didrāma-*, *dilaka-*, etc. It is possible that these forms represent the only reflexes of PIr. **aita-* in Khotanese, lost everywhere else. However, the more economical explanation would be their emergence in Late Khotanese due to assimilation from *t–d* to *d–d*, see Sims-Williams 1994: 51. Thus, the preservation of Proto-Iranian demonstrative stems such as **ima-*, **au̯a-*, and if possible **aita-* on the periphery of Khotanese demonstrative system may provide yet another piece of evidence for a complete restructuring of Khotanese demonstrative pronouns to a completely new type.

3. Khotanese innovations compared to Tocharian

In section 2 above, I argued that the Khotanese system of demonstrative pronouns is different from the typical Iranian system. In this section, I submit that it finds an almost identical structural counterpart in Tocharian, and this similarity is due to contact-induced convergence. The evolution of Khotanese demonstrative pronouns, which is unique to other Iranian languages, meets its parallels in Proto-Tocharian, cf. (1) PToch. **sə-m(ə), *sa-m(ə), *tə-m(ə)* > TA *säm, säm, täm*; TB *su, sā$_u$, tu* (masc., fem., neut.); (2) PToch. **sə-ṣ(ə), *sa-ṣ(ə), *tə-ṣ(ə)* > TA *säs, sās, täṣ* (masc., fem., neut.); PToch. **sæ, *sa, *tæ* > TB *se, sā, te* (masc., fem., neut.); (3) PToch. **sæ-n(ə), *sa-n(ə), *tæ-n(ə)* > TA *saṃ, sāṃ, taṃ*; TB *seṃ, sāṃ, teṃ* (masc., fem., neut.).

Both Tocharian A and B display the contrast between (1) anaphoric, (2) proximal and (3) distal pronouns, while Tocharian B also displays an additional medial series (see the analysis of the relevant deictic meanings in Kümmel 2013). It was pointed out by scholars that the Tocharian B distal demonstrative pronoun *samp* is of the innovative type, replacing the older distal demonstrative pronoun TB *seṃ*, a cognate of TA distal *saṃ* (see Pinault 2009, Kümmel 2013). Thus, a tripartite system of (1) anaphoric, (2) proximal and (3) distal demonstrative pronouns can be reconstructed for Proto-Tocharian. Apart from that, all the demonstrative pronouns of Tocharian are formed on the basis of two suppletive stems with several historically motivated modifications, which derive from PIE **só-*/**tó-* with possible allomorphs, such as PIE **tid*, **sim* (Pinault 2009: 240–241). According to Pinault, the system of Proto-Tocharian de-

monstrative pronouns was generalized to the **sV/*tV*[6] stem, while deictic contrast was marked by the final index and not by the stem itself (Pinault 2009: 242). For the complete paradigms of Tocharian A and B demonstrative pronouns see Tables 7 and 8.

	Case	Singular			Plural	
		Masculine	Feminine	Neuter	Masculine	Feminine
Anaphoric	nom.	*su*	*sā$_u$*	*tu*	*cey, cai*	*toṃ*
	obl.	*ce$_u$, cau*	*tā$_u$*	*tu*	*ceṃ*	*toṃ*
	gen.	*cwi, cpi*	*tāy*	*tuntse*	*ceṃts*	*toṃts*
Proximal	nom.	*se*	*sā*	*te*	*cey, cai*	*toy*
	obl.	*ce*	*tā*	*te*	*ceṃ*	*toy*
	gen.	*cwi, cpi*	*tāy*	*tentse*	*ceṃts*	—
Medial	nom.	*seṃ*	*sāṃ*	*teṃ*	*cey, cai*	*to(y)na*
	obl.	*ceṃ*	*tāṃ*	*teṃ*	*ceyna*	*to(y)na*
	gen.	*cwi, cpi*	*tāy*	*tentse*	*ceynaṃts, cainaṃts*	*to(y)naṃts*
Distal	nom.	*samp*	*somp*	*tamp*	*ceymp, caimp*	*toymp*
	obl.	*comp*	*tomp*	*tamp*	—	*toymp*
	gen.	*cwimp*	—	—	*cempaṃts*	—

Table 7. Tocharian B demonstrative pronouns (Pinault 2009: 222).

	Case	Singular			Plural	
		Masculine	Feminine	Neuter	Masculine	Feminine
Anaphoric	nom.	*säm*	*sām*	*täm*	*cem*	*tom*
	obl.	*cam*	*tām*	*täm*	*cesäm*	*tosäm*
	gen.	*cami*	*temi*	*tmis*	*cesmi*	*tosmāśśi*
Proximal	nom.	*säs*	*sās*	*täṣ*	*ceṣ*	*toṣ*
	obl.	*caṣ*	*tāṣ*	*täṣ*	*cesäs*	*tosäs*
	gen.	*caṣi*	—	—	*cessi*	—
Distal	nom.	*saṃ*	*sāṃ*	*taṃ*	*ceṃ*	—
	obl.	*caṃ*	*tāṃ*	*taṃ*	*cesäṃ*	*tosäṃ*
	gen.	*cani*	—	*tanis*	*cesni*	—

Table 8. Tocharian A demonstrative pronouns (Pinault 2009: 222).

Considering the above, two features make Khotanese demonstrative pronouns structurally similar to Tocharian and distinct from typical Iranian. First, it is their basic trifold division into anaphoric, proximal and distal. This situation is reconstructed for Proto-Tocharian, and well-preserved in Tocharian A, whereas within Iranian it is unique to Khotanese. Second, it is the use of the same two suppletive stems for all the series (PToch. **s-*, Khot. *ṣ-* occurs in nom.sg.m./f., while PToch. **t-*, Khot. *tt-* is reserved for all oblique and plural forms), while the deictic contrast is indicated by the suffixal indices. Furthermore, similarities are also found in the ety-

[6] In the masculine the *tV-* stem underwent palatalization, probably motivated by the risk of confusion between the masculine and the neuter (Pinault 2009: 231).

mology of these indices, as discussed in the subsection 3.2. Therefore, I claim that the unique evolutionary type of Khotanese demonstrative pronouns is due to Tocharian influence. To demonstrate that, in the following subsections I examine the evidence supporting the hypothesis of language contacts in the Tarim Basin as well as the etymology of Tocharian demonstrative pronouns, highlighting parallels in their formation that exist between Khotanese and Tocharian.

3.1 Language contact in the Tarim Basin

An areal perspective is required for the purposes of this study. Khotanese and the closely related Tumshuqese are known from documents composed in the first millennium CE in the Tarim Basin region. These languages are referred to as Saka or Sakan[7] and could be hypothetically traced back to a common proto-language. Khotanese was spoken to the south of the Taklamakan desert, Tumshuqese to the north. The speakers of the Sakan proto-language probably settled in the Tarim Basin in the first millennium BCE, no later than the time when Khotan itself was founded (3rd century BCE). Thus, M. Peyrot tentatively associates the Ākètălā/Aqtala archaeological culture (from 1000 BCE to 650–550 BCE) with the Sakan-speaking population (Peyrot 2018: 276). The distribution of Aqtala gray ceramics sites fits well within the area where the Khotanese and Tumshuqese languages used to be spoken, and even coincides with the boundary between Tumshuqese and Tocharian B in the north, see Figure 1.

