Российский государственный гуманитарный университет
Russian State University for the Humanities



Russian State University for the Humanities
Institute of Linguistics of the Russian Academy of Sciences

Journal of Language Relationship

International Scientific Periodical

№ 24, issue 1—2



Moscow 2026

Российский государственный гуманитарный университет
Институт языкознания Российской Академии наук

Вопросы языкового родства

Международный научный журнал

№ 24, часть 1—2

Москва 2026

Advisory Board:
William H. BAXTER (Ann Arbor, Michigan)
Václav BLAŽEK (Brno)
Frits KORTLANDT (Leiden)
Alexander LUBOTSKY (Leiden)
Alexander MILITAREV (Moscow)
Valentin VYDRIN (Paris)

Editorial Staff:
George STAROSTIN (Editor-in-Chief)
Anna DYBO (Managing Editor)
Tatyana MIKHAILOVA
Mikhail SAENKO
Ilya YAKUBOVICH
Anastasia Garnet ROMANTSEVSKY (Editorial Secretary)

Founded by Kirill BABAEV

© Russian State University for the Humanities, 2026

Редакционный совет:
В. Блажек (Брно)
У. Бэкстер (Анн Арбор)
В. Ф. Выдрин (Париж)
Ф. Кортландт (Лейден)
А. Лубоцкий (Лейден)
А. Ю. Милитарев (Москва)

Редакционная коллегия:
Г. С. Старостин (главный редактор)
А. В. Дыбо (заместитель главного редактора)
Т. А. Михайлова
М. Н. Саенко
И. С. Якубович
А. Г. Романцевски (ответственный секретарь)

Журнал основан К. В. Бабаевым

© Российский государственный гуманитарный университет, 2026

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. Г. С. Старостина. — М., 2026. — № 24/1—2. — x + 180 с.

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by George Starostin. — Moscow, 2026. — № 24/1—2. — x + 180 p.

ISSN 2219-3820

http://www.jolr.ru/
jolr.rsuh@gmail.com

Дополнительные знаки: С. Г. Болотов
Add-on symbols by S. G. Bolotov

Учредитель и издатель
Российский государственный гуманитарный университет
125047, г. Москва, вн. тер. г. муниципальный округ Тверской,
Миусская пл., д. 6, стр. 6

Свидетельство о регистрации СМИ
ПИ № ФС77-73409 от 03 августа 2018 г.
Периодическое печатное издание

Подписано в печать 12.10.2026. Выход в свет 19.10.2026
Формат 60×90/8. Уч.-изд. л. 14,0. Усл. печ. л. 23,5
Тираж 1050 экз. Свободная цена
Заказ № 2439

Отпечатано в типографии Издательского центра
Российского государственного гуманитарного университета
125047, Москва, Миусская пл., 6, стр. 6
www.rsuh.ru

Table of Contents / Содержание

Contributors / Сведения об авторах viii
Note for Contributors / Будущим авторам x

Issue 1 / Часть 1

Articles / Статьи

Artyom Badeev. Tarim-type demonstrative systems of Khotanese and Tocharian: towards the reconstruction of language contact in the Tarim Basin 3
[*А. О. Бадеев.* Указательные местоимения таримского типа в хотаносакском и в тохарских языках: к реконструкции языковых контактов в Таримском бассейне]

Anton Kogan. Etymological notes on borrowed vocabulary in the Tibetan dialects of Baltistan and Kargil 27
[*А. И. Коган*. Этимологические заметки о заимствованной лексике в тибетских диалектах Балтистана и Каргила]

Pavel B. Lurje. Once more on Sogdian-Yaghnobi linguistic relations 40
[*П. Б. Лурье*. Еще раз к вопросу о взаимосвязи согдийского и ягнобского языков]

Issue 2 / Часть 2

Articles / Статьи

George Starostin. Lexicostatistical Studies In Khoisan III/III: Reconstructing a Swadesh wordlist for Proto-Khoe (items 51–100). 81
[*Г. С. Старостин*. Лексикостатистические исследования по койсанским языкам III/III: реконструкция списка Сводеша для працентрально-койсанского (пракхойского) языка (элементы 51–110)]

Andrei F. Stoica, Xue C. R. Stoica. On some Romanian substrate words with Albanian comparanda: etymologies and implications 122
[*Андрей Ф. Стойка, Сюэ К. Р. Стойка*. О нескольких румынских субстратных словах с албанскими соответствиями: этимология и импликации]

А. В. Савельев. О критериях выявления и стратификации булгаро-чувашских заимствований в пермских языках II. 134
[*Alexander Savelyev*. On the criteria for identification and stratification of Bulghar-Chuvash loanwords in Permic II.]

Reports / Хроника

A. Garnet Romantsevski. The 21st Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics 175
[XXI традиционные чтения памяти С. А. Старостина (А. Гарнет Романцевски)]

Contributors

Artyom Badeev — researcher, School of Linguistics of the Higher School of Economics; assistant researcher of the Institute of Linguistics of the Russian Academy of Sciences (Moscow); badeev@iling-ran.ru

Anastasia Garnet Romantsevski — assistant researcher, Institute of Linguistics (Moscow); sicfaturlacrimans@yandex.ru

Anton Kogan — doctor of sciences (Philology); head of the Department of Asian and African languages, Institute of Oriental Studies of the Russian Academy of Sciences (Moscow); kogan_anton@yahoo.com

Pavel B. Lurje — candidate of sciences (Philology); research fellow / curator, Oriental Department at The State Hermitage Museum (Saint-Petersburg); lurje@panchkath.ru

Alexander Savelyev — candidate of sciences (Philology); senior researcher, Institute for Oriental and Classical Studies, Higher School of Economics / Institute of Linguistics (Moscow); a.savelyev@iling-ran.ru

George Starostin — candidate of sciences (Philology); chief research fellow, Institute for Oriental and Classical Studies, Higher School of Economics (Moscow); external professor, Santa Fe Institute (New Mexico, USA), gstarst1@gmail.com

Andrei F. Stoica — independent researcher; andreifstoica@outlook.com

Xue C. R. Stoica — independent researcher; andreifstoica@outlook.com

Сведения об авторах

Бадеев, Артем Олегович — сотрудник Лингвистической лаборатории по корпусным исследованиям НИУ ВШЭ; лаборант-исследователь Института языкознания РАН (Москва); badeev@iling-ran.ru

Гарнет Романцевски, Анастасия — лаборант-исследователь Института языкознания РАН (Москва); sicfaturlacrimans@yandex.ru

Коган, Антон Ильич — доктор филол. наук; заведующий Отделом языков Азии и Африки Института востоковедения РАН (Москва); kogan_anton@yahoo.com

Лурье, Павел Борисович — канд. филол. наук, заведующий Сектором Средней Азии и Кавказа (Отдел Востока) Государственного Эрмитажа (Санкт-Петербург); lurje@panchkath.ru

Савельев, Александр Владиславович — канд. филол. наук; старший научный сотрудник Института классического Востока и античности НИУ ВШЭ / Института языкознания РАН (Москва); научный сотрудник Института истории и археологии УрО РАН (Екатеринбург); a.savelyev@iling-ran.ru

Старостин, Георгий Сергеевич — канд. филол. наук, ведущий научный сотрудник Института классического Востока и античности ВШЭ (Москва); внешний сотрудник Института Санта-Фе (Нью-Мексико, США); gstarst1@gmail.com

Стойка, Андрей Ф. — независимый исследователь; andreifstoica@outlook.com

Стойка, Сюэ — независимый исследователь; andreifstoica@outlook.com

Note for Contributors

Journal of Language Relationship welcomes submissions from everyone specializing in comparative-historical linguistics and related disciplines, in the form of original articles as well as reviews of recent publications. Articles are published preferably in English or Russian, although publication of texts in other major European languages (such as French or German) may be considered by the Editorial Board.

All submissions should be uploaded electronically in MS Word and PDF format, using the online *Manuscript Submission Form* at the official website of the Journal (http://jolr.ru). Each article should be accompanied with information about the author(s) (names, affiliations, contact information), an abstract (not exceeding 300 words) and relevant keywords.

For more detailed guidelines on article submission and editorial policies, please see our website (http://jolr.ru) or address the Editorial Board directly at (jolr.rsuh@gmail.com. Inquiries may also be sent by regular mail to the official address of the Journal:

Journal of Language Relationship
Institute for Oriental and Classical Studies
Russian State University for the Humanities
125047 Moscow, Russia
Miusskaya Square, 6, bd. 6

Будущим авторам

Журнал *Вопросы языкового родства* принимает заявки на публикацию оригинальных научных статей, а также рецензий от всех, кто специализируется в области сравнительно-исторического языкознания и смежных дисциплин. Предпочтительные языки публикации — английский или русский, хотя по согласованию с редакционной коллегией возможна также публикация статей на других крупных европейских языках (французский, немецкий и т. п.).

Подача материалов для публикации (в форматах MS Word и PDF) осуществляется через *Электронную форму подачи рукописей* на официальном сайте журнала (http://jolr.ru). К каждой статье обязательно прикладываются сведения об авторах (имена, аффилиации, контактная информация), краткое резюме (не более 300 слов) и список подходящих ключевых слов.

Подробнее о требованиях к оформлению рукописи, редакционной политике журнала и т. п. вы можете узнать на нашем сайте по адресу: http://www.jolr.ru или непосредственно от редакционной коллегии по электронной почте (jolr.rsuh@gmail.com). По различным вопросам с редакцией журнала можно также связаться по обычной почте:

125047 Москва
Миусская площадь, д. 6, стр. 6
Российский государственный гуманитарный университет
Институт восточных культур и античности
В редакцию журнала «Вопросы языкового родства»

Issue 1 / Часть 1

Articles / Статьи

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-/*aṷa-, 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 *aṷa-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)* < *aiam/*iiam or *aṷa-; 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-* < *aṷa-* are found. The etymology of Digor distal nom. *je* (obl. *wo-* < *aṷa-*) 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* < *aṷa-* (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)* < *aṷahia(-na)*, prox.gen.sg.m. *ʾwy(n)* < *imahia(-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) *ai- : *ii-, *ima-; medial (2) *aiša-, *aita- (and perhaps *ta-); distal (3) *aṷa-, *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 **au̯a-*, 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, sām, 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.

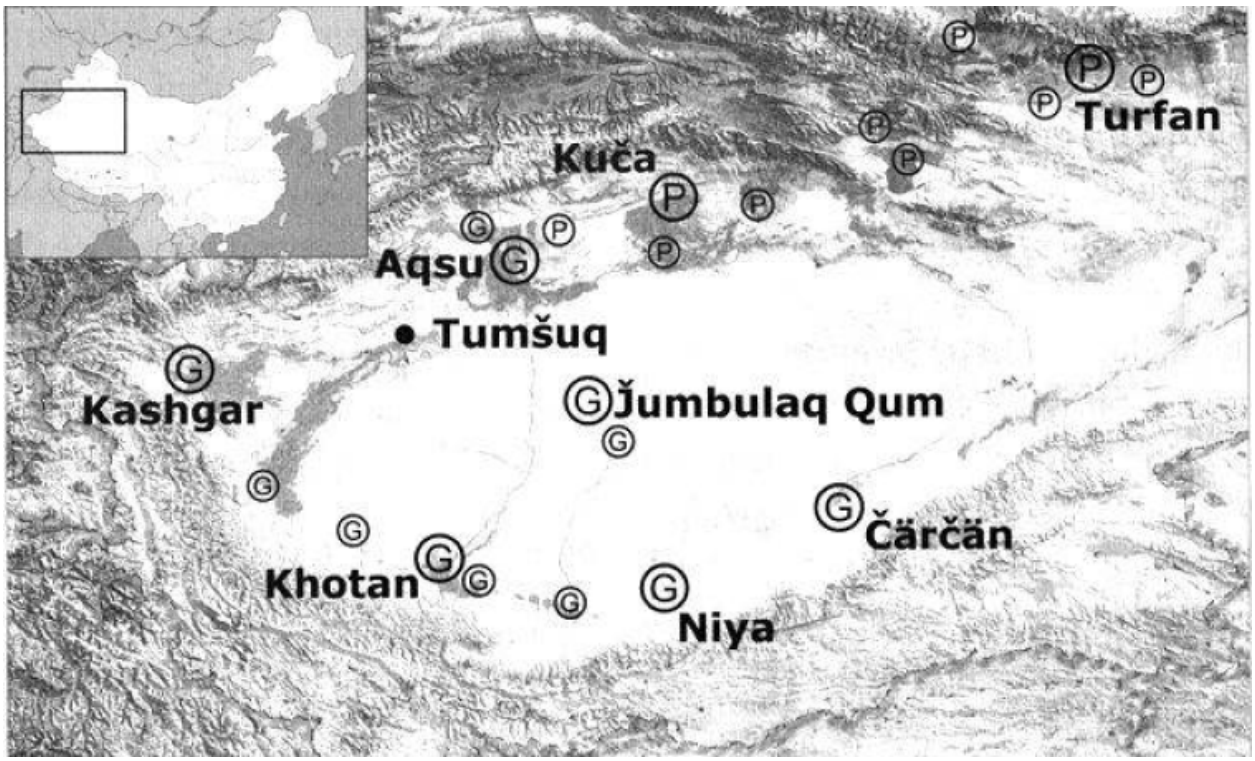


Figure 1. Finds of two ceramics types, gray © and painted ℗ in the Tarim Basin (Peyrot 2018: 276).

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 **aću̯a*, 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-ću̯ana-* '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 *$ć̱$[9]. Wakhi is the only living New Iranian language in this group,

[9] See also the alternative reconstruction PIr. *$ts̱$, as well as the discussion over *$ć̱$/*$ts̱$ reflexes in Iranian languages and evidence of Khotanese and Wakhi in favor of palatal *$ć̱$ 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.

Anton Kogan
Institute of Oriental Studies, Russian Academy of Sciences, Moscow; kogan_anton@yahoo.com

Etymological notes on borrowed vocabulary in the Tibetan dialects of Baltistan and Kargil

The results of recent studies have suggested that a major part of loanwords in the Tibetan dialects spoken in Kargil and Baltistan was adopted from an early form of Burushaski. This conclusion is, however, preliminary and needs to be refined. An attempt at such refinement is made in the present article. The author demonstrates that although Burushaski words really predominate in borrowed vocabulary of Balti and Purik, both dialects contain a considerable number of loans from certain Indo-Iranian sources. Based on historical-phonological data, as well as on analysis of lexical isoglosses, it is argued that these sources are likely to include Dardic Shina and Brokskat.

Keywords: Tibetan dialects; Balti language; Purik language; Burushaski language; Dardic languages; language contact; loanwords.

Introduction

The Tibetan dialects that will be dealt with in the present study are spoken mainly in a part of the Upper Indus Basin, nowadays divided between the Pakistani-controlled area of Gilgit-Baltistan and the Indian union territory of Ladakh. Their speakers differ from the rest of the Tibetans in religion (they are mostly Shia Muslims), as well as physical-anthropologically and, as recent studies have shown, genetically.[1] This fact suggests the existence of pre-Tibetan ethnic substratum, or substrata, in the region and naturally raises the question of possible substrate vocabulary in local Tibetan spoken varieties.

In my earlier publications (Kogan 2019; 2020) I have considered this question for both "Muslim" and "Buddhist" Tibetan dialects of the Upper Indus region. Etymological study of borrowed lexicon led me to conclude that in the varieties spoken in the greater part of Ladakh, possible substrate loans were adopted chiefly from an otherwise unknown Indo-Iranian lect possessing certain historical-phonological features of the Dardic group,[2] whereas for the dialects of Baltistan and Kargil some early form of Burushaski must have been the substrate language. These conclusions are, however, of the most general nature and in some points need further clarification. Among the unclear points is, e.g. the origin of Indo-Iranian loanwords peculiar to Balti and Purik. Such lexical items have been found in a fairly significant number in the course of my research (Kogan 2019). Furthermore, it must be noted that the predominance of Burushaski words in the loan vocabulary of the aforementioned spoken varieties, stated in (Kogan 2019), has been established preliminarily rather than conclusively, and should still be considered a hypothesis that needs testing.

[1] According to the latest genetic research, the genomes of Tibetan-speaking Balti people of Gilgit-Baltistan comprise only 22.6–26% Tibetan ancestry, which suggests that the influence of Tibet on the region involved only minor demic diffusion (Yang et al. 2020).

[2] Many of these loanwords are apparently common for Ladakhi, Purik and Balti dialects, which can be attributed to the fact that the route of Tibetan migration to what is now Kargil and Baltistan must have passed through the present-day Ladakh (Kogan 2019).