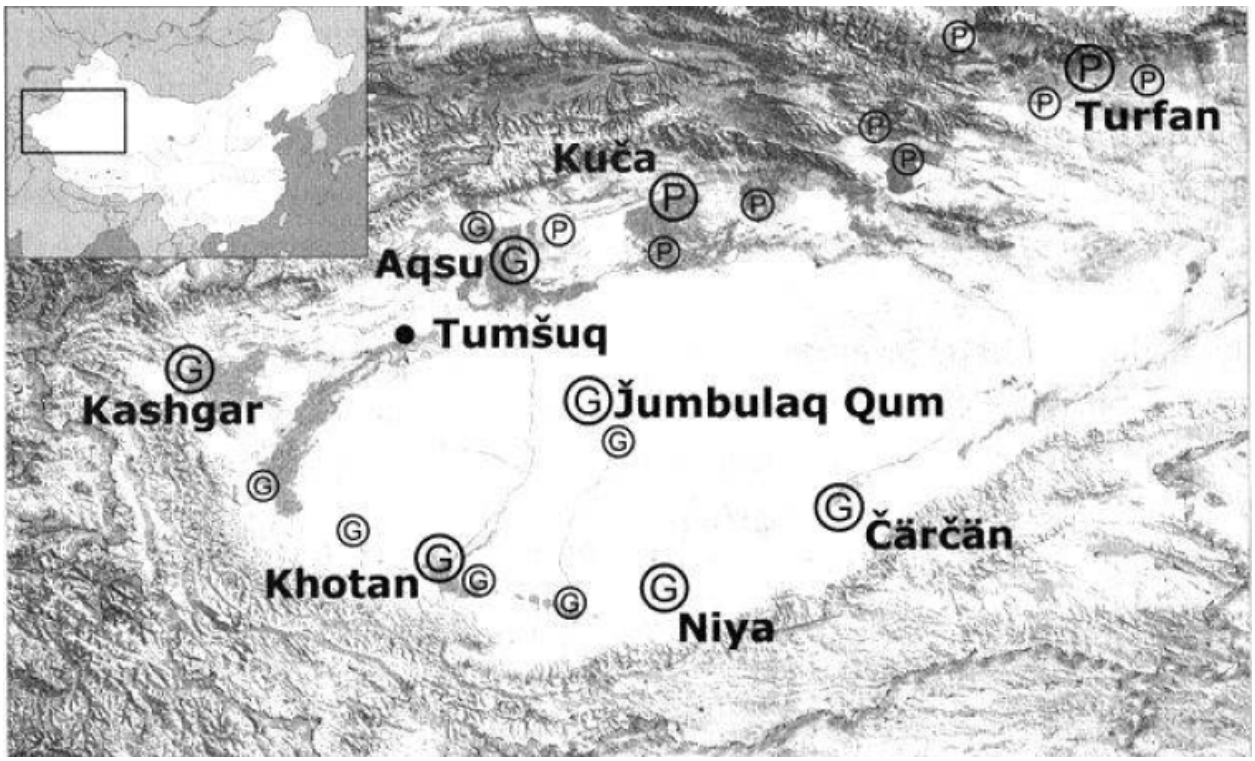

Figure 1. Finds of two ceramics types, gray © and painted ℗ in the Tarim Basin (Peyrot 2018: 276).

In turn, Tocharian A and B are two closely related Indo-European languages spoken in the Tarim Basin and represented by documents dated roughly from the fifth to the eleventh century CE. Contacts between Tocharian, on the one hand, and Iranian, on the other, have been noted by scholars. For instance, X. Tremblay (2005) describes these contacts on account of

[7] The term *Sakan* is introduced by X. Tremblay to avoid confusion of Khotanese and Tumshuqese (and probably Wakhi) with other Iranian languages called *Saka* in Greek and Achaemenid sources (Tremblay 2005: 422).

lexical borrowings at different stages of Tocharian. He concludes that Iranian lexemes began to be borrowed in the Proto-Tocharian period and continued to be borrowed after the split of Proto-Tocharian into Tocharian A and B; the major source of loanwords in Tocharian is Old Sakan (Proto-Tumshuqese-Khotanese) and its descendants (Tremblay 2005). The presence of Old Sakan borrowings in Proto-Tocharian indicates contact between them from at least 500 BCE. However, M. Peyrot argues against Tremblay's labeling of the oldest layer of Iranian borrowings as "Old Sakan". He provides several examples of Old Iranian loanwords in Tocharian that show phonological development uncharacteristic of Sakan, cf. TB *etswe* 'kind of horse' < PIr. **atsu̯a*, contra Old Sakan **aćṵa*, TB *waipecce* 'possessions' < PIr. **hwai-paθya-* (Peyrot 2018: 272, 280). In Sakan, the depalatalization of the Proto-Indo-Iranian cluster **ću̯* never occurred, contrary to most Iranian languages, where **ću̯* > **tsu̯* (Peyrot 2018: 272). Similarly, the palatalization of the cluster **θy* is not present in Sakan (as well as in Sogdian or Pamir languages), but is found in Ossetic and could be attributed to an archaic steppe dialect of Iranian (Peyrot 2018: 280). Several archaic Iranian borrowings are hard to attribute, since they do not show any dialectal features. Nevertheless, Peyrot admits there are cases of borrowings from Old Sakan into Proto-Tocharian, cf. TB *eñcuwo*, TA *añcu* 'iron' and Khot. *hīśśana-* < **ham-ćṵana-* 'iron' demonstrating the development of PIr. cluster **ću̯* specific to Sakan, and not to Old Steppe Iranian (Peyrot et al. 2022). This also correlates with the spread of iron in the Tarim Basin in the early first millennium BCE.

The most comprehensive study on Sakan loanwords in Tocharian is the recent paper by F. Dragoni (2023). It contains nearly 50 words of Old Sakan (Proto-Tumshuqese-Khotanese), Old Khotanese, Late Khotanese and Tumshuqese origin borrowed into Proto-Tocharian, Tocharian A and Tocharian B. Dragoni classifies this new corpus of borrowings according to their phonological and semantic properties, and supports Tremblay's initial claim that the earliest contacts between Sakan and Tocharian took place between 1000 and 500 BCE (Dragoni 2023: 254–258). Conversely, only 3 Tocharian loanwords have been attested in Khotanese and Tumshuqese (a few Tocharian B foreign words in Tumshuqese are treated separately). This list is not complete, but as it appears now, the lexical borrowings in the Tarim Basin were mostly unidirectional (Dragoni 2023: 251). Thus, the centuries-long contacts between Sakan and Tocharian are supported by the data of lexical interference. However, to the best of my knowledge, there are no studies yet concerning the development of grammatical or phonological features as a result of these contacts.

3.2 Reconstruction of Proto-Tocharian demonstrative pronouns

The system of demonstrative pronouns of Proto-Tocharian is based on the alternations of two PIE stems, **só-/*tó-*. However, the etymologically correct derivatives from Indo-European stems should be PToch. **sæ-/*tæ-*, whereas at least one half of Proto-Tocharian demonstrative pronouns stems are reconstructed with the reduced phoneme **ə*, cf. PToch. **sə-/*tə-*. This development could be explained by the assumption that the etymon of certain Tocharian demonstrative pronouns was an unaccented variant of PIE **só-/*tó-* (Adams 1988: 164). The PIE archetype **su-/*tu-* is reconstructed as well (Pinault 2009: 240), which is fortunate from the formal point of view, since it yields **sə-/*tə-* in Proto-Tocharian and does not lead to the palatalization of the preceding consonant, in contrast to PIE **se.* The stem PToch. **tə-* is also found in demonstrative adverbs, cf. TB *tne* < PToch. **tənæ* 'here, (there)'; TA *täpreṃ* 'so much, to such a degree' (Adams 2013: 259, 295). The two hypotheses do not directly contradict each other, but neither do they give an unambiguous answer to the question why in some cases PToch. **sæ-/*tæ-* and in the other PToch. **sə-/*tə-* are reconstructed.

The stems **sə-/*tə-* are also reconstructed for the Proto-Tocharian anaphoric demonstrative pronoun. Despite the different reflexes in the two languages, it is probable that TA *säm*, *sām*, *täm*; TB *su*, *sā$_u$*, *tu* derive from a single prototype, PToch. **sə-m(ə)*, **sa-m(ə)*, **tə-m(ə)*. This may be possible in the case of the varying reflection of PToch. **-m(ə)* in two Tocharian languages, cf. PToch. **-m(ə)* > TA *-m*, TB *-u* (Pinault 2009: 241). The reflection of PToch. **-m(ə)* as TB *-u*/*-w* is backed up by several other examples, cf. TB *-w* (1st person singular present ending), TA *-m* < PToch. **-m*; TB *erkau* 'cemetery' (plural *erkenma*), TA *arkämnā-ṣi* (adjective derived from plural form) < PToch. **erkam*, pl. **erkemənā*. Thus, PToch. **-m* > TB *-w*/*-u*. However, final *-m* in Tocharian B is preserved in certain words. The most convincing possible explanation of this phenomenon is given by S. Burlak: all Tocharian B forms with final *-m* derive from Proto-Tocharian forms with final **-mə*, cf. TB *ñem* 'name' < PToch. **ñæmə* (Burlak 2000: 164–165). Although the details of the change PToch. **-m(ə)* > TB *-u* remain unclear, the common origin of the anaphoric demonstrative pronoun in two closely related languages is more plausible than its development from two completely different markers.

Considering the origins of PToch. **-m(ə)*, several possible etymons have been proposed in the literature. First, I should state that the interrogative particle **mó* proposed by G. Dunkel (2014: 522) does not seem a plausible etymon, since there are no reliable Indo-European parallels in the formation of demonstrative pronouns with **-mó*. Moreover, it does not explain the unetymological PIE **-o* > PToch. **-ə*. The other etymon, which is a result of reanalysis of PIE **tom u* 'den dort' > **to mu* (Dunkel 2014: 522) is formally flawless, but presents an absolutely ad hoc scenario. Without completely rejecting the latter, I argue in favor of PIE **-m* as an etymon. Initially a case marker, it forms innovative stems of demonstrative pronouns in certain Indo-Iranian and Celtic languages, cf. PIE **i-m ó/em* > PIIr. **imám* 'diesen, den hier', PIE **si-m óm* > Ved. *simá-* 'eben er, er selbst'; PIE **si-m de* > Gall. *sinde* 'dieser', OIr. *sin* 'jenes', anaphoric (Dunkel 2014: 357). No parallels, however, are found with the Khotanese anaphoric demonstrative pronoun, because Khot. *ṣa-/tta-* clearly represents the combination of two PIr. stems **aiša-/*ta-*, which also has parallels in another Eastern Iranian language, namely Sogdian, see Section 2.3.

3.2.1. PToch. **sə-ṣ(ə)/*tə-ṣ(ə)* compared to Khot. *ṣäta-/ttäta-*

As discussed above in 2.3.1, the innovative reduplication pattern of the proximal demonstrative pronoun in Khotanese does not find any parallels in other Iranian languages. The Sanskrit reduplication of the type **só-só/*tód-tód* could not be a good parallel, since it is an instance of full reduplication, with case morphemes attached to both stems. Moreover, its usage in Sanskrit is restricted to delimitation or specificational meaning.

However, this pattern can be accounted for if we consider that demonstrative pronouns are formed similarly in Tocharian A. Khot. *ṣäta-/ttäta-* corresponds in semantics to the proximal demonstrative pronoun TA *säs*/*sāṣ*/*täṣ*, which allows to reconstruct PToch. **sə-ṣ(ə)*, **sa-ṣ(ə)*, **tə-ṣ(ə)*, with *-s* in *säs* being a result of depalatalization in Tocharian A (Kümmel 2013). PToch. **-ṣ(ə)* is widely considered a reflex of the deictic particle PIE **-so* or **-se*, thus **sə-ṣ(ə)* < PIE **só so/e*; **tə-ṣ(ə)* < PIE **tó so/e* (Van Windekens 1976: 410, Adams 1988: 162; Dunkel 2014: 738, hesitantly). This etymon is not very adequate, because it either does not explain the palatalization of *-s* (if **só so*), or does not have reliable Indo-European cognates at all (if **só se*).

In contrast, Tocharian B *se*/*sā*/*te* does not show the reflexes of the corresponding particle and yields PToch. **sæ*, **sa*, **tæ* < PIE **so*, **seh$_2$*, **tod*. Considering this, Pinault (2009: 240) proposes a different explanation, assuming that Proto-Tocharian nom.-acc.sg.n. was formed on

the basis of PIE **tid*[8] (possible allomorph of the regular **tód*), which yields PToch. **cäṣ*. According to Pinault, in Tocharian A the final sibilant *-ṣ* of nom.-acc.sg.n. was generalized as the marker of proximal demonstrative pronoun and somehow expanded to the whole paradigm, whereas Tocharian B generalized the form without a sibilant (Pinault 2009: 240). This hypothesis is flawless in terms of formal correspondences; however, it remains unexplained why the final sibilant of nom.-acc.sg.n. was expanded to all other case and gender forms. The most problematic is the following: the final **-d* in PIE **tid* should have palatalized in the earliest stages of Proto-Tocharian as suggested by several pieces of evidence, and without any paradigm leveling survived until the latest stage of Proto-Tocharian, to be dropped in Tocharian B and generalized as the proximal demonstrative pronoun marker in Tocharian A. The formation of palatalized and non-palatalized series of consonants happened relatively early in Proto-Tocharian, as suggested by several facts. For instance, palatalization happened before PIE **e* and not **u*, while in Proto-Tocharian they are not distinct phonemes, as both PIE **e* and **u* yield PToch. **ə* (Burlak 2000: 127). Moreover, the palatalization of the final **-d* in PIE **tid* should have happened before the loss of final **-d* in Proto-Tocharian, which is also considered to have happened very early. Finally, if **-ṣ* was the marker of nom.-acc.sg.n., it contradicts the hypothesis that the masculine **tV-* stem underwent unetymological palatalization motivated by the risk of confusion between the masculine and the neuter, as explained by Pinault (2009: 231).

I suggest a more appropriate etymon, which is PIE **to s̯io-*. PIE **s̯io-* has reliable reflexes of proximal deixis in various Indo-European languages, cf. Ved. *siá*, *syá* 'dieser'; nom.-acc.sg. OIr. *se* 'dieses' < **si̯iod*; dat.sg. OIr. *siu* 'diesem' < **si̯iōi̯*. Moreover, PIE **i̯* yields palatalization of the preceding consonants similarly as PIE **e* and **ī* do (Burlak 2000: 127), thus PIE **s̯io* > PToch. **ṣæ*. Therefore, PIE **to s̯io-* > PToch. **tæ ṣæ* (**tə ṣæ*) is a compound demonstrative in which **ṣæ* is a marker of proximal deixis. Further, it is possible that this marker was used in Proto-Tocharian, but not completely grammaticalized. In Tocharian A it was attached to the reflexes of Proto-Tocharian PToch. **sæ*, **sa*, **tæ*, in Tocharian B, however, the form without **ṣæ* was generalized as the proximal demonstrative pronoun.

Whatever etymology is more plausible, the forms of both Tocharian languages cannot be easily reduced to a single Proto-Tocharian one. However, these are the forms reconstructed on the basis of Tocharian A proximal demonstrative pronoun, namely PToch. **sə-ṣ(ə)*/**tə-ṣ(ə)* that represent the probable sources of calquing in Khotanese. It formally resembles the partial reduplication pattern with the first stem being fossilized, and the case endings added directly to *-ṣ-*/*-s-*, cf. obl.pl.m. *cesäs* (following nom.pl.m. *ceṣ*), gen.sg.m. *caṣi*. Also in the nominative, Khot. nom.sg.m. **ṣä-ṣä* and nom.sg.f. **ṣä-ṣa* strikingly resemble the Proto-Tocharian nom.sg.m. **sə-ṣ(ə)* and nom.sg.f. **sa-ṣ(ə)* (TA nom.sg.m. *säs*, nom.sg.f. *sas* respectively with assimilation). After the nominative, the whole paradigm could have been remodeled following the pattern of partial reduplication. If the Khotanese proximal demonstrative pronoun was actually built up by analogy with the proximal demonstrative pronoun of Proto-Tocharian (Proto-Tocharian-A?), this is an instance of pattern borrowing into Khotanese.

3.2.2. PToch. **sæ-n(ə)*/**tæ-n(ə)* compared to Khot. *ṣāra-*/*ttāra-*

Tocharian A distal and Tocharian B medial demonstrative pronouns derive from a Proto-Tocharian demonstrative series with primarily distal function, as pointed out in the subsection 3.2 above, cf. TA *saṃ*/*sāṃ*/*taṃ*, TB *seṃ*/*sāṃ*/*teṃ* < PToch. **sæ-n(ə)*, **sa-n(ə)*, **tæ-n(ə)* (Pinault

[8] A relatively rare PIE reflex. Except for possibly Tocharian, proposed for Proto-Germanic, cf. Old High German *diz*, *dizzi* 'this' (Dunkel 2014: 782).