The present article attempts to clarify the two aforesaid issues. For this purpose, I have conducted an etymological study of Burushaski and Aryan borrowed vocabulary in several Northwestern Tibetan dialects. As the main data source I used Richard Keith Sprigg's "Balti-English English-Balti dictionary" (Sprigg 2002).[3] It contains chiefly the material of Skardu and Khapalu dialects, spoken in Gilgit-Baltistan and often grouped together under the umbrella term "Balti". Some lexical evidence from the Purik dialect spoken in the Indian district of Kargil has also been analyzed. In the absence of dictionaries, such evidence was often drawn from grammatical descriptions (Bailey 1920; Zemp 2018). In the following section the results of the study are listed in alphabetical order. Balti and Purik words are cited in their original notation.

A list of probable Indo-Iranian and Burushaski loanwords in Balti[4] and Purik

1. Balti ***adiid***, Purik ***ădit*** 'Sunday' (Sprigg 2002: 20; Bailey 1920: 43). Cf. Burushaski, Shina *adít*, Brokskat *aditi* id. (Berger 1998b: 12; Sharma 1998: 130), OIA *āditya-* 'sun', *ādityavāra-* 'Sunday'. Like all the names of the days of the week in Balti, Burushaski and adjoining languages, this word is probably an early loan from Indo-Aryan (Kogan 2024).

2. Balti ***akaash*** 'perforce, unavoidably, obliged to' (Sprigg 2002: 20). Cf. Burushaski *akáṣ* '*Unglück, Pech, was man notgedrungen, wohl oder übel, machen muß*' (Berger 1998b: 14).

3. Balti ***angaaru***, Purik ***ăṅgāri*** 'Tuesday' (Sprigg 2002: 21; Bailey 1920: 43). Cf. Burushaski *aṅáro*, Shina *aṅgáaro*, Brokskat *angāro* id.,[5] OIA *aṅgāraka-* 'the planet Mars'. See also **1**.

4. Balti ***anggon*** 'courtyard (central square of a Baltistan house, open to the sky)' (Sprigg 2002: 21). Cf. Burushaski *agón* '"*Hof*" *um den Mond; Kreis (von Menschen); oberer Raum, Veranda, Empfangsraum*', Shina *hagón* id. (Berger 1998b: 13), OIA *aṅgana-* 'act of walking, courtyard', Pali *aṅgaṇa-* 'open space before palace', Hindi-Urdu *āgan* '(court)yard'.

5. Balti ***araṭap*** 'complete stranger, from another part', ***araṭap-la*** 'suddenly' (Sprigg 2002: 22). Cf. Burushaski *araṭap* '*plötzlich, unerwartet (vom Erscheinen einer Person), zufällig*', Shina *araṭáp* id. (Berger 1998b: 20).

6. Balti ***ashi*** 'responsibility, someone responsible', ***ashipa*** 'owner, master, husband, Lord' (Sprigg 2002: 22), Purik ***afi*** 'ownership', ***afi-pa*** 'owner' (Zemp 2018: 111). Cf. Burushaski (Nagir) *di-aśi* '*überlegen sein, überwältigen*' (Berger 1998b: 23). See also Kogan 2019.

7. Balti ***baan*** 'man or men who sing religious songs and foretell the future' (Sprigg 2002: 24). Cf. Kashmiri *wan-*, Indus Kohistani *ban-* 'to say', Kalasha *bandek* 'to teach; to announce so as to inaugurate, to order or command' (Trail & Cooper 1999: 27), OIA *vandate* 'praises, worships', Av. *vaṇd-* 'to praise, honor, worship'. See also Kogan 2019.

8. Balti ***ban*** 'fence' (Sprigg 2002: 25). Cf. OIA *bandha-* 'border, framework, damming', Shina *bănáa* 'boundary' (Bailey 1924: 131), Indus Kohistani *bʌnà* 'a boundary stone' (Zoller 2005: 301). See also the next word.

9. Balti ***bar-ban*** 'window (in a wall), glass-pane window' (Sprigg 2002: 26), Purik ***barban*** 'window' (Zemp 2018: 945). Cf. Brokskat ***barban*** id. For the element *bar-* cf. OIA *dvār-* 'door,

[3] In my previous works (Kogan 2019; 2020), I occasionally referred to Sprigg's dictionary but abstained from scrutinizing it in full, my main source of lexical data being Rebecca Norman's dictionary of Ladakhi (Norman 2010). In the present article, I analyze only those loanwords noted by Sprigg that are likely to be peculiar to Balti and Purik, i.e. are, except in certain special cases, not attested in other Tibetan spoken varieties.

[4] Lexical data labeled as Balti belong to the dialect of Skardu, unless otherwise indicated.

[5] Unless otherwise indicated, Brokskat lexical data have been drawn from Sharma 1998.

gate', Av. *duuar-* 'gate', Kalasha *durík*, Palula *darúṛi*, Indus Kohistani *dʌrī*, Shina *darií*, Kashmiri *dār* 'window'.[6] For the element *-ban* see the previous word. See also Kogan 2019.

10. Balti ***basanda*** 'dandelion' (Sprigg 2002: 27). Cf. OIA *vāsanta-* 'vernal, pertaining to spring', *vasanta-* 'spring' See also Kogan 2019.

11. Balti ***bat*** 'boiled mixture of germinated grain flour and ordinary flour (made during Ramzan)' (Sprigg 2002: 27), Purik ***bato*** 'rice' (Zemp 2018: 469). Cf. OIA *bhakta-* 'food; boiled rice', Khowar *bot* 'evening meal', Kalasha *batay* 'flour taken to be eaten in high pastures' (Trail & Cooper 1999), Shina *bat*, Bashkarik *batt,* Torwali *bāt,* Kashmiri *batɨ* 'boiled rice', Wakhi *bat* 'wheat flour gruel', Ishkashimi *bat*, Shughni *bāt*, Munji *bātək* 'ritual food made of wheat flour and butter, cooked in water or milk' (Steblin-Kamenskiy 1999: 95).[7] See also Kogan 2019.

12. Balti (Khapalu dialect) ***boṣoq*** 'calf' (Norman 2010: 604). Cf. Burushaski *buṣóoṣo* (Hunza, Nagir), *bóṣo* (Yasin) id. (Berger 1998b: 65). See also Kogan 2019.

13. Balti ***botu***, Purik ***bōdŭ*** 'Wednesday' (Sprigg 2002: 30; Bailey 1920: 44). Cf. Burushaski *bódo* (Hunza), *búdu* (Yasin), Shina *bóodo*, Brokskat *bodu* id., OIA *budha-* 'the planet Mercury', *baudha-, budhavāra-* 'Wednesday'. See also **1**.

14. Balti ***brespot***, Purik ***brešpăt*** 'Thursday' (Sprigg 2002: 32; Bailey 1920: 43). Cf. Burushaski *birésput*, Shina *brésput*, Brokskat *brasput* id., OIA *bṛhaspati-* 'name of a deity, the planet Jupiter', *bṛhaspativāra-* 'Thursday'. See also **1**.

15. Balti ***bwaar*** 'watermelon' (Sprigg 2002: 35). Cf. Shina *buăr* (Bailey 1924), Burushaski (Hunza, Nagir, Yasin) *buar* id. (Berger 1998b: 60). See also Kogan 2019.

16. Balti ***caka*** 'cheese' (Sprigg 2002: 39). Cf. Burushaski *caká* '*Quark, den man in einer Tasche hat abtropfen lassen*' (Berger 1998b: 82).

17. Balti, Purik ***cancan*** 'naked' (Sprigg 2002: 40; Zemp 2018: 926; Norman 2010: 288). Cf. Khowar, Kalasha *čan* id. See also Kogan 2019.

18. Balti ***chaoni*** 'shade among garden trees, courtyard (of large buildings)' (Sprigg 2002: 42). The source of this word must have been a compound with the first part inseparable from Indo-Iranian words for 'shade, shadow': OIA *chāyā-*, Pali *chāyā-*, Sogdian *sy'kh*, Classical Persian *sāya*, Yidgha *sāγo*, Kalasha *čhak*, Khowar *čhaγ*, Shina *čhāi*, Kashmiri *chāy* < Proto-Aryan **sćāi̯ā-* < PIE **sk'āi-*. Cf. also Av. *asaya-* 'without shadow'.

19. Balti ***chaḍu*** 'vile, unclean, unholy, abominable, filthy, foul, impure, lewd' (Sprigg 2002: 42). Cf. Brokskat *č(h)uṭu* 'bad, dirty, rotten'. Probably, together with Kashmiri *chĕṭun* 'to become impure, to become unclean, to become refuse and untouchable' < PII **sćr̥tta-*, the perfect participle of the verb **sćard-* 'to leave, eject, vomit'[8] (> OIA *chr̥ṇatti* 'leaves, vomits', *chardayati* 'vomits', *chardis-* 'sickness, nausea, vomiting'; Parthian *wsrd* 'poured?', Khotanese *e-saly-* 'to besmear'). The cognates of this word with the dentalization of the initial affricate are found in Ladakhi, Purik and Burushaski: Ladakhi *tshiṭu, tshitu* 'unclean, contaminated, disgusting; the Ladakhi Shia concept of contamination of food or water by a non-Muslim; contamination or violation of a spirit, sacred place or holy person, esp. if it is believed to cause illness or spiritual pollution' (Norman 2019: 323), Purik *tshe(r)ṭu* 'dirty' (Zemp 2018: 932), Burushaski *zuṭú* (Hunza-Nager), *zeṭú* (Yasin) '*unrein durch Pollution, Menstruation, oder wenn man nach dem Koitus nicht gebadet hat*' (Berger 1998b: 486). For more details, see Kogan 2025.

20. Balti ***chum*** 'bunch; muscle of thigh' (Norman 2010: 318). Cf. Burushaski *čhu* 'ear (of corn), bunch (of grapes)' (Lorimer 1938: 101), Indus Kohistani *cuī* 'a bunch of grapes or other fruits' (Zoller 2005: 166). See also Kogan 2019.

[6] Burushaski *dəri* 'window' (Lorimer 1938: 114) must have been borrowed from a Dardic source.

[7] The Wakhi, Ishkashimi, Shughni and Munji words, as well as Burushaski *bʌt* in *brʌsɛ bʌt* 'plain, boiled rice', *aluɛ bʌt* 'mashed potatoes' (Lorimer 1938: 73) must have been borrowed from a Dardic source.

[8] From PIE **sk'er-d-* (Pokorny 1959: 947–948).

21. Balti ***cuḍuk, culuk*** ‘small twigs, etc.; broken sticks for firewood’ (Sprigg 2002: 47). Cf. Burushaski *ćúuẏo* ‘*kleines Stückchen (von Holz, Stein, Fleisch usw.)*’ (Berger 1998b: 94). In Burushaski intervocalic *ẏ* regularly reflects earlier *ḍ* in loanwords (Berger 1998a; Kogan 2019).

22. Balti ***ḍanḍa*** ‘torch, stick on the top of which are fitted resinous spikes of wood for burning as a torch’ (Sprigg 2002: 183). Sprigg’s comparison with Hindi-Urdu *ḍaṇḍā* ‘staff', 'pole’ seems to be correct. Cf. also OIA *daṇḍa-* ‘stick, club, handle’, Shina *doo̯ṇŭ* ‘handle (of axe etc.)’, Brokskat *ḍun*, Indus Kohistani *ḍāṇḍ* ‘stick’ (Zoller 2005: 222), Kashmiri *dan* ‘hilt, handle (of hammer)’.

23. Balti ***das*** ‘stony unfertile plot of land’ (Sprigg 2002: 49). Cf. Burushaski *das* ‘*unbebautes Land, Steppe, Wüste, Ödland*’ (Berger 1998b: 116), Brokskat *dus* ‘ground, plain, land, field’, Shina *dās* ‘uncultivated tableland’ (Bailey 1924: 138), Khowar *das* ‘waste land never cultivated, desert, wilderness’. Sprigg’s etymology of the Balti lexeme (< Urdu *dašt* ‘desert, arid plain’) seems to be doubtful for phonological reasons. Turner’s comparison of the Khowar word with OIA *dhāsas-* ‘mountain’ (Turner 1966: 389) is highly questionable because of the semantic difference.

24. Balti ***daxon*** ‘sheet, scarf (Baltistan)’ (Sprigg 2002: 49). Cf. Ladakhi ***daon***, ***dahon***, ***dagon*** ‘head-scarf, a shawl or large scarf covering a woman’s hair according to Islam, *chador, hijab*’ (Norman 2019: 185), Burushaski *doón* ‘*Frauenkopftuch*’, Shina *doón* id. (Berger 1998b: 122). Since the Shina word has no Aryan etymology, it is reasonable to assume that it is a Burushaski loan. The Ladakhi word denotes one of the cultural realia peculiar to neighboring Muslim areas rather to Ladakh itself, and therefore may well have been borrowed from Balti.

25. Balti ***ḍen*** ‘sulphur’ (Sprigg 2002: 183). Cf. Burushaski *dáṅċil* ‘*Schwefel, findet sich in Schwefelquellen, die als heilkräftig gelten*’, Shina *daṅċál* id. (Berger 1998b: 114). The Shina word may be a loan from Burushaski. The final element *-ċil* in Burushaski is probably etymologically identical to *ċhil* ‘water’. The original meaning of the Burushaski word may thus have been ‘water of a sulphur spring’. See also Kogan 2019.

26. Balti ***ḍiangs*** ‘hard (of soft things that have become hard, bricks, food, etc.)’ (Sprigg 2002: 183). Cf. Burushaski *ḍaṅ* ‘*hart (auch Strafe); tapfer, ausdauernd*’ (Berger 1998b: 130), Khowar *ḍang* ‘hard, stiff’ (Bashir 2022: 42). Balti ***ḍuang***, ***ḍaang*** ‘high-flown; ostentatious (of speech); very tall (of small children)’ (Sprigg 2002: 183) may also belong here. See also Kogan 2019.

27. Balti ***dom-ba*** ‘trouble, want, necessity’ (Sprigg 2002: 51), ***dom*** ‘sadness, trouble, difficulty, adversity’ (Norman 2010: 493). Cf. Shina *dămizhār* ‘adversity, trouble’ (Bailey 1924), Burushaski *dʌm(i)jar* ‘trouble, inconvenience, worry’ (Lorimer 1938). The Shina word is most probably the source for the Burushaski one, and may be related to OIA *damayati* ‘tames, subdues’, Ossetic *domyn* ‘to tame; exhaust; demand’. Purik ***domba*** ‘bad coincidence’, ***dom*** ‘coincide unfortunately’ (Zemp 2018: 437, 931) may also belong here. Semantically, the Purik words may have been influenced by the reflex of Old Tibetan *’dom pa* ‘to come together’. See also Kogan 2019.

28. Balti ***ḍuhum***, ***ḍuhung*** ‘large hole in a field, old tree, among rocks’ (Sprigg 2002: 183). Cf. Bashkarik *ḍoṅ* ‘hole’ (Morgenstierne 1940), Indus Kohistani *ḍúṅg* ‘a pit, hole (in the earth), depression, hollow’ (Zoller 2005: 223). See also Kogan 2019.

29. Balti ***gam-pa*** ‘lower, step’, ***gam-pa-ing*** ‘downstairs’, ***gam-pa-po-ri*** ‘by the lower way’ (Sprigg 2002: 57). Cf. Burushaski *gamún* ‘*Unterteil; (am Baum) Wurzel, Stumpf*’ (Berger 1998b: 145), Shina *gabŭn* ‘foot of mountain, bottom, tree-trunk’ (Morgenstierne 1945: 92), Brokskat *gamuni* ‘down, below’, OIA *gambhan-* ‘depth’. Sprigg’s comparison with Tib. *gan* ‘nearness’ seems to be less plausible. The meaning ‘step’ in Balti may have developed due the influence of Tib. *gom pa* id. Balti ***gap*** ‘down’ (Sprigg 2002: 58), must belong to the same etymon. The Indo-Iranian sources of the Balti words with and without nasal, most probably, reflect a full and a zero grade derivative respectively.

30. Balti ***ganduru*** 'windpipe' (Sprigg 2002: 57). Cf. Brokskat *ganduru* 'throat'. Possibly, related to Kashmiri *gaḍur*, Khowar *gurūḷi*, OIA *gaṇḍa-* 'goitre'.

31. Balti ***ghaash***[9] 'epilepsy, phlegm, sputum' (Sprigg 2002: 61). Cf. Burushaski *ġaṣ* '*Schaum*' (Berger 1998b: 173). Sprigg's comparison with Urdu *γaš* 'a swoon, stupor, insensibility, fainting' is questionable. However, the meaning 'epilepsy' in Balti may have developed due to some semantic influence of the Urdu word.