2009: 238; Kümmel 2013). It has been pointed out in literature that PToch. *-*n(ə)* stands for the enclitic variant of PIE **áno-*, cf. OCS **onъ* 'that', Lith. *añs*, *anás* 'that' (Pinault 2009: 234; Adams 2013: 362; Peyrot 2017: 641). This enclitic, or rather deictic, particle is reconstructed as PIE **no*, which yields distal demonstrative pronouns in various languages, cf. Av. *a-na*, Ved. *a-ná́* < *é noh*$_1$, PAn. **anno-* 'jener' < PIE *é no*, PIt. **óllo* < PIE **ól no* (Dunkel 2014: 56–57). PIE **no* then should yield PToch. **næ*, not **nə*, unless accented and unaccented variants of the same Proto-Indo-European particle yield different reflexes in Tocharian, as suggested by Adams (1988: 164), cf. the accented PIE **nó-* > PToch. **næ* > TB *-ne*, TA *-ṃ* /n/ (Dunkel 2014: 56). Alternatively, the loss of the final vowel in Tocharian B is explained through the variant PIE **nu* (Pinault 2009: 235). However, no reliable reflexes of PIE **nu* are found elsewhere. Neither can one completely discard the possibility that the final *-*æ* from PToch. **tæ-næ* < PIE **to no* is lost in Tocharian B because of the influence of Tocharian A.

However, for this study it is mainly important that the structure of the Proto-Tocharian distal demonstrative pronoun could provide the source for the innovative formation of the distal demonstrative pronoun in Khotanese. The reflexes of the deictic particle PIE **no* 'that' and the locative particle PIE **no* (Dunkel 2014: 550) became rather indistinguishable in both Tocharian languages, cf. the homonymy between TA enclitic oblique third person singular pronoun *-ṃ* (= TB *-ne*), on the one hand, and TA locative marker *-ṃ* (= TB *-ne*) on the other (Adams 2013: 89, 362); cf. TB *eneṃ*, TA *ane* < PToch. **ene* 'inside, within; herein', TB *tane*, *tene*, *tne* < PToch. **tænæ*, **tənæ* 'here (there)'. Then, the early convergence of both reflexes in PToch. *-*næ* probably led to the demonstrative pronoun **tæ-næ* and the demonstrative adverb **tænæ*, **tənæ* being regarded as formed according to a similar model. From this perspective, the innovative formation of the Khotanese distal demonstrative pronoun *ṣāra-/ttāra-* on the model of Iranian demonstrative adverb **aiša-θra*/**ta-θra* has occurred by analogy with Proto-Tocharian distal demonstrative pronoun **tæ-næ* and demonstrative adverb **tænæ*, **tənæ.*

3.3 Summarizing Tocharian parallels to the deviant features of Khotanese

To sum up, the systems of demonstrative pronouns of Khotanese and Proto-Tocharian display both structural similarities and convergences in the formation of individual demonstratives. In those instances where convergences between Khotanese and Proto-Tocharian can be postulated, Khotanese shows innovations that find no parallel in other Iranian languages. Among the similarities between Khotanese and Proto-Tocharian are the following: (1) the functional distribution of demonstrative pronouns in the system as anaphoric, proximal and distal; (2) formation of all three series from two suppletive stems (PToch. **s-*/**t-*, Khot. *ṣ-*/*tt-*); (3) marking deictic contrast by means of suffixal indices; (4) partial reduplication of PKhot. **ṣa-ṣa-*/**ta-ta-* under the influence of the Proto-Tocharian (A) model **sə-ṣ(ə)*; (5) innovative formation of PKhot. **ṣāra-*/**tāra-* based on the model of the demonstrative adverb **aiša-θra*/**ta-θra* under the influence of the PToch. model linking demonstrative pronoun **tæ-næ* 'that' and demonstrative adverb **tænæ*, **tənæ* 'here, there'. On the basis of these findings, I propose a hypothesis that the Pre-Khotanese, or perhaps the Proto-Tumshuqese-Khotanese (Old Sakan) system of demonstrative pronouns was dramatically reshaped as compared to a typical Iranian system due to continuous language contacts (starting at least from 500 BCE). Proto-Tocharian represents the most likely source of these contact-induced changes. I will discuss possible diachronic scenarios in more detail in the next section 4.

4. The origins of the Tarim type of demonstrative system

As I demonstrated in sections 2 and 3 above, Khotanese demonstrative pronouns are distinct from those in other Iranian languages in several ways. In this regard, they display a diachronic profile distinct from typical Iranian. By comparing Khotanese innovations with those of Tocharian, these features can be identified as markers of the Tarim-type system of demonstrative pronouns. The other languages whose contacts with Khotanese are (well-)known do not show similar characteristics in the system of demonstrative pronouns. For instance, the existence of language contacts between Early Old Khotanese and Old Uyghur is supported by lexical borrowings, cf. OUygh. *balto* 'axe' < OKh. acc.sg. *paḍu* 'axe'; OUygh. *küräš* < OKh. *gūrāś-* 'to quarrel' (Dragoni 2023: 159). However, the language of the tenth to thirteenth centuries Uyghur texts demonstrates typically Turkic reflexes OUygh. *bu* 'this', *ol* 'that'; in conjunction with the deictic particle: OUygh. *ošbu* 'this one', *ošul* 'that one' (Shcherbak 1961: 128). Then, the Middle Indic language Gāndhārī introduced numerous lexemes into Khotanese, which, however, are mostly related to Buddhist concepts, cf. Khot. *ṣṣaddā-* 'faith' < Gāndh. *ṣadha-*, in contrast with Skt. *śraddhā*; Khot. *vīra-* 'hatred' < Gāndh. *vera-*, in contrast with Skt. *vaira-* (Loukota 2023: 26–27). Gāndhārī was used as the official language in Khotan starting from the first century CE, from the Kuṣāṇa rule, and even developed into a local variety of Tarim Gāndhārī with distinctive phonological features as opposed to "Metropolitan Gāndhārī" (Loukota 2023: 10). Similarly to Old Uyghur, Gāndhārī does not show distinctive features of the Tarim-type demonstrative pronouns, cf. Gāndhārī demonstrative pronouns (1) *sa-*/*ta-*, (2) *eṣa-*/*eda-*, (3) *aya-*/*ima-*/*a-* (Baums 2009: 222–224).

4.1 Possible diachronic scenarios

A specific Tarim type could not have developed unless there was a situation of continuous mutual influence between the languages. Then, there are several possible diachronic scenarios of this development. First, one language could influence another directly. Since the evolution of the Khotanese demonstrative system does not reflect the typical Iranian scenario of development, it would be fair to assume that Tocharian influence on Khotanese resulted in restructuring of Khotanese demonstrative pronouns from a typical Iranian system to a Tarim-type system. However, according to the data of lexical borrowings cited in section 3.1, the direction of borrowings was almost exclusively from Old Sakan, later Khotanese and Tumshuqese into Tocharian. Thus, according to the present data, direct Tocharian influence on Khotanese is less probable than vice versa, although not entirely unrealistic. If we assume that Tocharian languages were substrate languages for Khotanese, then the language shift from Tocharian to Khotanese accompanied by imperfect language learning could trigger calques in the system of demonstrative pronouns. Loanwords and structural interference are most common correlates of language contact in the situations of language maintenance and language shift respectively (see Guy 1990: 47–67).