32. Balti ***ghanḍel*** 'cross-eyed' (Sprigg 2002: 61). Cf. Burushaski *ġanḍér* '*krumm*', *-lćine ġanḍér* '*schielend*' (Berger 1998b: 169).

33. Purik ***ʁondil*** 'dove' (Zemp 2018: 87). Cf. Burushaski *ġóġundil* (Hunza-Nager), *ġunḍíl* '*Taube*' (Berger 1998b: 177).

34. Balti ***ghonii*** 'wickedness, evil' (Sprigg 2002: 61), Purik ***ʁoni*** 'evil, wickedness' (Zemp 2018: 87). Cf. Burushaski *ġuníqiṣ, ġuníkiṣ* '*schlecht (Menschen, Sachen, Reden), hässlich*' (Berger 1998b: 179-180). The element *-qiṣ/-kiṣ* in Burushaski is a suffix forming adjectives (Berger 1998a: 204). The connection between the Balti and Northwestern Tibetan words was admitted by Berger and Zemp.

35. Balti ***ghul*** 'rotting wood, tinder' (Sprigg 2002: 62). Cf. Burushaski *ġul* '*das, was nach dem Worfeln des Getreides als Abfall bleibt, Spreu; verfaultes Holz*', *ġulpo* '*verfaultes Holz*' (Berger 1998b: 177–178).

36. Balti ***ghuṭing*** 'wood pigeon' (Sprigg 2002: 62). Cf. Burushaski *kúuṭi* '*wilde Taube*', Shina *kúuṭi* id. (Berger 1998b: 249). The initial voiced uvular in Balti may have developed by analogy with the cognate of Purik *ʁondil* 'dove' (see **33**). The final nasal of the Balti word needs explanation.

37. Balti ***gra-a***, Purik ***gra***, 'goitre' (Sprigg 2002: 68; Zemp 2018: 935). Cf. Shina *gaáro*, Burushaski *ġáaro* id. (Berger 1998b: 164). See also Kogan 2019.

38. Balti ***hal*** 'goal' (Sprigg 2002: 74). Cf. Burushaski *hála* (Hunza, Nagir), *hal*, *hálu* (Yasin) '*Tor (beim Polo)*', Shina *hal*, Khowar *haḷ* id. (Berger 1998b: 187). See also Kogan 2019.

39. Balti ***hariṭham*** 'small centipede' (Sprigg 2002: 75), Purik "***ha-ri-tsam***" 'centipede' (Jäschke 1881: 595). Probably, somehow connected with Burushaski *híriman* (Hunza, Nagir), *hírmen* (Yasin) '*Insekt, dem Tausendfüssler ähnlich*' (Berger 1998b: 200). See also Kogan 2019.

40. Balti ***haral*** 'open-air sheep-fold (in the high mountains)' (Sprigg 2002: 75). Cf. Burushaski *haráay̯* '*Einzäunung, Hürde, Pferch; eingezäunter Lagerplatz für Vieh und Hirten auf der Hochweide*', Shina *haráai* id. (Berger 1998b: 192), Indus Kohistani *hírā* 'an enclosed or fenced area for sheep and goats where their feet are bound with strings in such a way that they can move around only slowly' (Zoller 2005: 409). The final *l* in Balti, as well as the final *y̰* in Burushaski may reflect an earlier cerebral.

41. Balti ***hasiri*** 'kitchen work', ***hasiri-khang*** 'kitchen', ***hasiri-pa*** 'cook' (Sprigg 2002: 75). Cf. Shina *hasĭriyĕ* 'cook' (Bailey 1924: 142). Probably, somehow connected with Burushaski *ási* (Hunza), *hási* (Nager, Yasin) '*Frühstück (in vornehmen Haushalten)*' (Berger 1998b: 21).

42. Balti ***hudia bya*** 'vaccinate' (Sprigg 2002: 76). Cf. Burushaski *(h)udá --t-* '*impfen*' (Berger 1998b: 203), Shina *hŭdā* 'inoculation' (Bailey 1924: 143), Khowar *hudá* 'vaccination' (Sloan 1981: 86). If the primary meaning of the word is 'grafting', it may reflect some derivative of PII **ud-dāya-* 'binding, fastening' (cf. OIA *dā-* 'to bind', *uddāna-* 'act of binding or fastening together', *saṃdāya-* 'rein, leash').

43. Balti ***jabal*** 'crow-bar' (Sprigg 2002: 78). Cf. Burushaski, Shina, Khowar *jabál* '*Steinbohrer, Brechstange*' (Berger 1998b: 220), Torwali *jabal* 'pickaxe', Pashto, Dari *jabal*, Nepali *jhampal* 'crow bar'. Perhaps, a local Wanderwort.

[9] In Sprigg's notation, *gh* stands for the voiced uvular fricative.

44. Balti ***kaari*** 'basket' (Sprigg 2002: 80). Cf. Shina *karĕ* (Bailey 1924: 146), Brokskat *kari* id., Indus Kohistani *kʌrèʈʰ* 'a flat fruit basket' (Zoller 2005: 103), Khowar *khareṭi* 'flat basket'. The Dardic words probably reflect PII **skăra-* (> OIA *khāra-* 'a particular measure' (a Prakritism?), OIA, Pali *khārī-* 'a measure of capacity', Lahnda, Punjabi *khārā* 'large basket made of reeds', Proto-Iranian **skăra-* > Pashto *škōr, škərəy* 'wicker basket'[10]). Burushaski *kharéṭi* '*kleiner Weidenkorb*' (Berger 1998b: 252) may be a Dardic borrowing.

45. Balti ***karanṭi*** 'mason's trowel' (Sprigg 2002: 82). Cf. Indus Kohistani *karʌɳḍī* 'a spatula' (Zoller 2005: 102), Pashto *karanḍəy*, Hindi-Urdu, Punjabi *karnī*, Marathi *karṇī* 'mason's trowel'. According to Turner (1966: 140), the Indo-Aryan words reflect OIA *karaṇa-* 'act, deed', Prakrit *karaṇa-* 'instrument'.

46. Balti ***kasna*** 'brass horn (musical)' (Sprigg 2002: 82). Sprigg reasonably compares this word with Hindi-Urdu *kāsā* 'bell-metal, brass'. Cf. also OIA *kaṁsa-* 'metal cup, bell-metal', *kāṁsya-* 'made of bell-metal', Prakrit *kaṁsa-* 'bell-metal, drinking vessel, cymbal', *kaṁsiā-* 'a kind of musical instrument', Woṭapuri *kasṓṭ* 'metal pot', Kashmiri *kenz* 'a kind of cup made of clay or copper'.

47. Balti ***kelak, kyalak*** 'twisted' (Sprigg 2002: 82). Cf. Khowar *koli*, Palula *kūulo*, Shina *kōlụ* 'curved, crooked' < PII **kaula-* (> Proto-Iranian **kaura-* > Khotanese *kūra-* 'crooked' (Bailey 1979: 62)). Brokskat *kolo* 'lame' may also belong here. The prototype of the Balti word in the donor language probably reflected an old formation with the *-k-* suffix. The vowel of the first syllable in Balti seems to have developed due to the influence of some derivative of Tibetan *'khyil-ba* 'to wind, to twist'. See also Kogan 2019.

48. Balti ***khaaḍi*** (Skardu), ***khaaṭii*** (Khapalu) 'basket' (Sprigg 2002: 85). Probably, a Burushaski loanword (cf. Burushaski *kharéṭi* '*kleiner Weidenkorb*' (Berger 1998b: 252)) and an etymological doublet of Balti *kaari* 'basket' (see **44**).

49. Balti ***khaṭ*** 'get fixed, be forced to stop' (Sprigg 2002: 86), Purik ***χaṭ*** 'get stuck' (Zemp 2018: 43). Cf. Burushaski *du-kháṭ-* '*gehindert werden, eingequetscht werden, gehemmt werden, sich verfangen, gefangen werden*' (Berger 1998b: 253). The contamination of the Burushaski loan with the reflex of Old Tibetan *khad* 'to stick fast, to be seized, stopped, impeded' (Jäschke 1881: 38) must have taken place in Balti and Purik. Balti ***khaṭ*** 'pole with a catch' (Sprigg 2002: 86) may also belong here.

50. Balti ***khuṭ*** 'short (speech, etc.); small (area)', ***kot ci***, ***koṭ ci*** 'a short while' (Sprigg 2002: 91, 93). Cf. Burushaski *khuṭ*, Shina *khuṭo*, Brokskat *χoṭo*, Indus Kohistani *khʌṭʌ̀ṽ* (Zoller 2005: 120), Khowar *iskurdi* 'short'. See also Kogan 2019.

51. Balti ***kóḍos*** 'rounded stone, cobble-stone' (Sprigg 2002: 93). Cf. Burushaski *guyos* 'grinding-stone' (Lorimer 1938: 176). See also Kogan 2019.

52. Balti ***lalu, laḍu, laṭu*** 'cucumber' (Sprigg 2002: 97, 98). Cf. Burushaski *láay* (*y* < *ḍ*), Shina *lā*, Kashmiri *lār* id., Palula *laṛu* 'pumpkin', Kalasha *laù* 'kind of melon or cantaloupe' (Trail & Cooper 1999: 182), Khowar *laṫù* 'species of melon' (Bashir 2022: 78), Dameli *laŕū* 'a kind of melon' (Morgenstierne 1942: 177).

53. Balti ***lashi*** 'a resinous wood used as a candle or torch because it burns slowly' (Sprigg 2002: 98). Cf. Kashmiri *làshi* 'a torch' (Grierson 1915-1932: 533), Shina (Gilgit dialect) *lăi* 'torch (unlit)' (Bailey 1924: 150). In the Gilgit dialect of Shina the change **ś* > *i* is regular in the intervocalic position. The Dardic forms seem to have developed due to contamination of the two Indo-Iranian roots: **ruś-*/**luś-* 'to shine' (> OIA *ruśant-* 'bright', Khotanese *rrus-* 'to shine', Pa-

[10] For Pashto *šk* < Proto-Iranian **sk*, cf. also, e.g. *škārəl* 'to incite against' < Proto-Iranian **skar-* 'to pursue' (cf. Av. *skār-* id.), *škawəl* 'to tear, break, rip off; rob' < Proto-Iranian **skap-*/**skaf-* 'to split, make a crack' (< PIE **(s)kep-* '*hacken, hauen*' (LIV: 555), cf. Greek *σκάπτω* 'I dig'). Pashto *ō* in *škōr* regularly reflects Proto(-Indo)-Iranian **ā*.

shai *luš* 'bright', Shina *lŭṣṭáikḭ* 'morning') and **las-* (> OIA *lasati* 'flashes, shines; sports'). See also Kogan 2019.

54. Balti ***lish byar*** 'stick tightly to' (Sprigg 2002: 103). Cf. OIA *ślíṣyati* 'adheres to', *ślēṣayati* 'connects', *ślēṣa-* 'adhering to', Av. *hąm.srišāiti* 'puts together', Classical Persian *sirišt*, Ossetic *sasm* 'glue', Shina *lĭṣ thoĭkḭ* 'to join', *lĭṣ boĭkḭ* 'to cling' (Bailey 1924: 150), Brokskat *leẓis* 'to close' (Ramaswami 1989: 117), "*lisit*" 'to stick' (Sharma 1998: 151), Palula *lišóo* 'to close, put together' (Liljegren, Haider 2011: 94), Indus Kohistani *lìṣ karáṽ* 'to close tightly; to seal' (Zoller 2005: 366), Kalasha *lišék* 'to close something tightly' (Trail & Cooper 1999: 183). Burushaski *liṣ* '*verschlossen, zugemacht*' (Berger 1998b: 266) is, most probably, a Dardic borrowing.

55. Balti ***mal*** 'plot of ground that one works oneself, place' (Sprigg 2002: 109), Purik ***mal*** 'land' (Zemp 2018: 684). Cf. Burushaski *mal* '*Feld*' (Berger 1998b: 275). Cf., however, also Tib. *mal* 'the place where a thing is, its site, situation' (Jäschke 1881: 412). Perhaps, in Balti and Purik we are dealing with two homonyms, the one with the meaning 'place' being inherited from Old Tibetan, whereas the other one with the meaning 'plot of ground, land' being a Burushaski loan.

56. Balti ***moncholok*** 'bark cut off for firewood' (Sprigg 2002: 118). Probably, a two-part compound. For the element *mon-*, cf. Burushaski *mon* (Hunza-Nager) '*Artemisie, Pflanze der Hochweide, stark riechend, als Brennmaterial verwendet*', *mun* (Yasin) '*Reisig*' (Berger 1998b: 290). For the second component (*-cholok*), cf. OIA *challi-*, Pashai *čalak* 'bark'.

57. Balti ***monṭhok*** 'clod of earth' (Sprigg 2002: 118). With dissimilation of middle consonants < **maṭ-ṭhok*? Cf. OIA *mr̥ttikā-* 'earth, clay', Shina *māṭḭ* 'clay', Brokskat *moṭi* 'mud'; Purik *ṭoqṭoq* 'lump' (Zemp 2018: 926), Burushaski *ṭhóko* '*Klumpen, Ballen (von Zucker, Salz, Butter)*' (Berger 1998b: 452), Indus Kohistani *ṭhokh* 'a clod (earth, salt)' (Zoller 2005: 220). For the second part of this compound see **85**. See also Kogan 2019.

58. Balti ***morok*** 'carving' (Sprigg 2002: 118). Possibly, connected with Burushaski *maráq* '*Krümmung, Biegung, Kurve*', Shina *marák* id. (Berger 1998b: 281).

59. Balti ***moṭo*** 'beans' (Sprigg 2002: 118). Cf. OIA *mukuṣṭ(h)a-* 'a kind of bean', Hindi-Urdu *moṭh* 'a kind of pulse or vetch', Shina "*muṭhù:*" 'beans' (Rajapurohit 2012: 103), Indus Kohistani *mōṭā* 'a type of dark red and kidney-shaped beans' (Zoller 2005: 345), Kashmiri *muṭh* 'the vetch Phaseolus aconitifolius'.

60. Balti ***mulak***, ***mulu***, Purik ***mulaq*** 'turnip' (Sprigg 2002: 118; Zemp 2018: 944). Cf. Burushaski *múlo*, Shina *muúlo*, Brokskat *mulo* id., OIA *mūla-* 'root', *mūlaka-* 'radish', Shina *mŭlī*, Khowar *mūḷ* 'root', Pashai *mūluk*, Hindi-Urdu *mūlī* 'radish'. See also Kogan 2019.

61. Balti ***mun*** 'short stubby dry branch, cut off; large thorn; blunt-nosed' (Sprigg 2002: 118). Cf. OIA *muṇḍa-* 'shaved, bald, lopped, hornless; blunt; trunk of lopped tree', Khowar *mun* 'stump of tree', *kər-mvn* 'crop-eared', Shina *mūn* 'stump or bole of tree, stump of amputated leg or arm, maize stubble', Palula *múṇḍo* 'blunt', Indus Kohistani *mùṇḍʌr*, *mùṇḍur*, *mùṇḍraʰ* 'the stump of a tree, a piece of a wooden log, a club, mallet', *mũ̀ṇḍ*, *mũṇḍ* 'a stump of tree' (Zoller 2005: 340, 344). Burushaski *mun* '*Stumpf, Pflanze, deren Blätter oder Oberteil abgeschnitten sind*' (Berger 1998b: 294) must be a Dardic loanword.

62. Balti ***mush*** 'squeegee type of rake' (Sprigg 2002: 119). Cf. Burushaski *máuṣ* (Nagir), *máuṣk* (Yasin) '*zahnloser Rechen für Erde und Dung*' (Berger 1998b: 284), Khowar *máuṣ* 'wooden hoe' (Strand 2006). See also Kogan 2019.

63. Balti ***múyu***, ***myo*** 'mouse', ***munzhur*** 'small mole-like mouse' (Sprigg 2002: 119). Cf. OIA *mūṣ-*, *mūṣā-*, *mūṣika-* 'mouse, rat', Persian *mūš*, Palula *mũ̀ṣo* 'mouse', Shina (Guresi dialect) *mūẓu* 'rat', *mŭẓai* 'mouse', Brokskat *mūži* 'mouse'. The variant *myo* in Balti may have developed due to the influence of *byo ba* 'mouse, rat'. See also Kogan 2019.

64. Balti ***paghon***, ***phraghon***, Purik ***phoGon*** 'spoon' (Sprigg 2002: 127, 130; Norman 2010: 602). Cf. Burushaski *khʌpun*, Khowar *khipini*, Shina *khapai*, Brokskat *khapi*, Indus Kohistani

khapʌ̃ŋ̃ ‘spoon, ladle’ (Lorimer 1938: 228; Zoller 2005: 123). In Balti and Purik a metathesis of stops must have taken place. See also Kogan 2019.

65. Balti, Purik ***payu*** ‘salt’ (Sprigg 2002: 127; Norman 2010: 558). Cf. Burushaski (Hunza, Nagir, Yasin) *bayú*, Brokskat *payu*, Shina *pažúu* id. The intervocalic *ž* in Shina probably appeared due to the influence of the verb *paáž-* ‘to cook, bake’ (Berger 1998b: 45). See also Kogan 2019.

66. Balti ***phaḍa*** ‘bald’ (Sprigg 2002: 128), Purik ***phaṭa*** ‘*Glatze*’ (Zemp 2006: 32). Cf. Brokskat *phaṭa*, Shina (Drasi dialect) *phaṛaro*, Burushaski (Hunza, Yasin) *baṭa* ‘bald’. According to H. Berger (Berger 1998b: 44), the word is connected with Burushaski *baṭ* ‘skin’. See also Kogan 2019.

67. Balti ***phalu*** ‘fruit, necklace, bead’ (Sprigg 2002: 128). OIA *phala-* ‘fruit, seed of a fruit, spot on dice’, Hindi-Urdu, Nepali *phal* ‘fruit’, Kalasha *phălak*, Khowar *pholok*, Palula *phalŭṛī*, Shina *phal*, Kashmiri *phol* ‘grain’. It cannot be ruled out that the coexistence of the meanings ‘fruit’ and ‘bead’ in Balti is the result of double borrowing, the word *phalu* with the former meaning being an Urdu loan, whereas with the latter one it may have been adopted from some Dardic language where its prototype originally meant ‘grain’. Burushaski *phal ‘(Getreide-) Korn; ein kleines Stückchen (Fleisch)*’ (Berger 1998b: 320) must also have been borrowed from a Dardic source.

68. Balti ***-phoṭo*** in *mik-phoṭo* ‘eye-socket, orbit’ (Sprigg 2002: 115), *go-phoṭo* ‘skull, cranium’ (Sprigg 2002: 67)[11], Purik ***-photo*** in *mikphóto* ‘eye socket’ (Zemp 2018: 135). Cf. Shina *phọ̣t* ‘peel, rind’ (Bailey 1924: 157), *phóṭo* ‘*Maiskolben, Schale*’ (Berger 1998b: 334), Kalasha *phoṭ* ‘eggshells; skins or peels of fruit such as grapes, apples, peaches, figs, jujube, pea pods and nut meats’ (Trail & Cooper 1999: 242). Burushaski *phóṭo* ‘*Eichel (des Penis, wenn groß)*’ (Berger 1998b: 334) may also belong here.

69. Balti ***phuṭ*** ‘chaff’ (Sprigg 2002: 131). Cf. Kalasha *phoṭ* id. (Trail & Cooper 1999: 242), Shina *phọ̣t* ‘peel, rind’ (Bailey 1924: 157), *phóṭo* ‘*Maiskolben, Schale*’ (Berger 1998b: 334). Perhaps, an etymological doublet of the previous word.

70. Balti ***qaar*** ‘blanket’ (Sprigg 2002: 133). Cf. Burushaski *qáar* ‘*Schal, dünne Decke*’, Shina *káar* id. (Berger 1998b: 340).

71. Balti ***rat*** ‘flat, level’ (Sprigg 2002: 135), Purik ***rat*** ‘level’ (Zemp 2018: 939). Cf. Burushaski (Hunza, Nagir, Yasin) *rat* ‘*flach, eben, glatt*’ (Berger 1998b: 364). See also Kogan 2019.

72. Balti ***rkat*** ‘to cut down with a sword’ (Sprigg 2002: 140). Cf. PII **kart-* ‘to cut’ > OIA *kr̥ntati, kartati* ‘cuts’, Av. *kərəṇtaiti* ‘cuts’, *karəta-* ‘knife’, Hindi-Urdu *kattā* ‘curved knife’, *kattī* ‘sword, knife, dagger’, Kalasha *katār*, Khowar *kuter* ‘knife, dagger’, Palula *kaṭōro* ‘dagger’, Shina *khăṭa̯rų*, Brokskat *kaṭar, kuṭar* ‘knife’. See also Kogan 2019.

73. Balti ***sang*** ‘be light’ (Sprigg 2002: 147), Purik ***saŋsaŋ*** ‘bright; wide awake’ (Zemp 2018: 758). Cf. Burushaski *saṅ* ‘*Helligkeit*’, Shina *saṅ* id. (Berger 1998b: 374), Brokskat *saŋ* ‘light’. According to Berger, the Burushaski word is etymologically connected with *sa* ‘sun’. Tib. *(b)sang* ‘cleanse’ cited by Sprigg is semantically more distant.

74. Balti ***shingsher***, Purik ***shiṅsheř*** ‘Saturday’ (Sprigg 2002: 152; Bailey 1920: 41). Cf. Burushaski, Shina *śimśér*, Brokskat *šiŋseru* id., OIA *śanaiścara-* ‘the planet Saturn’, *śanaiścaravāra-* ‘Saturday’. See also **1**.

75. Balti ***shuguru***, Purik ***shŭkŭr*** ‘Friday’ (Sprigg 2002: 153; Bailey 1920: 38). Cf. Burushaski *śúkuro*, Shina *śúkur*, Brokskat *śúkuru* id., OIA *śukra-* ‘the planet Venus’, *śukravāra-* Friday’. See also **1**.

76. Purik ***sụ*** ‘nose’ (used as a synonym of *snamtshul* id. < Old Tibetan *sna-mtshul*) (Zemp 2018: 938). Cf. Burushaski *ṣũ ɛtʌs* ‘to smell (v.t.)’ (Lorimer 1938: 332), Shina *ṣũ thoĭki̯*, Indus Kohistani *ṣũ karʌ́tõ* id. (Bailey 1924; Zoller 2005: 391). See also Kogan 2019.

[11] Cf. Balti *mik* ‘eye’, *go* ‘head’.

77. Balti ***ṭangka***, ***ḍangga*** 'shelter for the night, room boarded all round' (Sprigg 2002: 183, 184), Purik ***ṭangka*** 'partially roofed daytime pen for animals; open or partially roofed courtyard on the upstairs of a house' (Norman 2010: 359). Cf. Burushaski *ḍaṅóo* '*oberirdischer Speicher, mit Tür zum Haus*' (Berger 1998b: 131), Shina *ḍǎṅō* 'native storehouse' (Bailey 1924). But cf. Tibetan *dang-ra* 'stable for cattle' (Jäschke 1881: 249). See also Kogan 2019.

78. Balti ***ṭek*** 'button' (Sprigg 2002: 184). Cf. Burushaski *ṭʌk*, Shina *ṭǎk*, Indus Kohistani *ṭikū*, Khowar, Kalasha *tak* id. See also Kogan 2019.

79. Balti ***thalo*** 'dish' (Sprigg 2002: 164), Purik ***thali*** 'plate' (Zemp 2018: 939). Cf. OIA *sthālī-*, *sthāla* 'earthen dish', Hindi-Urdu *thālī* 'dish', Shina *tāl-* 'iron tray', Brokskat *thali* 'metal plate', Kashmiri *thāl* 'metal platter'. Burushaski *tháal* '*Schüssel*' (Berger 1998b: 435) is an Indo-Iranian borrowing.

80. Balti ***thar rkyang*** 'lie down with arms and legs outstretched' (Sprigg 2002: 164), Purik ***thar*** 'wide (asunder)' (Zemp 2018: 757). The second component of the Balti compound verb is identical to Balti *rkyang* 'stretch (out)' (Sprigg 2002: 141). For the 1st component, cf. Burushaski *thar* '*offen, weit, unbegrenzt*', Shina *thar* id. (Berger 1998b: 438), *thŭroĭki* 'to open' (Bailey 1924: 166). According to Turner (1966: 789), the Shina words are etymologically related to OIA *starati* 'strews, covers', *stara-* 'layer'. However, the change **st* > *th* in Shina is attested in intervocalic but not word-initial position. Therefore, it seems more reasonable to connect the Shina lexemes with OIA *āstarati* 'spreads, covers', *āstara-* 'covering', *āstāra-* 'spreading'.

81. Balti ***ṭhing*** 'upright, straight up' (Sprigg 2002: 184). Cf. Burushaski *ṭiŋ* 'peak (of mountain), top (of hill); stretched up, erected', Khowar *ṭhing* 'high place; steep slope' (Lorimer 1938: 352). See also Kogan 2019.

82. Balti ***ṭhom-bos*** 'embankment, retaining wall (to prevent a sloping field from being washed away)' (Sprigg 2002: 184). Cf. Burushaski *ḍom* '*Erdhaufen*' (Berger 1998b: 133), *buśúṣ* '*Zaun aus Wacholderzweigen und Dornen um bebautes Land; Einzäunung zum Fangen von Wölfen*' (Berger 1998b: 65). The element *ṭhom-* of the Balti word seems to reflect the Burushaski prototype without the initial sonorization, seen in the modern Burushaski form. The reflection of Burushaski *buśúṣ* as *bos* in Balti is probably due to the restrictions imposed by the rules of Tibetan phonotactics.

83. Balti ***thup*** 'darkness' (Sprigg 2002: 166), Purik ***ṭhup*** 'dark(ness)' (Zemp 2018: 61). Cf. Burushaski *thap* 'night'; Shina (Chilasi dialect), Indus Kohistani *thʌph* 'dark; extinguished (light, fire)' (Zoller 2005: 113). Balti ***ṭhap-ṭhup*** 'dawn, dusk, twilight' (Sprigg 2002: 184) may belong here too. See also Kogan 2019.

84. Balti ***too*** 'iron plate (for making bread-cakes (chapaties)' (Sprigg 2002: 167), Purik ***to*** 'cooking plate' (Zemp 2018: 443). Cf. OIA *tāpaka-* 'cooking-stove', *tāpikā-* 'frying pan', Hindi-Urdu *tāvā* 'frying pan', Middle Persian *t'pk*, Classical Persian *tāba* 'frying pan', Shina, Khowar *tāo* 'griddle', Brokskat *to* 'iron plate for baking bread', Indus Kohistani *tʌv* 'metal plate for baking bread' (Zoller 2005: 233), Kashmiri *tāw* 'frying pan'. Burushaski *dau.o* 'griddle' (Lorimer 1938: 122) must be an old Indo-Iranian borrowing with the sonorization of the initial dental stop.

85. Balti ***ṭok*** 'swelling, bump on the head, small hill' (Sprigg 2002: 185), Purik ***ṭoqṭoq*** 'lump' (Zemp 2018: 926). Cf. Burushaski *ṭhóko* '*Klumpen, Ballen (von Zucker, Salz, Butter)*' (Berger 1998b: 452), Shina *ṭhọ̆kŭ* 'hill' (Bailey 1924), Indus Kohistani *ṭhokh* 'a clod (earth, salt)' (Zoller 2005: 220). See also Kogan 2019.

86. Balti ***tsandar***, Purik ***tsăndrăl*** 'Monday' (Sprigg 2002: 169; Bailey 1920: 40). Cf. Burushaski *ċandúra*, Shina *ċandúro*, Brokskat *candral*, OIA *candravāra-* id. See also **1**.

87. Balti ***tseenu*** 'pantaloons, trousers' (Sprigg 2002: 169). Maybe somehow related to Shina *tsanalŭ* (Gilgit dialect) 'native trousers' (Bailey 1924: 166), *tsʌléni* (Dras dialect) 'trousers' (Rajapurohit 2012: 99), Brokskat *canalu* 'trousers (pāyjāmā)'. Further etymological connections are not clear.

88. Balti ***xapxap*** 'burning, smarting, peppery' (Sprigg 2002: 175). Connection with Burushaski *qhap*, Shina *kháap* '*Zunder*' (Berger 1998b: 351), Indus Kohistani *khăp* 'tinder' (Zoller 2005: 125) seems possible.

89. Balti ***xar-a*** 'tree whose wood is hard', ***xar-xar*** 'rough, harsh' (Sprigg 2002: 175), Purik ***ʁarʁar*** 'rough' (Zemp 2018: 87). Cf. OIA *khara-* 'hard, sharp, pungent', Pali *khara-* 'hard, rough, cruel, sharp', Indus Kohistani *khʌrō* 'rough, hard' (Zoller 2005: 124), Classical Persian *χārā* 'a very hard stone, flint, rock', *χār* 'thorn'.

90. Balti ***xor*** 'move, slide out of place (walls only)' (Sprigg 2002: 176). Cf. Burushaski *ġōōr man-* '*(Mauer) langsam zusammenfallen*' (Berger 1998b: 170).

91. Balti ***xuṭu*** 'deep (things only; rare)' (Sprigg 2002: 177). Cf. Burushaski *ġuṭúm* '*tief, eingebuchtet*' (Berger 1998b: 183), Shina *gŭṭŭmŭ* 'deep' (Bailey 1924: 142).

92. Balti ***zan-zos*** 'wife, housewife, female partner', Purik ***zanzos*** 'wife; family' (Sprigg 2002: 180; Norman 2010: 842). Cf. OIA *jani-*, Av. *ǰaini-* 'woman, wife', Kashmiri *zən'* 'woman', Palula *jeeni* 'female person' (Liljegren, Haider 2011: 76), Indus Kohistani *zhāl* 'marriage' (Zoller 2005: 202). See also Kogan 2019.

Discussion

The majority of the above-listed Balti and Purik words (74 out of 92) have obvious or probable etymological parallels in Burushaski. Some of these 74 words, however, have cognates in neighboring Indo-Iranian (primarily, Dardic) languages too, and are etymologizable in them. Such lexical items must be considered Aryan loanwords common for Burushaski and some Northwestern Tibetan dialects. With this fact in mind, I put forward the following two equally obligatory criteria for identification of Burushaski borrowings in the above list: (i) presence of a probable cognate in Burushaski; (ii) absence of Indo-Iranian etymology.[12] Words strictly meeting these criteria number in total 46. The remainder of the list includes 41 apparent Indo-Iranian loans, two compounds consisting of a Burushaski and an Aryan part (see **56** and **57**), and three probable Wanderwörter (see **43, 52** and **87**), whose ultimate source is yet to be established. Technically, the Burushaski-Aryan compounds may be classified as belonging both to Indo-Iranian and Burushaski etymological stratum.

Based on the above figures, one can state a certain predominance of Burushaski elements in the loan vocabulary peculiar to Balti and Purik, and it can thus be argued that the hypothesis put forward in (Kogan 2019) receives additional corroboration.

Regarding the Aryan loanwords on the list, it should be noted that they seem to have been borrowed from more than one source. As stated above, names of the days of the week (see **1, 3, 13, 14, 74, 75, 86**) were probably adopted from some Indo-Aryan lect.[13] Most of the remaining 34 Indo-Iranian loans remarkably have etymological parallels, inter alia, in at least one of the two closely related Dardic languages: Shina and Brokskat. As many as 14 of them

[12] The 2nd criterion needs some clarification. Aryan languages of the area under study were to a greater or lesser extent influenced by Burushaski, and this influence sometimes considerably affected their lexicon. Given this fact, Indo-Iranian etymology should not be understood as simply presence of parallels in any Indo-Iranian lects. Rather it means the existence of cognates in Old and Middle Indo-Iranian languages, and/or in modern Aryan languages other than those that were in close contact with Burushaski, i.e. Shina, Khowar, Indus Kohistani and Wakhi.

[13] Besides the seven words mentioned here, the Indo-Aryan loan vocabulary in Balti and Purik also includes recent borrowings from Urdu. Cf., e.g. *balo* 'earring' (Sprigg 2002: 25) < Urdu *bālā* 'large earring', *gan* 'sledgehammer' (Sprigg 2002: 57) < Urdu *ghan* id. However, such lexemes do not present any problems as regards etymology and historical phonology, and therefore are not discussed in the present work.

(see **9, 10, 22, 29, 44, 47, 48, 50, 54, 60, 63, 72, 79, 84**) find cognates in both Brokskat and Shina, 11 (see **4, 8, 11, 18, 27, 42, 53, 59, 61, 67, 80**) in Shina only, and 2 (see **19, 30**) in Brokskat only. The Aryan part of one of the two Burushaski-Aryan compounds (see **57**) has both a Shina and a Brokskat cognate. No other modern Indo-Iranian language shows a comparable number of parallels to the above listed words, and this fact can hardly be accidental. Given the fact that Shina and especially Brokskat are still poorly documented, it seems more than likely that the real figures are even higher than those cited here.