Second, Khotanese could develop such a system of demonstrative pronouns independently before establishing contacts with Tocharian and then influence Tocharian. This could happen at the stage of Peripheral Iranian, after the predecessor of Khotanese, Tumshuqese and Wakhi separated from other Iranian in its phonological development and before the speakers of Old Sakan migrated to the Tarim Basin. These three languages are labeled as Peripheral Iranian mainly because of their unique development of Proto-Iranian aspirated stops (*p^h, *t^h, *k^h) as well as of the PIr. cluster *$ć\underline{u}$[9]. Wakhi is the only living New Iranian language in this group,

[9] See also the alternative reconstruction PIr. *$ts\underline{u}$, as well as the discussion over *$ć\underline{u}$/*$ts\underline{u}$ reflexes in Iranian languages and evidence of Khotanese and Wakhi in favor of palatal *$ć\underline{u}$ reconstruction (Peyrot 2018: 271).

which however appears to have been "Pamirized", i.e. strongly influenced by areal features peculiar to the Pamir languages (that is, Shughni-Roshani, Ishkashimi, etc.). This scenario, however, does not explain the motivation of Khotanese innovations that are not found in any Iranian language outside the Tarim Basin. Moreover, Wakhi demonstrative pronouns do not show similarities to Khotanese innovations either, as discussed later in this section.

A more cautious proposal involves parallel development of demonstrative systems. Over centuries of language contact, Khotanese and Tocharian demonstrative pronouns may have developed a common typological profile. As an instance of parallel development, one can consider the Pamir languages, e.g. Shughni, Wakhi, Ishkashimi, Munji. They are not closely related; however, they display a high level of convergence, despite not having a significant amount of lexical borrowings between each other. Their demonstrative systems developed in a parallel fashion from PIr. **ima-*, **aita-*, **au̯a-* into a tripartite system distinguishing proximal, medial and distal deixis, cf. Shugh. (1) *yam* (obl. *mi*, *mam*); (2) *yid* (obl. *di*, *dam*); (3) *yu*, *yā* (obl. *wi*, *wam*); Wakhi (1) *yəm*; (2) *yət*; (3) *ya(w)*, *yow*; Munji (1) *ma* (obl. *mān*, *māy*); (2) *ya* (obl. *yān*, *yāy*); (3) *wa* (obl. *wān*, *wāy*); Ishk. (1) *am(i)* (obl. *man*); (2) *ad(i)* (obl. *dan*); (3) *aw(i)* (obl. *wan*) (Edelman 1990: 222–228). This scenario also includes a common substrate influence, namely a Burushaski type substrate which was suggested in the literature (see Edelman 1980, Liljegren 2020). A common substrate for Khotanese and Tocharian cannot be excluded either, although the language of the Tarim Basin Bronze Age population is unknown, and its traces in Khotanese and Tocharian vocabulary remain undiscovered.

4.2 Uralic substrate influence hypothesis

Yet another argument in favor of Proto-Tocharian influence on Khotanese demonstrative system can be provided by the hypothesis of Uralic (Proto-Samoyedic) substrate origin of the characteristic features of the Tarim-type demonstrative system. Certain features of Proto-Tocharian, uncharacteristic for Indo-European languages, are explained in Peyrot 2019: 103–105 through parallels with Proto-Uralic and especially its Proto-Samoyedic branch. Among these features are: merger of three Proto-Indo-European series of stops into a single series of voiceless stops; agglutinative case inflection; parallels in the vowel system; object marking expressed on the verb; the use of converbs denoting simultaneity (Peyrot 2019: 103–105). On the basis of mostly archaeological data, it is assumed that contacts between Proto-Tocharian and Proto-Samoyedic speakers took place in the area associated with the Afanasievo culture in Southern Siberia, probably in the late Afanasievo period (Peyrot 2019: 108–109). The estimate for the split of Proto-Uralic lies around 2500 BCE, when Late Proto-Uralic was divided into Samoyedic and Finno-Ugric centers (Häkkinen 2023: 90). The assumed migration of the speakers of Proto-Tocharian from Southern Siberia southward is associated with the end of the Afanasievo culture period approximately in 2500 BCE (Peyrot 2019: 75). However, there is no necessity to claim that the speakers of Proto-Tocharian left Southern Siberia all together and at once in 2500 BCE. Probably, the process of their migration southward down to the Tian Shan mountains stretches at least from 2500 to 2000 BCE (Anthony 2007: 311). Therefore, the contact situation must be dated to no later than 2500–2000 BCE, toward the end of the Afanasievo culture and its succession by the Okunevo culture.

Based on the Uralic (Proto-Samoyedic) substrate hypothesis, I suggest that there are parallels between Proto-Tocharian and Proto-Samoyedic demonstrative pronouns that can be motivated by the abovementioned prehistoric contacts, similarly to the other deviant features of Tocharian. Proto-Uralic demonstrative pronouns are reconstructed as follows in Abondolo 1998: 25–26: proximal PU **tF* (**tV* with a front vowel), distal **tB* (**tV* with a back vowel), ana-

phoric **ćF* (**ćV* with a front vowel), cf. Fin. *tä*= in *tämä* 'this', *tuo* 'that', *se* 'it, that (anaphoric)'. Both Proto-Uralic proximal and distal stems contain the *t-* consonantal base, while the consonantal base of the anaphoric demonstrative pronoun is different. These stems are reconstructed as follows in Raun 1988: 562: PU **ći ~ će* 'this', **tä* 'this', **to* 'that'. Several other demonstrative stems are reconstructed as well, cf. PU **i ~ *e ~ *a* 'this', **u ~ *o* 'that'; **nä ~* (**ne ~* ?**ni*) 'dieser; ? der, jener', **no* 'jener' (Rédei 1988: 67, 300, 306; Abondolo 1998: 26). The proto-forms of demonstrative suffixes are: PU **mɜ*, **kɜ*, **tɜ*, **ŋkɜ* (Rédei 1988: 514).

Based on forms and functions of Samoyedic demonstrative pronouns as mentioned in Wagner-Nagy 2023, Helimski 1998, Nikolaeva 2014, Mus 2023, Sorokina 2010, Kazakevič 2022, Klumpp & Budzisch 2023, Simoncsics 1998, Klumpp 2022, as well as on the demonstrative proto-forms reconstructed in Janhunen 1977 and Rédei 1988, I propose the reconstruction of the Proto-Samoyedic demonstrative system (Table 9). In Nganasan, three series of demonstratives are distinguished: proximal *əm-* 'this here', distal *takəə* 'that over there', *taanʲiə* 'that one', and anaphoric *tə-* 'this/that, it' (Helimski 1998: 500; Wagner-Nagy 2023: 765). In Nenets, proximal, anaphoric, and several distal forms of demonstrative pronouns are found, cf. *tʲuku* 'this', *taki* 'that (deictic)', *tʲexa* 'that (distant, or behind something)', *tajkuj* 'that one over there', *tʲiki* 'that (anaphoric)' (Nikolaeva 2014: 141; Mus 2023: 876–877). In Enets, there are several forms of demonstratives, analyzed in terms of proximity and remotenes in Sorokina 2010: 236, cf. En. *eke* 'this (close to speaker)', *tʃike* 'this (distance neutral)' *tɛxɛ* 'that (away from speaker)', *toj* 'that (away from speaker and addressee)'. In Selkup, proximal, distal, anaphoric, and medial deixis is distinguished, cf. Selk. *tav* 'this (nearest to the speaker)', *tona* 'that (distant from the speaker)', *na* 'this/that (anaphoric)', *ɨlna* 'this (nearest to the hearer)' (Kazakevič 2022: 794; Klumpp & Budzisch 2023: 912–913). According to Kuznetsova et al. 2002: 277, two distinct anaphoric forms are found in Selkup: *na* 'this' refers to an object that was just mentioned, while *tɪna* 'that, that very one' refers to an object mentioned earlier in the conversation. In Kamas, proximal *dy* 'this here', *dym* 'and this (even closer to speaker)', distal *idə* 'and that (even farther from speaker) and *ʃø* 'yonder', and anaphoric (neutral) *dĭ* 'this (neutral in terms of proximity, functions in anaphoric contexts)' are found (Simoncsics 1998: 588; Klumpp 2022: 827). The abovementioned forms are labeled in Table 9 as proximal, distal, and anaphoric/ neutral.