In historical phonology, some of the Aryan borrowings on the above list possess certain features clearly distinguishing them from Indo-Iranian lexical elements, common for Balti, Purik and Ladakhi. E.g., in **18** and **19** the PIE consonant cluster **sk*' (> Proto-Aryan **sć*) is reflected as *ch* (the voiceless aspirated palatal affricate), whereas in loanwords found in all Northwestern Tibetan varieties including Ladakhi, the regular reflex is the dental affricate *tsh* (Kogan 2020).[14] The reflection of PIE **sk*' identical to the one attested for the words **18** and **19**, is characteristic of Indo-Aryan and a number of Dardic languages including Shina and Brokskat. The same holds true for the change **st* > *th* in historically intervocalic position, assumed for **80**. This phonological development is peculiar to Eastern Dardic, as well as Middle and New Indian. Thus, for **18, 19** and **80** both Indo-Aryan and Dardic origin is theoretically possible, but the fact that exact cognates of the latter two loans are found only in Brokskat and Shina respectively, points to Eastern Dardic as their most likely source.

All the above suggests that at least a part of Aryan loans in Balti and Purik definitely could not have been borrowed from the same donor-language as Indo-Iranian loan vocabulary in Ladakhi. The latter, as was demonstrated in Kogan 2019, 2020, was adopted from a lect which, although it belongs to the Dardic group, cannot be classified with its Eastern sub-branch.

The question of whether the above-listed Indo-Iranian loanwords were borrowed from Shina or Brokskat or both, is an interesting issue for future research. However, certain facts already known to us may offer some hints as to the probable answer. In Brokskat one can detect historical-phonological innovations that are unusual for Dardic languages but have parallels in Balti. An obvious example of such an innovation is the change **n* > *ŋ* in the palatalizing position: *ŋilo* 'blue', cf. OIA *nīla-* 'dark blue, dark green', Kati *nīlə*, Prasun *nīl* 'blue', Kalasha *nīḷa*, Palula *nīlo*, Shina *nīlų*, Kashmiri *nyūl* 'green, blue', Hindi-Urdu *nīlā* 'blue'; *ŋin-dis* 'to weed', cf. OIA *nirdātṛ-* 'weeder, reaper', Pali *niḍḍāyati* 'cuts out, weeds', Wakhi *nɛždun*, Shina *nīn*, Khowar *niḷáṅg*, Kalasha *nəhŕún*, 'weeding', Kashmiri *nĕnd* 'weeding a rice-field', Hindi-Urdu *nirānā* 'to weed'; *ŋiliẓis* 'to hide', cf. OIA *nilīyatē* 'hides oneself, disappears', Shina *nilížóĭkį* 'to hide oneself'. In Balti the Old Tibetan palatalized nasal *ny* changes to the dorsal nasal *ng*:[15] *ngespa* 'sin' < Old Tibetan *nyes pa*, *ngima* 'sun' < Old Tibetan *nyi ma*, *ngis* '2' < Old Tibetan *gnyis*, *ngishu* '20' < Old Tibetan *nyi shu*; *ngya* 'fish' < Old Tibetan *nya*, *ngyamngyam* 'of the same age' < Old Tibetan *mnyam pa* 'equal, same'.

The intervocalic change *m* > *ŋ (ng)*, regular for Balti[16] (cf. *bo-ngo* 'daughter, girl' < Old Tibetan *bu mo*, *bya-nga* 'sand' < Old Tibetan *bye ma*, *lo-nga* 'leaf' < Old Tibetan *lo ma*, *o-nga* 'milk' < Old Tibetan *o ma*) is also attested in Brokskat. There it is, however, found in one single word: *čiŋar* 'iron', cf. Shina *čiměr*, Palula *čímar*, Kashmiri *cɨmur*, Torwali *čimu*, Bashkarik *čimɛr*, Khowar *čúmur*, Gawar-Bati *ċimár*, Pashai *čimă̆r*, Burushaski *ćímar, ćhumár* id.

[14] Of particular interest in this connection is the etymology **19**. The Indo-Iranian source word of Balti *chaḍu* 'vile, unclean, unholy, abominable, filthy, foul, impure, lewd' appears to have been borrowed into Ladakhi and Purik as well, where, however, the respective loanwords have initial dental *tsh.*

[15] All the Balti examples are drawn from Sprigg 2002.

[16] According to Backstrom (1994: 34), this change took place "intervocalically when the following vowel was word final and the consonant preceding the preceding vowel was not *ɲ* or *ŋ*".

It seems highly probable that the two aforementioned historical-phonological isoglosses, shared by Balti and Brokskat, are contact-induced. Remarkably, both sound changes took place in Balti but not in Ladakhi and Purik[17] with which Brokskat is in contact nowadays. This fact suggests that in the past Brokskat-speakers (the Brokpas) must have lived somewhere in the present-day Baltistan or in closest proximity to it,[18] and it is there that they encountered the speakers of Northwestern Tibetan. If this be the case, it is only natural that Balti vocabulary was to a certain extent influenced by Brokskat.

Abbreviations

Av. — Avestan; OIA — Old Indo-Aryan; PIE — Proto-Indo-European; PII — Proto-Indo-Iranian; Tib. — Tibetan.

References

Backstrom, Peter Charles. 1994. *A phonological reconstruction of Proto-Western Tibetan. MA thesis*. Arlington: The University of Texas at Arlington.

Bailey, T. Grahame. 1920. *Linguistic Studies from the Himalayas* (*Asiatic Society Monographs 18*). London: Royal Asiatic Society.

Bailey, T. Grahame. 1924. *Grammar of the Shina language* (*"Royal Asiatic Society, Prize Publication Fund", vol. VIII*). London: Royal Asiatic Society.

Bailey, Harold W. 1979. *Dictionary of Khotan Saka.* Cambridge University Press.

Bashir, Elena. 2022. *A digital Khowar-English dictionary with audio: 1,000 words / compiled by Elena Bashir with Maula Nigah and Rahmat Karim Baig*. *2nd ed*. Chicago: Digital Dictionaries of South Asia, University of Chicago. Available online at: https://dsal.uchicago.edu/dictionaries/khowar-2nd/A_Digital_Khowar-English_Dictionary_with_Audio-second_edition.pdf (accessed 11.06.2025).

Berger, Hermann. 1998a. *Die Burushaski-Sprache von Hunza und Nager. Teil I. Grammatik.* Wiesbaden: Harrassowitz Verlag.

Berger, Hermann. 1998b. *Die Burushaski-Sprache von Hunza und Nager. Teil III. Wörterbuch Burushaski-Deutsch, Deutsch-Burushaski. Unter Mitarbeit von Nasiruddin Hunzai.* Wiesbaden: Harrassowitz Verlag.

Francke, August H. 1906. The Dards at Khalatse in Western Tibet. *Memoirs of the Asiatic Society of Bengal* 1(19): 413–419.

Grierson, George A. 1915–32. *A Dictionary of the Kashmiri Language. Vol. I–IV* (*"Bibliotheca Indica". Work No. 229*). Calcutta: Royal Asiatic Society of Bengal.

Jäschke, Heinrich A. 1881. *A Tibetan-English dictionary with special reference to the prevailing dialects.* London: Routledge & Kegan Paul.

Kogan, Anton. 2019. On possible Dardic and Burushaski influence on some Northwestern Tibetan dialects. *Journal of Language Relationship* 17(3–4): 263–284.

Kogan, Anton. 2020. Notes on the historical phonology of Indo-Iranian loanwords in Northwestern Tibetan dialects. *Journal of Language Relationship* 18(3–4): 261–275.

Kogan, Anton. 2024. On the etymological stratification of borrowed Indo-Iranian vocabulary in Burushaski. *Journal of Language Relationship* 22(1–2): 21–42.

Kogan, Anton. 2025. Aryan loanwords in Burushaski as a data source for the reconstruction of language contact in the Upper Indus basin. *Journal of Language Relationship* 23(1–2): 20–28.

Liljegren, Henrik, Naseem Haider. 2011. *Palula vocabulary.* Islamabad: Forum for Language Initiatives.

[17] In Purik, however, there is an interesting phonetic phenomenon, which may well have existed earlier also in Balti and prepared the ground for the regular phonological change *ny* > *ng* there. The Purik palatal nasal has several pronunciational variants, one of which is the palatalized dorsal nasal (Zemp 2006: 37).

[18] In this connection, it is worth remembering that according to the hypothesis put forward by August Hermann Francke, the Brokpas are the remnant of the pre-Tibetan population of Ladakh (Francke 1906). As I have shown in Kogan 2019, this viewpoint cannot be upheld.

Lorimer, David L. R. 1938. *The Burushaski language. Vol. III. Vocabularies and index.* Oslo: H. Aschehoug & Co. (W. Nygaard).

Morgenstierne, Georg. 1940. Notes on Bashkarik. *Acta Orientalia* XVIII(3/4): 206–257.

Morgenstierne, Georg. 1942. Notes on Dameli, a Kafir-Dardic Language of Chitral. *Norsk Tidskrift for Sprogvidenskap* XII: 115–198.

Morgenstierne, Georg. 1945. Notes on Burushaski Phonology. *Norsk Tidsskrift for Sprogvidenskap* 13: 61–95.

Norman, Rebecca. 2010. *A Dictionary of the Language Spoken by Ladakhis.* Ms. (unpublished draft).

Norman, Rebecca. 2019. *Rang-Skat: A Ladakhi-English Dictionary.* Leh: Melong Publications.

Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch.* Bern / München: Francke Verlag.

Rajapurohit, B. B. 2012. *Grammar of Shina Language and Vocabulary (Based on the dialect spoken around Dras).* Mysore: Central Institute of Indian Languages.

Ramaswami, Natarajan. 1989. *Brokskat-Urdu-Hindi-English Dictionary.* Mysore: Central Institute of Indian Languages.

Sharma, D. D. 1998. *Tribal Languages of Ladakh, Part One. A concise Grammar and Dictionary of Brok-skad.* New Delhi: Mittal Publications.

Sloan, Mohammad Ismail. 1981. *Khowar-English dictionary.* Peshawar: Printing Corporation of Frontier Ltd.

Sprigg, Richard K. 2002. *Balti-English, English-Balti dictionary.* New York / London: Routledge.

Steblin-Kamenskiy, Ivan M. 1999. *Etimologicheskij slovar' vaxanskogo jazyka.* Sankt-Peterburg: Peterburgskoje vostokovedenije.

Strand, Richard F. 2006. *Khowar lexicon.* Available online at: https://nuristan.info/lngFrameL.html (accessed 11.06.2025).

Trail, Ronald L., Gregory R. Cooper. 1999. *Kalasha Dictionary with English and Urdu.* Islamabad: National Institute of Pakistan Studies, Quaidi-Azam University, Summer Institute of Linguistics.

Turner, Ralph L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages.* Oxford University Press.

Yang, Xing-Yan, Allah Rakha, Wei Chen, Juzhi Hou, Xue-Bin Qi, Quan-Kuan

Shen, Shan-Shan Dai, Sulaiman Xierzhatijiang, Tojiddin Najmudinov, Elena Manilova, Behruz Khudoidodov, Xi Chen, Wei-Kang Yang, Atif Adnan, Ruo-Han Zhao, Yong-Gang Yao, Bing Su, Min-Sheng Peng, Ya-Ping Zhang. 2020. Tracing the Genetic Legacy of the Tibetan Empire in the Balti. *Molecular Biology and Evolution* 38(4): 1529–1536.

Zemp, Marius. 2006. *Synchronic and diachronic phonology of the Tibetan dialect of Kargil.* Bern: Lizentiatsarbeit im Fach Historisch Vergleichende Sprachwissenschaft.

Zemp, Marius. 2018. *A grammar of Purik Tibetan.* Leiden / Boston: Brill.

Zoller, Claus P. 2005. *A Grammar and Dictionary of Indus Kohistani. Volume 1: Dictionary.* Berlin / New York: Mouton de Gruyter.

А. И. Коган. Этимологические заметки о заимствованной лексике в тибетских диалектах Балтистана и Каргила

Результаты недавних исследований дали основания полагать, что большая часть лексических заимствований в тибетских диалектах, распространённых в Каргиле и Балтистане, была усвоена из ранней формы языка бурушаски. Данный вывод, однако, является предварительным и нуждается в уточнении. Попытка уточнить его делается в настоящей статье. Автор демонстрирует, что хотя слова бурушаски действительно преобладают в иноязычной лексике балти и пурик, оба эти диалекта содержат значительное число заимствований из ряда индоиранских источников. Данные исторической фонетики, а также анализ лексических изоглосс позволяет утверждать, что в число этих источников входили дардские языки шина и брокскат.

Ключевые слова: тибетские диалекты; балти язык; пурик язык; бурушаски язык; дардские языки; языковые контакты; лексические заимствования.

Pavel B. Lurje
State Hermitage Museum; lurje@panchkath.ru

Once more on Sogdian-Yaghnobi linguistic relations

The article deals with the questions of interrelation of two Eastern Iranian languages which are particularly close to each other: Sogdian and Yaghnobi. After a brief introduction on the trivia of extinct literary Sogdian of the 1st millennium CE and on the living tongue of Yaghnob in the mountains of Central Tajikistan, I review the previous scholarship on their similarities and dissimilarities. At the end of this chapter a special paragraph is devoted to the reasonability of bilateral comparison of Sogdian and Yaghnobi. Chapter 3 is devoted to the key differences between these two languages and the probable date of their separation, these features being the Sogdian "Rhythmic law", the position of the augment in the imperfect, the third plural and first plural ending, *y*-umlaut diphthongs and the development of the clusters **δr* and **θr*. I advocate dating the key isoglosses to the 3rd century BCE — 3rd century CE: after the formation of Sogdian writing system and Greek influence and before the "Ancient Letters", the first substantial group of Sogdian texts (before 313 CE). In the next chapter I deal with shared innovations of the two tongues that happened after the separation of proto-languages. Many of these features were triggered by close contact with Turkic speakers. Chapter 5 focuses on some cultural words and extra-linguistic, chiefly archaeological data. I argue that proto-Yaghnobis played a minimal role in the development of Sogdian trade and cosmopolitan high culture. In the chapter 6 I consider literary Sogdian a koine used by speakers of different closely related dialects. I try to show that some isoglosses visible in certain Sogdian texts were rooted in the dialectal background of their authors, sometimes perhaps the dialects of eastern Sogdiana and Ustrushana, which were closer to Yaghnobi than Samarkand standard was, and I conclude that proto-Yaghnobi had its own (perhaps, quite special) position within the dialects of Sogdian city-states. The article is followed by a list of some 150 Sogdian — Yaghnobi lexical correspondences which came to light after the publication of the dictionary in the book Jagnobskie teksty (1957). They are partially based on unpublished archival materials of B. A. Alborov and M. N. Bogoljubov, in a half of cases they were first noticed by this author. The indexes are also available in the form of online supplementary materials.

Keywords: close language relation; Eastern Iranian; Sogdian language; Yaghnobi language; isoglosses; absolute dating of language features.

1. Generalities

Sogdian is an extinct Eastern Iranian language of the Middle period. It is known from written texts in Sogdian (mostly translated compositions of Buddhists, Manichaeans and Christians) as well as a set of indirect attestations (*Nebenüberlieferung*) in the neighboring textual traditions, such as Chinese or Arabic. The language is known from the texts dated to almost the whole of the first millennium CE, which have been mostly found outside Sogdiana proper (the valleys of Zarafshan and Kashka-darya rivers in the parts of present-day Uzbekistan, Tajikistan and Turkmenistan), most significantly in Dunhuang and Turfan in North-Western China, where Sogdians established their trading colonies. The language is generally written in three writing systems: the so-called National Sogdian script (all Buddhist, Zoroastrian (?) and secular texts as well as a good half of Manichaean fragments and several Christian ones), Manichaean script and Syriac script for Manichaean and Christian texts, respectively. All these

writing systems ultimately derive from the Aramaic quasi alphabet (*abjad*), and given the ambiguity of vocalization of the basically consonantal script, they are published in unvocalized transliteration and not in transcription; the records of Sogdian in Indian Brāhmī script are quite limited. The Sogdian texts are relatively numerous: my impression is that they are fewer than the surviving Middle Persian and Khotanese writings but more numerous than those in Parthian, Bactrian, Chorasmian, Tumshuqese, Ṭabarī and Alanic: the other Middle Iranian languages known to us so far.