Language	Proximal	Distal	Anaphoric/Neutral
Nganasan	*əm-*	*takəə, taanʲiə*	*tə-*
Nenets	*tʲuku*	*taki, tʲexa*[10], *tajkuj*	*tʲiki*
Enets	*eke*	*tɛxɛ*[11], *toj*	*tʃike*
Selkup	*tav*	*tona*	*na, tɪna*
Kamas	*dy, dym*	*ʃø, idə*	*dĭ*
Proto-Samoyedic	**tə̊-, *tü-, *ɜ-*	**tå-, *tä- ~ *te-*	**ti-, *nɜ-*

Table 9. Samoyedic demonstrative pronouns with Proto-Samoyedic reconstructions.

Ngan. *takəə*, *taanʲiə*, *tə-*, Nen. *tʲuku*, *taki*, *tʲexa*, *tajkuj*, *tʲiki*, En. *tɛxɛ*, *toj*, *tʃike*, Selk. *tav*, *tona*, *tɪna*, Kam. *dy*, *dym*, *ʃø*, *dĭ* reflect six Proto-Samoyedic demonstrative stems, palatalized and non-palatalized correlates, cf. PS **tə̊-*, **tü-*, **tå-*, **tä- ~ *te-*, **ti-* (Janhunen 1977: 144, 150, 152, 160,

[10] Both Nen. and En. *VxV* come from the older **VkV* (Sorokina 2010: 31).

[11] En. *ɛ* is a variant of Enets phoneme *a* after palatal consonants (Sorokina 2010: 26).

167). Similarly in Rédei 1988: 505, 513–514, 526–527, the above forms of Samoyedic demonstrative pronouns are listed as the reflexes of PU **ta*, **to*, **tä* (~ **te* ~ **ti*). Contrary to Janhunen 1977: 153, who tentatively suggests Kam. *ʃø* < (?) PS **te*, it is probable that Kam. *ʃø* derives from PS **tåj*, as a result of regressive palatalization, cf. similar development of Kam. *ʃo* 'Birkenrinde' < PS **tôj* (Janhunen 1977: 145–146). Therefore, Kam. *ʃø* should be considered cognate with the demonstratives of the same series En. *toj*, Nen. *taj=* in *tajkuj*.

Ngan. *əm-*, En. *eke*, Selk. *na* reflect two other Proto-Samoyedic demonstrative stems, cf. PS **ɜ-*, **nɜ(-)* (Janhunen 1977: 19, 105). The reflexes of the PS **ɜ-* are found in the proximal demonstrative pronouns En. *eke*, Ngan. *əm-* (Janhunen 1977: 19). The reflexes of PS **nɜ(-)* are found in Selkup anaphoric *na* (Janhunen 1977: 105). However, Proto-Samoyedic proximal **tü-* does not entirely disappear at least in Enets, cf. En. *tʲuku* 'all, some', a cognate of Nen. *tʲuku* 'this' (Sorokina 2010: 248). Similarly, PS **ti-* yields Selk. *tɪma* 'that, that very one' (Janhunen 1977: 160). The proto-form of Kam. *idə* is not easily identifiable. It can be related to En. *inô* 'dorthin' < PS **i-no* 'dieser da' (Janhunen 1977: 26).

To sum up, the demonstrative pronouns in Samoyedic languages reflect Proto-Samoyedic proximal, distal and anaphoric/neutral proto-forms. The core Proto-Samoyedic system should be reconstructed as based on the proximal – distal – anaphoric composition, similarly to Proto-Uralic. It should be noted that the interpretation of the deictic pronouns in Proto-Samoyedic is not final and that further discussion is needed[12]. However, it can be assumed that six stems, palatalized and non-palatalized correlates PS **tə̂-*, **tü-*, **tå-*, **tä-* ~ **te-*, **ti-* constitute the core of Proto-Samoyedic demonstrative system. Compared to Proto-Uralic as reconstructed by Abondolo (1998: 25–26), Proto-Samoyedic has lost Proto-Uralic opposition between exophoric **tV* and non-exophoric **ć* but preserved the proximal — distal — anaphoric composition of the system. Thus, PU **ć* > PS **s* (Zhivlov 2023: 150) is not involved in the formation of Samoyedic demonstratives, and all three series are built (the composition is similar to Proto-Uralic) on the basis of the same consonantal base **t-*, cf. PU **ta*, **to*, **tä* (~ **te* ~ **ti*). Compared to Tocharian and, more widely, the Tarim-type demonstrative systems, two characteristics of Proto-Samoyedic are crucial: composition of the three series (proximal, distal and anaphoric) and formation of all three series on the basis of the same consonantal base **t-*. Similarly, Proto-Tocharian reconstructions specify a proximal – distal – anaphoric composition, with all three series formed on the basis of **s-*/**t-* stems. If the common features of the demonstrative systems of Tocharian and Samoyedic could be attributed to contacts between Proto-Tocharian and Proto-Samoyedic discussed in this section (as well as several other odd features of Tocharian resulting from such contacts), it is yet another argument in favor of Tocharian influence on the Khotanese demonstrative system.

5. Conclusion

Compared to the typical Iranian composition of the demonstrative system as proximal (1) **aḭ-* : **i̯i-*, **ima-*; medial (2) **aiša-*, **aita-* (and perhaps **ta-*); distal (3) **au̯a-*, **hāu-*, Khotanese demonstrates restructuring to a completely new type: (1) proximal **ṣa-ṣa-*/**ta-ta-*; (2) distal **ṣa-θra-*/**ta-θra-*; (3) anaphoric **ṣa-*/**ta-*; with all stems formed from **aiša-* and **ta-*. The deviant features of Khotanese find their parallels in Proto-Tocharian. Among the similarities are the following:

[12] Cf. also etymologically related forms of the extinct Mator and Koibal languages, the deictic properties of which are perhaps impossible to establish: Mat. *тынъ* 'вот', *дана* 'там', *ты* 'он', *тинь* 'они', *тугокъ* 'только (soviel)', *магай* 'сегодня'; Koib. *дё* 'вот', *тыганъ* 'там', *со* 'сюда', *дэ* 'он', *дугунъ* 'здесь' (Janhunen 1977).

(1) the functional distribution of demonstrative pronouns in the system as anaphoric, proximal and distal; (2) formation of all three series from two suppletive stems (PToch. *s-/*t-, Khot. ṣ-/tt-); (3) marking deictic contrast by means of suffixal indices; (4) partial reduplication of PKhot. *ṣa-ṣa-/*ta-ta- under the influence of the Proto-Tocharian (A) model *sə-ṣ(ə); (5) innovative formation of PKhot. *ṣāra-/*tāra- based on the model of the demonstrative adverb *aiša-θra/*ta-θra under the influence of the PToch. model linking demonstrative pronoun *tæ-næ 'that' and demonstrative adverb *tænæ, *tənæ 'here, there'. By comparing Khotanese innovations with those of Tocharian, these features can be identified as markers of the Tarim-type demonstrative system. Over centuries of language contact, Khotanese and Tocharian may have developed a common typological profile, absent in the neighboring languages. Another possibility of historical development includes Proto-Samoyedic substrate influence on Proto-Tocharian. The composition of Proto-Samoyedic demonstrative system is reconstructed as proximal – distal – anaphoric, with all the three series formed from the same consonantal base *t-, cf. PS demonstrative stems *tə̂-, *tü-, *tå-, *tä- ~ *te-, *ti-. Thus, the characteristic features of the Tarim-type demonstrative system probably have a substrate origin similarly to the other deviant features of Tocharian that resulted from prehistoric contact with Proto-Samoyedic.

Abbreviations

ACC — accusative case; DAT — dative case; DIST — distal deixis; F — feminine gender; GEN — genitive case; INSTR — instrumental case; LOC — locative case; M — masculine gender; N — neutral gender; NOM — nominative case; OBL — oblique case; PL — plural number; PROX — proximal deixis; SG — singular number.

Av. — Avestan; Gall. — Gallic; Gil. — Gilaki; EIr — Eastern Iranian; En. — Enets; Fin. — Finnish; Ishk. — Ishkashimi; Kam. — Kamas; Khot. — Khotanese; Khw. — Khwarezmian; Koib. — Koibal; Kurm. — Kurmanji; Lith. — Lithuanian; Mat. — Mator; Maz. — Mazanderani; MP — Middle Persian; Nen. — Nenets; Ngan. — Nganasan; NP — New Persian; OCS — Old Church Slavonic; OInd. — Old Indian; OIr. — Old Irish; OKh. — Old Khotanese; OP — Old Persian; PAn. — Proto-Anatolian; Parth. — Parthian; PIE — Proto-Indo-European; PIIr. — Proto-Indo-Iranian; PIt. — Proto-Italic; PIr. — Proto-Iranian; PKhot. — Pre-Khotanese; PS — Proto-Samoyedic; PToch. — Proto-Tocharian; PU — Proto-Uralic; Selk. — Selkup; Shugh. — Shughni; Skt. — Sanskrit; Sogd. — Sogdian; Sor. — Sorani; TA — Tocharian A; TB — Tocharian B; Tumsh. — Tumshuqese; Ved. — Vedic Sanskrit; Yazgh. — Yazghulami.

References

Abaev, Vasily I. 1989. *Istoriko-ètimologičeskij slovar' osetinskogo jazyka. Tom IV. U-Z*. Leningrad: Nauka.

Abondolo, Daniel. 1998. Introduction. In: Daniel Abondolo (ed.). *The Uralic Languages*: 1–42. London / New York: Routledge.

Adams, Douglas Q. 1998. *Tocharian historical phonology and morphology*. New Haven / Connecticut: American Oriental Society.

Adams, Douglas Q. 2013. *A Dictionary of Tocharian B.: Revised and Greatly Enlarged. Vol. 10*. Amsterdam / New York: Rodopi.

Anthony, David W. 2007. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton: Princeton University Press.

Badeev, Artyom O. 2024a. Demonstrative pronouns in Old Iranian languages: main types of usage and the reconstruction of the Proto-Iranian system. *Journal of Language Relationship* 22/1-2: 75–105.

Badeev, Artyom O. 2024b. K istorii suppletivnogo sklonenija dal'nego demonstrativa v digorskom dialekte osetinskogo jazyka. In: Danil A. Alekseev, Irina A. Khomchenkova (eds.). *PrOLEGomeny ko vsjakoj buduščej lingvistike, moguščej vozniknut' v smysle nauki. Sb. statej k 35-letiju O. I. Beljaeva*: 19–25. Moscow: Buki-Vedi.

Bailey, Harold W. 1979. *Dictionary of Khotan Saka*. Cambridge / London / New York / Melbourne: Cambridge University Press.

Bogolyubov, Mixail N. 2012. *Trudy po iranskomu jazykoznaniju*. Moscow: Vostočnaja literatura.

Boyle David, Anne. 2014. *Descriptive grammar of Pashto and its dialects. Vol. 1*. Göttingen: Walter de Gruyter.

Chyet, Michael L. 2003. *Kurdish-English dictionary = Ferhenga Kurmancî-Inglîzî with selected etymologies by Martin Schwartz.* New Haven / London: Yale University Press.

Degener, Almuth. 1989. *Khotanische Suffixe*. Stuttgart: Franz Steiner Verlag Wiesbaden GmbH.

De Vaan, Michiel. 2015. Reduplicated demonstratives in ancient Indo-European languages. *Transactions of the Philological society* 113(1): 38–52.

Dragoni, Federico. 2023. *Watañi Lāntaṃ: Khotanese and Tumshuqese Loanwords in Tocharian*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Durkin-Meisterernst, Desmond. 2014. *Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch)*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Edelman, Joy I. 1980. K substratnomu naslediju central'no-aziatskogo jazykovogo sojuza. *Voprosy jazykoznanija* 5: 21–32.

Edelman, Joy I. 1986. *Sravnitel'naja grammatika vostočnoiranskix jazykov: Fonologija*. Moscow: Nauka.

Edelman, Joy I. 1990. *Sravnitel'naja grammatika vostočnoiranskix jazykov: Morfologija. Èlementy sintaksisa*. Moscow: Nauka.

Emmerick, Ronald. E. 1968a. *Saka Grammatical Studies*. London / New York / Toronto: Oxford University Press.

Emmerick, Ronald. E. 1968b. *The Book of Zambasta. A Khotanese Poem on Buddhism*. London: Oxford University Press.

Emmerick, Ronald. E. 1992. *A Guide to the Literature of Khotan, 2nd ed.* Tokyo: The International Institute for Buddhist Studies.

Emmerick, Ronald. E. 2009. In: Gernot Windfuhr (ed.). *The Iranian Languages*. London: Routledge: 377–415.

Emmerick, Ronald. E. (†). 2024. *A Handbook of Khotanese*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Fattori, Marco. 2024. *Il persiano tardoachemenide. Edizione dei testi e profilo storico-linguistico*. Doctoral dissertation. Università degli Studi "La Sapienza" di Roma.

Guy, Gregory R. 1990. The Sociolinguistic Types of Language Change. *Diachronica* 7(1): 47–67.

Häkkinen, Jaakko. 2023. On locating Proto-Uralic. *FUF* 68: 43–100.

Herzenberg, Leonard G. 1981. Xotanosakskij jazyk. In: V. S. Rastorgueva (ed.). *Osnovy iranskogo jazykoznanija. Sredneiranskie jazyki*: 233–313. Moscow: Nauka.

Hitch, Douglas A. 2016. *The Old Khotanese Metanalysis*. Doctoral dissertation. 2016.

Janhunen, Juha. 1977. *Samojedischer Wortschatz: Gemeinsamojedische Etymologien*. Helsinki: Castrenianumin toimitteita.

Kazakevič, Olga. 2022. Selkup. In: Marianne Bakró-Nagy et al. (eds.). *The Oxford Guide to the Uralic Languages*: 777–816. Oxford University Press.

Klingenschmitt, Gert. 2005. *Aufsätze zur Indogermanistik*. Hamburg: Verlag Dr. Kovac.

Klumpp, Gerson. 2022. Kamas. In: Marianne Bakró-Nagy et al. (eds.). *The Oxford Guide to the Uralic Languages*: 817–844. Oxford University Press.

Klumpp, Gerson, Josefina Budzisch. 2023. Selkup. In: Daniel Abondolo, Riitta-Liisa Valijärvi (eds.). *The Uralic languages. Second edition*: 897–938. London / New York.

Konow, Sten. 1941. *Khotansakische Grammatik: mit Bibliographie, Lesestücken und Wörterverzeichnis*. Leipzig: Otto Harrassowitz.

Kudrinski, Maksim, Ilya Yakubovich. 2016. Sumerograms and Akkadograms in Hittite: Ideograms, Logograms, Allograms, or Heterograms? *Altorientalische Forschungen* 43(1–2): 53–66.

Kümmel, Martin J. 2013. About the demonstrative system of Tocharian B. In: Melanie Malzahn et al. (eds.). *International Conference on Tocharian Manuscripts and Silk Road Culture* (*Vienna, June 25–29th, 2013*). DOI: 10.13140/2.1.3703.5207.

Kupfer, Katharina. 2002. *Die Demonstrativpronomina im Rigveda*. Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridge University Press.

Kuznetsova, Adriana I., Olga A. Kazakevich, Ekaterina V. Grushkina, Evgeniy A. Helimski. 2002. *Sel'kupskij jazyk. Učebnoe posobie dlja pedagogičeskix kolledžej i vysšix učebnyx zavedenij.* Sankt-Peterburg: Prosveščenie.

Liljegren, Henrik. 2020. The Hindu Kush–Karakorum and linguistic areality. *Journal of South Asian languages and linguistics* 7(2): 239–285.

Loukota, Diego. 2023. *Ne Hāḍe Vajrropamä Vaśärä: Indic Loanwords in the Khotanese Book of Zambasta and the Chronology of the Spread of Buddhism to Khotan*. Bochum: BuddhistRoad.

Maggi, Mauro. 2009. Khotanese Literature. In: Ronald E. Emmerick, Maria Macuch (eds.). *Literature of Pre-Islamic Iran. Companion Volume I*. London / New York: I. B. Tauris & Co Ltd: 330–417.