The translated Sogdian texts are very well understood; some words or phrases in the original, secular compositions, and in inscriptions especially, remain hard to explain, which is partially due to the fact that they are usually written in an ambiguous shorthand script. The Sogdian dictionary (Gharib 1995) includes the lexicon of the main texts published before 1990s, it was more recently complemented by special dictionaries of Manichaean and Christian texts (Sims-Williams & Durkin-Meisterernst 2012, 2022; Sims-Williams 2016a, 2021a). The recent reference grammars are relatively brief book chapters (Sims-Williams 1989a; Yoshida 2009a),[1] so the older language descriptions (especially Gershevitch 1954) as well as a number of individual journal articles and even commentaries to text editions are to be consulted as well.

Few Sogdian texts can be dated exactly; the chronology is largely based on the palaeography, orthography, historical and archaeological contexts. The history of development of Sogdian can be outlined as follows. The earliest (and shortest) texts are coin legends (since the 2nd century BCE?) and some very short inscriptions. The texts on bricks from Kul-tobe in southern Kazakhstan perhaps date to the first centuries of common era. More extensive and better understood are the so-called "Ancient Letters" from a watch-tower in Dunhuang predating 313 CE, recently published together in a handy volume (Sims-Williams 2023); the visitor inscriptions of Upper Indus valley (Sims-Williams 1989b, 1992) are perhaps slightly later. There are several inscription and secular texts of the 6th century from Mongolia, Xian and Turfan areas. The better-preserved Buddhist texts come mainly from Dunhuang (less from Turfan) and are supposedly dated to the 6th–8th centuries. The documents from Mount Mugh, found not so far from Yaghnob, predate 722. These texts are written in a partially archaizing orthography, which reflects the older stage of the language. For example, the very widespread Sogdian suffix coming from Old Iranian **-aka*, and by then pronounced as */-e/*, could be spelt as <*-k*>, <*-'k*>, <*-'y*> or <*-y*>. The historical and archaeological data suggest that the climax of Sogdian culture falls in the 7th — early 8th centuries, too.

The Manichaean texts, found almost exclusively in the Turfan area, are supposedly dated to the 8th — very early 11th century. Those written in the National script display virtually the same orthographic features as the Buddhist and secular ones. The ones in the Manichaean script are free of ideograms and historical spellings and are more accurate in rendering the actual pronunciation of the time. The Christian texts of the 9th–10th century from Shüipang (Bulayïq) monastery near Turfan demonstrate further simplification of the language reflecting what one may call colloquial pronunciation; the manuscript E5 seems to document the latest stage. The Syriac orthography permitted a better rendering of vowels through not only the *matres lectionis* ', *w* and *y* but also diacritic points.

The Sogdian fragments in a variant of the Brāhmī script are few but nevertheless quite important for understanding the actual pronunciation in Turfan around the late 9th to 10th century (Maue & Sims-Williams 2024). The latest texts in the 'National' script are the so-called 'Turko-Sogdian' ones from Dunhuang and Turfan; the inscriptions from Semirechie (the latest

[1] The Japanese 'Lectures on Sogdian grammar' by Yutaka Yoshida (2022) came to my notice too late to be included in full.

dated one is from 1026 CE, exactly 1000 years ago! — Livshits 2015: 290) have many similar features. Though written basically in Sogdian, they show a simplified grammar and profound influence from Uighur, which concerns not only personal names and lexica, but also phraseology, orthography, and even the sound system (Sims-Williams 2008; Yoshida 2009b; Sims-Williams & Hamilton 2015).

Yaghnobi is a living Eastern Iranian language for perhaps 10 000 to 20 000 native speakers (Panaino 2013a: 463; Mirzozoda apud Novák 2018: 27), but it does not have a literary tradition. All the speakers are bilingual with Tajik and some are tri- or quadrilingual with Uzbek and Russian. Yaghnob is an elevated mountain valley on the left-hand tributary of upper Zaraf-shan in central Tajikistan, which is difficult to access and poor in amenities. All the surrounding areas (including the uppermost and lower parts of the Yaghnob valley) are inhabited by Tajiks. In some publications, even the recent ones, Yaghnobi is classified as one of the Pamiri languages (Wendtland 2009). There are more than 200 km as a crow fly between Yaghnob and Yazghulom, the closest Pamiri speaking area, the distance that could be crossed in not less than a dozen days in the pre-modern period. However, since Pamiri is an areal term of heterogenous Iranian tongues, the ecologies and lifestyle of Yaghnob and Pamir are similar, and some lexica is shared (Novák 2013: 214),[2] this designation can be justified at least partially.

The inhabitants of the Yaghnob valley are agriculturalists cultivating a very limited variety of crops on poor soils in the cold climate and practicing transhumance cattle breeding. All Yaghnobis are Sunny Muslims and are famed for their zeal among the neighbours. The scarce limit of arable land in the valley pressed many Yaghnobis to move outside, to Varzob to the south and Ghonchi to the north; in 1970s the Soviet government forcefully resettled Yaghnobis to the cotton fields of Zafarobod district near Khujand (Loy 2005), but since late 1980s, some of them decided to return to their homeland (with no roads, electricity, shops, schools or hospitals). Nowadays, one half of the older villages are partially populated, at least in summer (many people, especially those retired and children, prefer to spend summer in Yaghnob and return to Zafarobod, Dushanbe or other places in winter). Presently, the government of Tajikistan and international humanitarian foundations make efforts to establish a natural and ethnographic reserve in Yaghnob.

The resettlement and vulnerability to modern life notwithstanding, and contrary to pessimistic predictions from the earliest descriptions on (Akimbetov 1881, Panaino 2013a: 465), the Yaghnobi language demonstrates an outstanding vitality, and in 2008 I was able to collect folk-tales with a handful of previously unattested non-Tajik words in the text (Lurje 2010: 15–16, 2015b).[3] The language is relatively well studied, beginning with the pioneering description of Alexander Kuhn made during the so-called 'Iskandarkul expedition' of 1870 and thereafter in Samarkand.[4] The main descriptions are Khromov 1972, 1987; Bielmeier 1989; Novák 2010: 220–247. The main dictionaries and glossaries have been prepared by V. A. Livshits and A. K. Pisarchik (in Andreev & Peshchereva 1957), A. L. Khromov (1972), Sayfiddin Mirzozoda (2008). More recently a Yaghnobi — Czech dictionary, aggregating all the previous publications with some novel material and using a uniform transcription has been published (Novák 2010).

Despite the small number of speakers, Yaghnobi has a dialect division. The principal ones are Western and Eastern dialects. The most significant differences between them (Khromov

[2] Cf. also many terms discussed by Filippone (2013).

[3] "Having survived this long, it seems quite likely that Yaghnobi will continue to be spoken by bi- and trilingual speakers who will use Yaghnobi at home" (Durkin-Meisterenst 2008: 415)

[4] See Iskander-t'ur'a 1881. The photographs of three Yaghnobis who were taken to Samarkand for linguistic records can be found in the 'Turkestan album' (Kun 1871: I, fol. 20).

1972: 97–105) include (1) the articulation of inherited **θ*: eastern *s*, western *t*; (2) the realisation of **ā* + *y*-umlaut diphthongs: eastern *ɛ*, western *ai*; (3) third person singular present ending: eastern *-či*, western *-tišt*; (4) first person plural imperfect and simple preterit ending: eastern *-īm*, western *-ōm* (Novák 2013: 27), as well as certain lexical features. There is no clear borderline between the Eastern and Western dialects and the isoglosses cut across the valley in different spots, so one can talk of transitional dialects and a variety of the Kul side-valley (Khromov 1972: 97–105). A peculiar speech register, usually called "secret language" has been encountered many times (Klimchitskij 1938); I would rather classify it as a sort of language game, not unlike North European kennings, spiced with some humour.[5]

Over the 150 years of documentation of Yaghnobi some changes can be observed. The area covered by the language in the valley apparently did not shrink.[6] In the early records the cluster *ft* was sometimes rendered as *vd*: *xšīvd*, now *xᵘšíft* 'milk' and *xt* as *γt*: *piraγta* for *piráxta*, *ūγta* for *úxta* (Bogoljubov 1956: 121; 888), some words (e. g. many inherited terms for colours) fell out of use. The preservation of the inherited *θ* in some early records cannot be proven (Bogoljubov 1956: 87–88).

However, as Rüdiger Schmitt remarked, 'die Erforschung dieser Sprache [d.h. Yaghnobi — PL] und insbesondere ihres Wortschatzes stand leider nicht unter einem günsteren Stern' (Schmitt 2010: 95). Many important works remained unpublished, including *Jagnobskie étjudy* of Carl Salemann,[7] the field studies of Robert Gauthiot of 1913, the recordings of B. A. Alborov of early 1930s (Lurje 2010: 12–15; Filippone 2013: 635), the doctoral dissertation of M. N. Bogoljubov (Bogoljubov 1956), many text records of Yu. N. Lebedeva of 1955–56, the university diploma of V. S. Nishanov on Yaghnobi syntax (1997). Even the recent dissertation of Ľubomír Novák and his article on historical phonology of Yaghnobi and Sogdian remain unpublished (Novák 2013, 2014).[8]

For this article, apart from published material, I had an opportunity to utilize the somewhat amateurish, but very important lexical data of B. A. Alborov, the brilliant dissertation of M. N. Bogoljubov, as well as my field records of the trip to Yaghnob in 2008 and some stray notes made afterwards in Tajikistan in the middle of intense archaeological work. I am grateful to Zarinæ Albortæ, the niece of Alborov, who told me of the existence of his manuscript, to Arsenij P. Vydrin of the Institute of linguistic studies, who made the copy for me in Vladikavkaz, and to my Hermitage colleague Alexej Mikhajlovich Bogoljubov, who provided me with a copy of his late father's dissertation in 2011. I received important advice from Nicholas Sims-Williams, Yutaka Yoshida, Antonio Panaino, Arsenij Vydrin, Adam Benkato, and especially Ilya Yakubovich. Of course, all the responsibility for this paper lies on its author.

---

[5] An unnoticed curious example comes from B. A. Alborov's records (Alborov 1934: 53): Tajik *mullo* is glossed (in modernized transcription) as *γuš čak dórna*, literally "(the one who) grabs (by) ear". One should keep in mind that a village mullah was also normally a teacher, and a teacher used to fiddle with ears of his pupils.

[6] Of course, toponymy suggests that dialects close to Yaghnobi were widespread in the whole valley and around it (Khromov 1966b). Cf., e. g., the village and cave *Makševat*/*Maxševad* on Iskandar-darya, which evidently comes from Yaghnobi *mákša* 'calf'; the micro-toponymy of Marγēb in the Tajik-speaking middle course of Yaghnob river is mostly Yaghnobi in origin (Lurje 2010); Bogoljubov noticed an interesting dialect variation between toponyms *Muštif* in Yaghnob and *Muždif* in Matcha (Bogoljubov 1956: 122–123).

[7] The data were used by Geiger (1898) and the first part of the glossary was published by Émile Benveniste (1955). During his second trip to Central Asia in 1908, Salemann was accompanied by A. A. Staël von Holstein, who made photographs and recordings of Yaghnobi and other local languages on phonograph cylinders. The cylinders and photos are kept in the archive of the Institute of Oriental Manuscripts in Petersburg.

[8] Both are available on the Internet: https://independent.academia.edu/ʟubomírɴovák; https://www.researchgate.net/profile/Lubomir-Novak (accessed on 02.02.2026).

2. Previous studies on the relations of Sogdian and Yaghnobi

Yaghnobi became known to the Western scholarship more than three decades before the first Sogdian texts came to light, just when the area was incorporated into emerging Russian Turkestan in 1870. The first, often amateurish records of the language, which was immediately recognized as an (Indo-)Iranian vernacular, were put together by Carl Salemann. Using his unpublished *Jagnobskie étjudy*, Wilhelm Geiger wrote the description of the language in the then-standard reference book *Grundriss der iranischen Philologie* on the eve of the 20th century (Geiger 1898).

When the first European expeditions started to uncover medieval texts in Khotan, Turfan and Dunhuang and to transfer them to the libraries in their countries on the eve of 20th century, the description of Yaghnobi was at the desks of the first decipherers. F. C. Andreas, who was the first to identify the "Pahlavi dialect" as Sogdian, noticed its similarity to Yaghnobi, and announced his discovery in the letters to F. W. K. Müller and C. Salemann (see now Sims-Williams 2022: 486–487). Some Yaghnobi forms (quotes as Pāmir-Dialekte) were provided in the first edition of Sogdian (= "Pahlavi-Dialekt") fragments by Müller (Müller 1904: 110), and Salemann used extensively the Yaghnobi data in his reedition of first Sogdian texts published by Sachau (Salemann 1907).

It was Andreas himself, however, who realized very soon that Yaghnobi, or Neusoghdisch, cannot be a direct continuation of what he called Mittelsoghdisch in any of its known forms (Andreas 1910: 307).[9] Although he did not explain his reasoning, he could not help noticing that the development *$\theta r > š$, which was his key argument for identifying Sogdian (apud Müller 1908: 2–3), is alien to Yaghnobi.

Robert Gauthiot, the pioneer of Sogdian studies in France, undertook a research trip to Yaghnob with Heinrich Junker in Summer 1913. The untimely death on the battleground during World War I prevented him from publishing the wealth of his data. In a short report he noticed that the language of Samarkand which extended as far as China could not be a parent of a humble Montagnard Yaghnobi tongue and that the two languages diverge in many points. Moreover, even the more archaic (so-called) Buddhist Sogdian could not be identical to ancient Yaghnobi, but rather was similar to it, and the more developed Christian Sogdian was more divergent. It was interesting to study, Gauthiot remarked, whether they are two different languages or dialects of a single one (Gauthiot 1913: 672, 678).[10]

---

[9] "Diesen [Yaghnōbī — PL] wird man aber, soweit sich das schon jetzt beurteilen läßt, nicht als den Abkömmling eines der beiden uns erhaltenen mittelsoghdischen Dialekte ansehen dürfen"

[10] "Il n'y avait pas de doute que l'humble parler montagnard d'aujourd'hui et la langue répandue au moyen âge de Samarqand jusqu'en Chine ne fussent parentes; mais il y avait intérêt à rechercher si l'intelligence de l'une pouvait être facilitée par la connaissance de l'autre et si l'on pouvait parler du yagnôbî et du sogdien comme de deux dialectes alliés de très près, ou de deux parlers d'une même langue"; "Que, sur bien des points, le sogdien et le yagnôbi aient divergé, cela n'a rien d'étonnant. Le sogdien d'écriture sogdienne, souvent appelé à tort bouddhique, qui est le plus archaïque, n'apparaît lui-même pas comme identique à la forme ancienne du yagnôbi, mais comme très voisine. A plus forte raison, le sogdien dit « chrétien » diverge-t-il, lui qui est plus évolué".

In his earlier paper, Gauthiot seconded the opinion of Andreas (Gauthiot 1911: 139). Although he noticed the difference in development of *θr (1911: 145–148), the similarity of Sogdian *wγwšw [wxwšw]* and Yaghnobi *uxš* 'six' led him to the conclusion that the difference of Buddhist Sogdian and Yaghnobi is largely chronological and both preserve archaic features (Gauthiot 1911: 159–160).

In the first part of his Sogdian grammar, Gauthiot noted that Yaghnobi shares most of the features of Sogdian, but not in the form attested in the documents from China, which means that Yaghnobi is a descendant of a Sogdian dialect. In particular, he observes the different development of *θr, the absence of a prothetic vowel in Yaghnobi, the similarity in the development of *$\underset{\circ}{r}$ (Gauthiot 1914: 25, 42–45, 90–95) and several other important details.

Paul Tedesco, having discovered the Sogdian "Rhythmic law" and the difference between Sogdian "light" and "heavy" stems, immediately noticed that the Yaghnobi inflexion is identical with that of the Sogdian heavy stems (Tedesco 1926: 123 and elsewhere). Some scholars see it as the main difference between the two languages (Bielmeier 1989: 480). In an earlier paper, Tedesco considered the difference of 3Pl *-ant* in Sogdian against *-or* in Yaghnobi as the most significant feature among others (Tedesco 1923: 294–295).