Martínez, Javier, Michiel de Vaan. 2014. *Introduction to Avestan*. Leiden / Boston: Brill.

Mus, Nikolett. 2023. In: Daniel. Abondolo, Riitta-Liisa Valijärvi (eds.). *The Uralic languages. Second edition*: 853–896. London / New York: Routledge.

Nikolaeva, Irina. 2014. *A Grammar of Tundra Nenets*. Berlin / Boston: De Gruyter Mouton.

Ogihara, Hirotoshi, Ching Chao-Jung. 2017. Some observations on the Tumshuqese documents newly published in China. In: Alberto Cantera, Maria Macuch (eds.). *Iranica 25, Zur lichten Heimat: Studien zu Manichäismus, Iranistik und Zentralasienkunde im Gedenken an Werner Sundermann*: 453–482. Wiesbaden: Harrassowitz Verlag.

Peyrot, Michaël. 2017. Slavic *onъ*, Lithuanian *anàs* and Tocharian A *anac*, *anäṣ*. In: Bjarne Simmelkjær Sandgaard Hansen et al. (eds.). *Usque ad radices. Indo-European studies in honour of Birgit Anette Olsen*: 633–642. Copenhagen: Museum Tusculanum Press.

Peyrot, Michaël. 2018. Tocharian B etswe 'mule' and Eastern East Iranian. In: *Farnah. Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*: 270–283. Ann Arbor / New York: Beech Stave Press.

Peyrot, Michaël. 2019. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European Linguistics* 7(1): 72–121.

Peyrot, Michaël, Federico Dragoni, Chams B. Bernard. 2022. The spread of iron in Central Asia: on the etymology of the word for "iron" in Iranian and Tocharian. *Bulletin of the School of Oriental and African Studies* 85(3): 403–422.

Pinault, Georges-Jean. 2009. On the formation of the Tocharian demonstratives. In: Elisabeth Rieken, Paul Widmer (eds.). *Pragmatische Kategorien. Form, Funktion und Diachronie. Akten der Arbeitstagung der Indogermanischen Gesellschaft 24*: 221–245. Wiesbaden: Dr. Ludwig Reichert Verlag.

Rastorgueva, Vera S., Joy I. Edelman. 1982. Giljanskij, mazanderanskij (s dialektami šamerzadi i velatru). In: Vera S. Rastorgueva (ed.). *Osnovy iranskogo jazykoznanija. Novoiranskie jazyki: zapadnaja gruppa, prikaspijskie jazyki*: 447–554. Moskva: Nauka.

Rastorgueva, Vera S., Joy I. Edelman. 2000. *Ètimologičeskij slovar' iranskix jazykov. Tom 1*. Moskva: Vostočnaja literatura.

Raun, Alo. 1988. Proto-Uralic Comparative-historical Morphosyntax. In: Denis Sinor (ed.). *The Uralic Languages: Description, History and Foreign Influences*: 555–571. Leiden / New York / København / Köln: Brill.

Rédei, Károly. 1988. *Uralisches Etymologisches Wörterbuch. Band I. Uralische und finnisch-ugrische Schicht*. Budapest: Akadémiai Kiadó.

Shcherbak, Aleksandr M. 1961. *Grammatičeskij očerk jazyka tjurkskix tekstov X–XIII vv. iz Vostočnogo Turkestana*. Moskva: Izdatel'stvo AN SSSR.

Simoncsics, Péter. 1998. Kamassian. In: Daniel Abondolo (ed.). *The Uralic Languages*: 580–601. London / New York: Routledge.

Helimski, Eugene. 1998. Nganasan. In: Daniel Abondolo (ed.). *The Uralic Languages*: 480–515. London / New York: Routledge.

Sims-Williams, Nicholas. 1994. The triple system of deixis in Sogdian. *Transactions of the Philological Society* 92(1): 41–53.

Skjærvø, Prods. O. 2002. *Khotanese Manuscripts from Chinese Turkestan in the British Library: A Complete Catalogue with Texts and Translations*. London: The British Library.

Sorokina, Irina P. 2010. *Èneckij jazyk*. Sankt-Peterburg: Nauka.

Tremblay, Xavier. 2005. Irano-Tocharica et Tocharo-Iranica. *Bulletin of the School of Oriental and African Studies* 68(3): 421–449.

Van Windekens, Albert J. 1976. *Le tokharien confronté avec les autres langues indoeuropéennes. Vol. I. La phonétique et le vocabulaire*. Louvain: Centre International de Dialectologie Générale.

Wagner-Nagy, Beáta. 2023. Nganasan. In: Daniel. Abondolo, Riitta-Liisa Valijärvi (eds.). *The Uralic languages. Second edition*: 753–792. London / New York: Routledge.

Wendtland, Antje. 2009. The Position of the Pamir Languages within East Iranian. *Orientalia Suecana* LVIII: 172–188.

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In: Daniel Abondolo, Riitta-Liisa Valijärvi (eds.). *The Uralic languages. Second edition*: 117–175. London / New York: Routledge.

Литература

Абаев, В. И. 1989. *Историко-этимологический словарь осетинского языка*. Том IV. U-Z. Ленинград: Наука.

Бадеев, А. О. 2024b. К истории супплетивного склонения дальнего демонстратива в дигорском диалекте осетинского языка. В: Д. А. Алексеев, И. А. Хомченкова (ред.). *ПрОЛЕГомены ко всякой будущей лингвистике, могущей возникнуть в смысле науки. Сб. статей к 35-летию О. И. Беляева*: 19–25. Москва: Буки-Веди.

Боголюбов, М. Н. 2012. *Труды по иранскому языкознанию*. Москва: Восточная литература.

Эдельман, Д. И. 1980. К субстратному наследию центрально-азиатского языкового союза. *Вопросы языкознания* 5: 21–32.

Эдельман, Д. И. 1986. *Сравнительная грамматика восточноиранских языков: Фонология*. Москва: Наука.

Эдельман, Д. И. 1990. *Сравнительная грамматика восточноиранских языков: Морфология. Элементы синтаксиса*. Москва: Наука.

Кузнецова, А. И., О. А. Казакевич, Е. В. Грушкина, Е. А. Хелимский. 2002. *Селькупский язык. Учебное пособие для педагогических колледжей и высших учебных заведений*. Санкт-Петербург: Просвещение.

Расторгуева, В. С., Д. И. Эдельман. 1982. Гилянский, мазандеранский (с диалектами шамерзади и велатру). В: В. С. Расторгуева (ред.). *Основы иранского языкознания. Новоиранские языки: западная группа, прикаспийские языки*: 447–554. Москва: Наука.

Расторгуева, В. С., Д. И. Эдельман. 2000. *Этимологический словарь иранских языков. Том 1.* Москва: Восточная литература.

Щербак, А. М. 1961. *Грамматический очерк языка тюркских текстов X–XIII вв. из Восточного Туркестана*. Москва: Издательство АН СССР.

Сорокина, И. П. 2010. *Энецкий язык*. Санкт-Петербург: Наука.

А. О. Бадеев. Указательные местоимения таримского типа в хотаносакском и в тохарских языках: к реконструкции языковых контактов в Таримском бассейне.

В статье рассматриваются основы указательных местоимений хотаносакского языка (< восточноиранские языки), не имеющие морфологических параллелей в других иранских языках. Анализируются дейктические значения, этимология и процессы эволюции основ. Девиантные черты хотаносакских указательных местоимений обнаруживают параллели в тохарских языках. В их число входят: функциональное распределение указательных местоимений в системе в виде анафорической, ближней и дальней серий; образование всех трех серий от двух супплетивных основ; обозначение дейктических значений при помощи суффиксальных показателей; а также морфологическое строение отдельных указательных местоимений. Вышеупомянутые особенности могут рассматриваться в качестве маркеров системы указательных местоимений Таримского типа. В статье утверждается, что многовековые языковые контакты в Таримском бассейне могли привести к развитию в хотаносакском и тохарских языках общих типологических черт, отсутствующих в соседних языках.

Ключевые слова: хотаносакский язык; тохарские языки; указательные местоимения; языковые контакты; уральские языки; самодийские языки.