In his article on Sogdo-Yaghnobi correspondences, S. I. Klimchitskij noticed the similarity of potential constructions in the two languages and added 38 unnoticed lexical pairs (Klimchitskij 1937). The linguistic relationship between Sogdian and Yaghnobi constituted a significant topic in the dissertation of M. N. Bogoljubov (Bogoljubov 1956: 148–149 et passim); happily this part of the text was published, albeit in an abridged form (Bogoljubov 1957a). Although Bogoljubov called Yaghnobi "Neo-Sogdian" (новосогдийский) in brackets in the title of his dissertation, he was far from equating the two tongues. In his view, the key differences are: Sogdian *δ* and Yaghnobi *d* vs. Old Iranian **d* (Bogoljubov argues that in Proto-Yaghnobi, as in a part of eastern Sogdian dialects, there was no fricativization of historical **d*); the difference in the rendering of **θr*: *θr* or *š* in Sogdian vs. initial *tir-/sar-* and non-initial *l* in Yaghnobi; the lack of the development **dr>δr>ž* in Yaghnobi; the lengthening of *-a-* before *-rt*. Bogoliubov's observation on the voicing assimilation of historical preverbs in Yaghnobi is also very important. He thought that Yaghnobi originated from an eastern territorial dialect of Sogdian. In the *avtoreferat*, Bogoljubov adds the difference in the position of the augment and the 3pl ending *-or* in Yaghnobi, as well as the difference in some core lexical items (Bogoljubov 1957b: 19–20, 30).

The same year was marked by the publication of Yaghnobi texts collected by M. S. Andreev and E. M. Peshchereva in 1924 and 1927 with the appendix of a dictionary prepared by Andreev, V. A. Livshits and A. K. Pisarchik; the dictionary occupies almost one half of the book (Andreev & Peshchereva 1957: 225–390). It was Livshits who added brief etymological notes to the dictionary, and most of them are on Sogdian, with some 500 forms provided (see Index A in the supplementary materials). The later Yaghnobi dictionaries are generally void of etymological remarks,[11] so *Jagnobskie teksty* remain the main source of comparison of the lexica of two languages. Later, Livshits outlined some differences between the two languages and noted that Proto-Yaghnobi was substantially different from literary Sogdian. He thought it could continue the unattested Ustrushana Sogdian dialect (Livshits 1962a: 152). He also raised doubts on the postulated development of non-initial **-θr* into *(l)l* in Yaghnobi: both examples (*púlla* 'small, child' from **puθra-* and *ol* 'fire' from **āθr-*) can be questioned (Livshits 1970 note 29).[12]

A long chapter on Sogdian-Yaghnobi dialect relations was included by A. L. Khromov in his description of the Yaghnobi language (Khromov 1972: 119–135). He critically reviewed Bogoljubov's viewpoint and reached the conclusion that it is impossible to decide whether Yaghnobi *d* was articulated as /δ/ in the proto-language or not.[13] He follows Livshits in doubting the

[11] A few stray notes were included in the glossary of Khromov's book (Khromov 1972: 163–189). Novák (2010) adds references to the language of origin of words in the entries, e. g. (S) for Sogdian, (A) for Arabic, etc; many entries do not have such remarks.

[12] More recently, Novák added another example: *kat(t)ōlá*, *kattalá* '(1) stone shelter made with no wood; (2) ruin(s)', which he believes to come from **kata-āϑra-ka-* 'house-fire', Matcha Tajik *katol* (Novák 2013: 115). I would rather see here a metathesis of Persian *kalāt* 'fortification'; this word, a substandard borrowing from Arabic *qal'a(t)*, is very common in the south-east of the Iranian world.

[13] One can add an argument in favour of proto-Yaghnobi **δ*: the verb 'to pass', *gᵘdár-/dᵘgár-* is borrowed from Tajiki *guzaštan, guzar-*'idem'. In Persian, however, it is always spelled with *δ* (گذشتن), it is one of the few cases where so-called *dhāl-e Fārsī* remained as /z/ and not /d/, and apparently it was pronounced *guδar-* in some Mavarannahr

development *-*θr* > *l* and adds some other observations on Yaghnobi: the lack of voicing of clusters with initial *n-*; the lack of future tense, the connection between the Sogdian durative particle *'štn* and the ending *-išt* in Yaghnobi present and durative past. He followed Livshits in explaining Yaghnobi as a continuation of Ustrushana dialect. In his articles, Khromov discussed some lexical correspondences between two languages (Khromov 1966a, 1970, 2007).

I can hardly name any major edition of Sogdian texts, where some Yaghnobi data would not be included in the commentary. Émile Benveniste devoted a special paper to Sogdian *'štn* and Yaghnobi *-išt* (Benveniste 1966); the (late) Sogdian absolutive *-ky(')* and Yaghnobi gerundive in *-ki* were connected by Yutaka Yoshida (apud Sims-Williams & Hamilton 2015: 54–56); many Yaghnobi excurses can be found in Martin Schwarz's contributions on Sogdian (e. g. Schwartz 1970, 1982, 2021, 2024). Some 150 Yaghnobi parallels (albeit far from exhaustive and often in simplified transcription) are provided in the etymological section of entries in B. Gharib's *Sogdian Dictionary* (Gharib 1995) and are put together in the Index B (included in the supplementary materials).

Analysing the 3pl endings in *-āra* in Iranian, Xavier Tremblay suggested in a footnote that Proto-Yaghnobi had once undergone the "Rhythmic law", but later it was lost (as in the late Christian Sogdian manuscript C5). He based this observation on the Eastern Yaghnobi 3sg.prs ending *-či*, which, he believes, is parallel to Western *-t* and comes from the Sogdian light-stem ending *-ty* (Tremblay 2006: 276–277 n. 50). Tremblay's idea was met with some objections (Panaino 2008: 5 quoting the opinion of Elio Provasi), in particular one should note that Eastern *-či* functionally corresponds to the Western full form *-tišt* of the full paradigm, not *-t* of the dependent paradigm and, moreover, the "intermediate" forms *-čěš, -tčě* are attested in the speech of Kul side-valley (Andreev & Peshchereva 1957: 236–237; Bogoljubov 1956: 269–271).

Some 20 years ago international field research in Yaghnob resumed. A large team directed by Antonio Panaino of Bologna University worked in summers 2007–2012, and the present author joined the project in 2008. The results, promptly published, barely touch upon the language history, but are very important for understanding historical relations between Sogdiana and Yaghnob (Basello, Guizzo, & Ognibene 2008; Panaino, Gariboldi, & Ognibene 2013). In particular, the authors review critically the romantic concept of Yaghnobis being descendants of Sogdian refugees who refused to accept Islam. In the framework of the mission, a workshop on Yaghnobi was organized in Ravenna in September 2008. Nicholas Sims-Williams and the present author spoke on the topics of Sogdian-Yaghnobi relationship, and that talk ultimately realized in this paper.

Around the same time, Ľubomír Novák of Charles' University collected field linguistic data in Yaghnob. He ended up with the already named Yaghnobi-Czech dictionary, a series of articles and a large portion of his dissertation on the Problem of Archaism and Innovation in the Eastern Iranian Languages. In the dissertation, he discusses the relations of the two languages, and in an additional (equally unpublished) paper, he further elaborates upon the question. He notes that the 1pl ending is Eastern Yaghnobi *-īm(išt)* vs. Western Yaghnobi *-ōm(išt)*, and while the first one shares innovation with Sogdian *-'ym*, the latter is a continuation of the Old Iranian imperfect *-*āma* (Novák 2013: 27). He reconstructs moreover the history of stress in Yaghnobi and Sogdian. According to his interpretation, the first two stages were common to both languages, but Stage III was an exclusive Sogdian feature, while Stage IV a late Sogdian feature. His conclusion needs to be quoted in full:

Persian dialects, too. It might have been borrowed in Yaghnobi at the time when intervocalic /δ/ was still pronounced in this Tajik word and the Yaghnobi change of /δ/ into /d/ had not yet taken place. Vladimir Aronovich Livshits pointed out this fact in one of our conversations more than 20 years ago. Another relevant example in Yaghnobi is *(a) damón* 'then, by that time', if it comes from Arabic *dhamān* □ﻣ□'time'.

> "… the development of Yaghnōbī comes from a deeper history so the language cannot be considered as a descendant of Sogdian or any of its dialects. As I stated above, Sogdian itself is defined by its further development after the shift to *Stress III*. If Yaghnōbī and Sogdian are to be linked to a common ancestor, we need to think about a *Proto-Sogdic protolanguage that developed as one of the dialects of Central Asian Scythians. The development of *Proto-Sogdic has to be from an earlier date as Yaghnōbī also preserved some morphological features, that appear to be peripheral in Iranian (e.g. 3rd pers.pl. verbal ending *-ōr* (< **-ār* < **-r̥*, i.e. 3rd pers.pl.ind.perf. ending). Also Yaghnōbī nominal inflection shows that it developed independently of the Sogdian *Rhythmic Law* — if a similar feature occurred, Yaghnōbī light stems should be inflected in a two-case system, the same as those of Late Sogdian (see Table 1). During their growth the Sogdic languages/dialects were still in a linguistic contact, as may be illustrated by various common tendencies in their development. Yaghnōbī phonology appears to be rather archaic in some aspects when compared to Sogdian; some basic phenomena of development of both languages may be set even to the period of the turn of the Christian era or even earlier." (Novák 2014: 10–11)

Novák also provides a South Asian variety of Swadesh word-lists for Yaghnobi and Sogdian (Novák 2013: 161–214) and concludes that a part of Yaghnobi lexica is best explained as a common Pamiri heritage. More recently, Ela Filippone examined the terms of body-parts in Yaghnobi with many etymological remarks (Filippone 2013), and Adam Benkato made an interesting comparison of copulas in Sogdian and Yaghnobi (Benkato 2020; cf. Vydrin 2025).

The similarities of Sogdian and Yaghnobi within Iranian linguistic unity are striking and incorporate virtually all levels of the language: historical phonetics (preservation of non-initial voiceless stops; absence of depalatalization of [tʃ] > [ts]; partial preservation of *xt*, *ft* groups; quality differentiation of *ā*/*ă* (Sims-Williams 1981)); morphology (similar plural ending and nominal declension, similar independent and enclitic personal pronouns, preservation of imperfect and augment, *-ki* gerundive, multiple copulas), syntax (similar formation of potential forms) as well as a long list of lexical correspondences. The geographic proximity of Sogdiana and Yaghnob cannot be ignored as well. Consequently, the comparison of these two languages (with only stray notice of other Iranian and non-Iranian tongues), seems a justified approach. The author does not intend to discuss in this paper the position of Sogdian and Yaghnobi on a wider scope.[14]

3. The features separating Sogdian and Yaghnobi and their probable dates

In this chapter I discuss the features of historical phonology, morphology and phonotactics that separate Yaghnobi from Sogdian, and try to locate them in history. These main features include (a) the stress patterns and the "Rhythmic Law"; (b) the position of the augment; (c) the third plural ending; (d) the shifts **δr* > *ž* and **θr* > *š* in Sogdian; (e) the survival of *ā*+ *y*-umlaut diphthongs in Eastern Yaghnobi and (f) the first plural simple past ending *-om(išt)* in the Western Yaghnobi dialect. As a chronological marker, I will often analyse the data of the "Ancient Letters", the earliest sizeable, well understood and exactly dated (before 313) group of Sogdian texts (Sims-Williams 2023).

3a. Sogdian "Rhythmic Law" and Yaghnobi verb stress pattern

As has been shown, the absence of the "Rhythmic Law" has been considered the main distinction separating Yaghnobi from Sogdian, while in Sogdian it is a key feature influencing its mor-

[14] In particular, the recent revision of division into Eastern and Western Iranian (Korn 2016, 2019) is not relevant for the discussion below.

phology. It has been extensively studied, and Paul Tedesco, Ilya Gershevitch and Nicholas Sims-Williams made the main contributions to its understanding (Tedesco 1926; Gershevitch 1954: 72–81; Sims-Williams 1982, 1984). Let me quote a brief description from the most recent encyclopaedia entry:

> "For understanding of the Sogdian phonology and morphology, the so-called "rhythmic law" is of vital importance […]. According to this law, the retention or loss of proto-Sogdian final vowels depends on whether they bore a stress; in its turn, the position of the stress in words of more than one syllable is determined by the quantity of those syllables: all stems in Sogdian, whether nominal or verbal, containing a long vowel are heavy and bear a stress, while a stress falls on the endings of light stems that consist only of short vowel(s)." (Yoshida 2016)

The "Rhythmic Law" can be traced in the development of all grammar and heavily affects the verbal paradigm, for example; but it is most visible in the nominal declension, where the light-stem nouns follow the Old Indo-European patterns with modest innovations, while the heavy-stem ones have a simple two-case and two-number agglutinative paradigm. All the Yaghnobi nouns follow the second pattern.

In the latest Sogdian texts, however, the light-stem declension started to disintegrate, and in the Christian manuscript E5 (old label C5), the light-stem m.nom. with final *-i* was often taken as a base for new heavy-stem declension (Sims-Williams 1982: 69–70).[15] In some of the latest Turko-Sogdian inscriptions from Semirechie, the bare light-stem bases seem to be attested: *zmn* for *zmnw* 'time' (Livshits 2015: 272) and perhaps *δst* for *δsty* 'hand' in a newly discovered inscription (Kol'chenko & Lurje, in print).[16]

As discussed above, there are no visible traces of the "Rhythmic Law" in Yaghnobi. In principle, one can hypothesize that in Yaghnobi there had once been a sort of the "Rhythmic Law", and later the heavy stem pattern, a much simpler one and serving a wider word class, was generalized, and the light stem disappeared without traces.[17] However, the stress pattern suggests the opposite.

In Yaghnobi the stress (Khromov 1972: 16–17) can fall on the ultima or more commonly the penultima; the final *-a* is unstressed, even in most of Tajik loans.[18] The enclitic pronouns and the oblique marker *-i* are unstressed, too. The verbal pattern is more complicated. If the verb root has a long vowel (i. e., *ō, ē* or *ū*), or the short vowel of the root is followed by two consonants, the stress remains on the root. If the root has a short vowel followed by a single consonant, the endings with a long vowel (except 1sg.prs *-omišt*) bears stress. If the vowel of the ending is short, the stress remains on the root, even when it features a short vowel followed by a single consonant. Since the oblique marker *-i (-y, -ĕ*) is considered short, the basic pattern for both nouns and verbs is essentially the same.

It is also very similar to proto-Sogdian stress just before the application of the "Rhythmic Law": the stress fell on the root (*barytone)* when it contained a long vowel or a diphthong, and

[15] The similar fossilization of f.nom. in -' had even a wider usage (Sims-Williams 2017: 86); one example in Sogdian Brahmi fragment (Maue & Sims-Williams 2024: 45).

[16] Alternatively, *δst xp* is a compound word meaning 'by (one's) own hand'; cf. also (Henning apud Gershevitch 1954: 77 note) on *jmnw-/jwmn-* in Manichaean Sogdian. Both examples remain doubtful. The expression *βγ* 'o, sir' at the end of a phrase in late Manichaean has been explained as an enclitic to the last meaningful word (Sundermann 2016: 32).

[17] Even in the standard Sogdian, the light stem declension was a sort of recessive gene, while the heavy stem was a dominant one: many borrowings and new formation follow the heavy-stem pattern, even without long syllables or diphthongs (Sims-Williams 1984: 208).

[18] Interestingly, the penultimate stress is attested in Tajik vernaculars of the Yaghnob valley (Novák 2018)

if there weren't any long vowels, the stress fell on the ending (*oxytone*).[19] It seems thus that the difference of the Yaghnobi pattern from the Proto-Sogdian one lies in two facts: in the former the syllable with a short vowel closed by two consonants also had two morae and bore stress;[20] and if neither root nor the ending had two morae, the stress remained on the root.

In the Yaghnobi verbal present paradigm, the short vowel was dropped, so the endings of 2pl *-t(išt)/-s(išt)*[21] and 3sg *-t(išt)* lack the vowel part, while 1sg *-om(išt)*, 1pl *-im(išt)* and 3pl *-or(išt)* preserved the long vocalic part; the 2sg is simply *-∅(išt)*. In the imperfect 1sg *-im*, 1pl *-om/-im* and 3pl *-or* are explainable virtually in the same way as the present endings, and so does *-∅* of 3sg (as in the Sogdian imperfect of heavy stems). Less clear are 2Sg *-i* and 2pl *-ti/-si* (Khromov 1987: 681–683), the former could develop from OIr. **-ah* or rather Opt. **-aiš*, or even be influenced by Tajik *-ī*. The latter can be a product of analogy according to the following proportion (the result underlined):

	Present	Imperfect
2sg	*-∅(išt)*	*-i*
2pl	*-t(išt)/-s(išt)*	*-ti/si*

This model can work for the nominal declension as well, so all the endings safe genitive-dative in *-é* (from **-hya*) and locative *-ī* (from **-yā*) were lost, while the last two merged forming the oblique case ending (as in the heavy stems of Sogdian), however, in view of the declension of plurals in *-t* (see below, 4a), I would not exclude a parallel development.

The "Rhythmic Law" operates holistically in the standard Sogdian grammar, but in the "Ancient Letters" (before 313), it seems that some light stem words lack any endings. However, this feature can be also explained graphically: the final short vowels could remain unrecorded in the script (Sims-Williams 2021: 176). This situation is known from Semitic quasi-alphabets such as Imperial Aramaic (the predecessor of the Sogdian script) or Chorasmian recorded in Arabic-based script.

In essence, my explanation is similar to the one of Novák (2014),[22] with four stages of stress, but it is perhaps a more economical one.

3b. The position of the augment

The augment is an old Indo-European particle **H₁é-*, Indo-Iranian **(H)a-*, which preceded the verb root to indicate past tense. It was a facultative feature, and is absent in many languages, including all Iranian ones, save Avestan (rare), Old Persian, Sogdian, Chorasmian, Yaghnobi and Tumshuqese (Ogihara 2019).

The augment in Sogdian and Yaghnobi had different destinies. In Sogdian it appears in the imperfect and usually manifests itself as a lengthening of the final vowel of the preverb. In simple verbs with two initial consonants, the augment *-y-* was often placed between them, a feature based on analogy (Gershevitch 1954: 95). In Yaghnobi *a-* is present in front of the verb in the simple and durative past, be it a simple or a (historically) prefixed verb.

[19] This explanation in fact does not contradict the one proposed by Bogoljubov and his examples of the loss of short syllables in the root (Bogoljubov 1957b: 10–12; Khromov 1987: 991).

[20] Note that Gershevitch's explanation of roots with short vowel and final consonant clusters beginning in *r-* or *n-/m-* as an exception to the "Rhythmic law" has been abandoned in favour of newly formed *R* and *N* diphthongs (Sims-Williams 1984: 209–13).

[21] The final *-išt* is absent in the dependent present, which is a more archaic form.

[22] Who quoted my handout of 2008.

In Sogdian, the simple verb does not have an augment. However, due to the tendency to record prothetic vowels with *aleph*, this *aleph* can be mistaken for an augment (Benveniste, 1929: 45; Gershevitch 1954: 95). The augment only after the preverb (or pseudo-preverb) is well attested for the standard Sogdian. In the "Ancient Letters" the imperfect verbal forms seem to be absent,[23] but the imperfects *wytr* and *pr''γz('nt)* are attested in an early document from Niya and a mid-5th century one from Turfan (L.M.II.ii.09 R14, 17 and M96: 66, 2, 3 in Sims-Williams 2023: 74, 80). The first cannot help us with regards of augment: the present stem *wytr* < **wi-tr̥ya-* is also normally spelt with *y* in the first syllable, but the second one shows an internal augment in contrast with the present stem *prγ'z* (< **fra-gāz-*).[24]

Summing up, we cannot date exactly the separation of Sogdian and Yaghnobi with regard to the augment; the 5th century is the latest possible date. Before that, it seems that placement of augment was subject to dialectal variation: before the root in proto-Sogdian and before the word-form in proto-Yaghnobi. For an analogy, cf. the position of durative prefix **(ha)mē* with preverbs in Modern Persian *bar-mi-āyad* and Tajik *me-bar-oyad* 'rises'.

3c. The third plural ending

In Sogdian, the third plural ending of all the attested tenses is *-(')nt*, as in the majority of Iranian and Indo-European languages. In Yaghnobi, on the contrary, *-ant* is restricted to the subjunctive mood, while the relevant ending is *-or* in all the other temporal forms. In the present and past durative of the main clause this ending is now pronounced *-ošt*, which comes via *-oišt* (Andreev & Peshchereva 1957: 300) from *-orišt*, the form documented in early records. In some western villages *kunōrišt* has been attested up to 1940s (Bogoljubov 1956: 271).

The third plural ending with *R* is also attested elsewhere in Iranian: Avestan athematic optative *-ārə, -ārəš* and athematic indicative, injunctive and perfect preterit *-ārə*, Gothic *-ərəš*;[25] Khotanese active conjunctive *-āru*, optative *-īru*, middle present *-āre*, conjunctive *-āru/-ānde*, optative *-īru*, perfect *-re*;[26] Chorasmian indicative *-'r(y)*, imperfect and injunctive *'r̄*, conjunctive *-'r^{i}*, optative *-yr*[27] (Tremblay 2006). So Yaghnobi is not alone to display variation of the elements *NT* and *R* in 3pl.

As said above, Sogdian does not know the 3pl in *R*, unlike the related tongues. One naturally wonders if this fact can be explained by insufficient documentation. Yet, In the "Ancient Letters" alone, I have counted more than 20 forms of 3pl ending in *-'nt, -'ynt, -ynt* or *-'y'nt*. They include present, optative, intransitive and transitive preterit, optative intransitive preterit.[28] Imperfect *pr''γ'z'nt* (noticed above) appears in a later, 5th century document from Karakhoja, Turfan.[29] To sum up, I think that the third plural verbal endings went separate ways in Sogdian and Yaghnobi before the 4th century CE. Before that the common proto-language might have had the *R* endings in some temporal or modal forms and the *NT* endings in the other ones, as in some Middle Iranian sister languages.

[23] For the past tense, normally the preterit form (sometimes called perfect) is used, as in several Christian Sogdian manuscripts (Livshits & Khromov 1981: 488), which is perhaps a dialectal feature.

[24] See (Sims-Williams 2023: 81) on shortening of the root vowel.

[25] Vs. *-nt, -an* or *-at* (<**n̥t*) in other moods.

[26] In other moods it is *-īndi, -ānde, -āndu*

[27] Here *r* varies with *l* in the so-called "athematic" endings.

[28] Of course, in the latter cases it is a 3pl form of copula and the root *δ'r-*, respectively.

[29] The plural endings of subjunctive haven't been attested in Sogdian and are considered to be substituted by injunctive (Yoshida 2022: 356).

3d. The shift of *δr > ž and *θr > š in Sogdian

In many Sogdian words, Old Iranian *dr > δr developed into ž and *θr into š, respectively.[30] These consonant cluster simplifications were not universal, as in not a few cases we see a metathesis into rδ or rθ (Gershevitch 1954: 44, 47, 67). This sound-change is alien to Yaghnobi where word-initially these clusters turn into *dir-* and *tir-/sar-*, and word-internally **dr* becomes *rd*.[31]

The date of this change in Sogdian is difficult to establish, particularly because the historical *δr* and *θr* were often archaizingly spelled as <δr> in the national Sogdian script. This fact indicates, first of all, that the written tradition emerged when they were still pronounced as *δr* and *θr*. I have not met any clear case of spellings with *z* or *š* in the "Ancient Letters" or other early Sogdian texts. Quite suspicious though is the spelling *'tδrtyk* 'third' (from **θrityaka-*) in AL 3.26[32] (Sims-Williams 2023: 41). In the later texts, the sequence *'tδr* was used to record *š-* (or perhaps *čš, ṣ*?): apart from the same *'tδrty(k), 'čštyk, 'štyk* 'third' in Buddhist Sogdian we see the spelling *'tδrzm'k* 'anger', variant of *δrzm'k/y*, perhaps */(š)žmē/*, a peculiar development of Old Iranian **aižma-* Avestan *aēšma-* (Gershevitch 1946: 138) and *'tδrmnw* 'demon, Māra, Ahriman', Manichaean and Christian *šmnw /šamnu/*, which comes (perhaps as a loan) from the unattested Avestan **ašå-mainiiuš* 'the worse spirit' (Sims-Williams 2021b: 59). This spelling would suggest that 'third' was already pronounced as */ə(č)ṣtīk/* in the early 4th century.

In the Upper Indus inscriptions, we meet the names *myšpr[n]* 'glory of Mithra' (very indistinct) and *tyδrβntk* 'glory of Tīštriia' (Lurje 2010a: 258, 399). The latter could be a pseudohistoric spelling based on *myδr /Miš/* 'Mithra', or a simplification of **tyšδr-* for **Tīš-θrya- > Tī-ʸšš*; in both scenarios, the change of *θr > š already happened. The dating of this group of texts is a difficulty, the hypothesis of Étienne de la Vaissière that they correspond to the time of Hunnish (*xwn*) dominance in Sogdiana in the first half of the 5th century (Vaissière 2005: 82–83) is very attractive but cannot be verified.

In any case, the change *δr > ž* has already happened in Sogdian in the eve of the 7th century: the name *δrymtβntk* is rendered as 史射勿槃陀 Shi Shewupantuo, Early Middle Chinese *ʑiah-mut-ban-da*,[33] on a tomb-inscription of 609 CE (Lurje 2010a: 181).

Another approximation of the change *dr > ž* can be gleaned from loan-words. The goddess *δrymt*, or *jymt*, the patron of the 11th month (and the theophoric base of the mentioned *Shewupantuo*), was borrowed from Greek *Δημήτηρ*, as becomes clear from the forms *Δηματρ-*, *Δρηματ-* in Bactrian (Sims-Williams & de Blois 2005: 165). The borrowing might have happened in the Hellenistic period of the history of Sogdiana. Another ultimately Greek borrowing, from *δραχμή*, appears in Sogdian as *δrxm-*, without either traces of change into *ž* or metathesis into */ərδaxm-/*. This word does not appear in early Sogdian (Sims-Williams 2016b: 106–107), and in all probability it was borrowed from Middle Persian *drahm*, after the silver drachms of Peroz and Kawad (since late 5th century), their imitations and Bukhar-khuda coins (since mid-7th century) became the most common silver currency in Sogdiana. So, the change

[30] Somewhat similar development can be observed in Shughni group and partially in Chorasmian (Edelman 1986: 172–175).

[31] On a dubious development of the non-initial **θr* into *-ll-* see p. 43 above. It is altogether strange that no clear cases of development of the common cluster **-θr-* are documented in Yaghnobi in the non-initial position. The only relatively safe example is *dĭrót, darós* 'sickle' from Old Iranian **dāθra-* (Rastorgueva & Edelman 2003: 440; Wendtland 2009: 178), where metathesis outstepped the long *ā*; although S *δr''š-* (if here, cautiously Benveniste 1979: 93) makes the development more complicated (cf. also under *šyš-* in the Appendix). There are plenty of words featuring *l* (excluding clear loans) in Yaghnobi but none of them has a satisfactory etymology.

[32] Alongside *'δrtyk* in AL 5.32.

[33] Here and below the reconstructed Early Middle Chinese pronunciation follows Pulleyblank 1991 s. vv.

dr > *δr* > *ž* operated in Sogdian between the 2nd century BCE and perhaps the 4th century CE (while in the late 5th or 6th century it already did not function), and the separation of proto-Yaghnobi should be dated prior to it.

3e. The survival of *ā*+*y*-umlaut diphthongs in Eastern Yaghnobi

When in late 1990s Mikhail Nikolaevich Bogoljubov taught Sogdian at the Oriental Department of the Saint-Petersburg State University, he rigidly refrained from pronouncing anything[34]; for him Sogdian was only "reading for the eyes". When asked why he used to say that, he would point out that even in Yaghnobi it is unclear whether the diphthongs became monophthongs or not. The example of Yaghnobi *raụ̂t, roụ̂t* from **rautah-*, Sogdian *rwt*, can now be dismissed in favour of *r'w'th* (Schwartz 1970: 301 with reference to Khromov; see appendix), and Yaghnobi *xaụ̂k* 'marmot' from Old Iranian **hūka*, Middle Persian *hūk*/*xūg* 'pig' is problematic from both phonetic (otherwise unattested *guṇa* form ***hauka-*) and semantic viewpoints. Furthermore, there are numerous counterexamples with Yaghnobi *ū* from **au* (Khromov 1987: 654).

The situation with secondary *Y* diphthongs is different: if there was long *ā* in the root, and a suffix featuring *-y-* after a consonant, group of consonants or short syllable, we find *ɛ* in Eastern Yaghnobi, but *ay* in Western Yaghnobi. Bogoljubov provided a long list of correspondences (Bogoljubov 1956: 134–137): *wɛš*/*wayš* < **vāstrya-* 'grass', *méšin*/*máyšin* 'serum' < **māṣ(t)nya-*, *nipéšin*/*nipáyšin* 'grandson' < **napās(t)nya*, *mɛn*/*mayn* 'village' < **nmānya-*. This development is very common in the verbal stems in *-aya-*: *nipɛ́d-*/*nipáyd-* 'to put'< **ni-pādaya-*, *pinéš-*/*pináyš-* 'to lose' < **apa-nāsaya-*, *γuréš-*/*γuráyš* 'to awaken' < **vi-grāsaya-*, *sɛn-*/*sayn-* 'bring up' <*sānaya-*, *piréž-*/*piráyž-* 'run away' < *apa-rāzaya-*, *bidéž-*/*bidáyž-* 'to wrap' < **upa-dārzaya-*.

This fact cannot reflect a secondary re-diphthongization: the oblique case of *-a* words has *-ay* (e. g. *póday*) in Western Yaghnobi and *-e* (*póde*) in Eastern Yaghnobi (Khromov 1972: 19) and moreover, most of the words discussed have transparent etymologies,[35] so Western Yaghnobi *ay* is a very old feature.

Before examining this feature in Sogdian, let us look at some names and toponyms in foreign transmission. The toponymy of Sogdian or proto-Yaghnobi origin in the upper reaches of Zarafshan is controversial: we see *mayn* in the names of boroughs *Sari Mayn* of Iskodar[36] and *Mayn* in Kurut,[37] but *Tomen*, *Tomin* or *Tūmen* to the east of Ayni, *Veškant*, *Vešab* and *Vēšist*[38] which are most likely based on Sogdian *wyš*, Yaghnobi *wɛš*/*wayš* 'grass'. The Arabic rendering of Sogdian names and place-names is very ambiguous in this matter too, cf. e. g. al-Bayrūnī or al-Bīrūnī for Iranian *Bērūnī* (Blois 2024: 174–176).

Chinese renderings are also ambiguous: the early 7th century Sogdian king *šyšpyr*, from **sraišta-paira-* 'best faith' or rather **θrāšaya-paira-* 'disseminating faith' (Cheung 2007: 290–91, 397; Lurje 2010a: 380) is Shasebi (沙瑟毕, EMCh. *ʂaɨ*/*ʂɛ:-ʂit-*pjit*) and, surprisingly, Arabic *Šašbīr*. The Sogdian word for 'lady, queen' *xwt'ynh*, which might have a diphthong of a similar origin (**xwa-tāw-nī-*, **xwa-tāw-nyā-*) is rendered (as a part of personal names) as *Hudianpantuo* (呼典畔陀; EMCh. *xɔ-tɛn'-ban^{h}-da*) in Chinese and *Hkhutteṃ-Wām* in Tocharian A (Lurje 2010a: 447 with references). The Arabic rendering *Ḥudhayna* (for *Khudayna,* خُدَينة > حُذَينة), as said

[34] Any reader of the edition of the Mount Mugh economic documents (Bogoljubov & Smirnova 1963) can notice the same approach.

[35] Yaghnobi *nɛs*/*nays* 'nose' is, however, a complication, see Bogoljubov 1956: 136–137.

[36] The place where a cist with well-preserved mummies was found, see https://fergana.media/articles/139212/ (access 10.01.2026).

[37] My record of 07.07.2024.

[38] Recorded as *'y-wšy-šth* and *wyšβ-* in the Mount Mugh documents.

above, is also ambiguous. The word *rēžkēn* 'desiring', from **rāzaya-* in Sogdian in Brāhmī script shows *ē* (Maue & Sims-Williams 2024: 91). Cf. on the contrary *nai d_1[ē]* for *nyδk nyδ'y*, Old Iranian **naida-ka-* 'no-one'.

In trying to examine whether such diphthongs occurred in the bulk of Sogdian texts, one is hindered by the consonantal scripts where the vowels are quite roughly rendered by prototypically consonant letters', *w, y*. Thus, in National script, we see *wyš, np'yδ*, etc., with the vowel rendered as *y* or *'y*, as is virtually the case of any long *ē* or *ī.* At any event, there is no trace of the etymological *-y* in the following syllable.[39]

The common rendering of long vowels by combinations with aleph in non-initial position, namely *'y* for *ī, ē* and *'w* for *ō, ū*, is quite unusual in itself for other scripts derived from West Semitic abjads. One can wonder if at the time of the script formation the diphthongs **ai* and **au* had not yet contracted into *ē* and *ō*. However, even in the earliest texts we see *sp'δny* < **spāda-nī*, *nwkmyδnc* < **nawaka-maiθana-*, *nxšpyk* < **nixšap-iyaka-*, *šwtt* < **šuta-* but *kš'yk* < **kiš-iya-ka-* (Sims-Williams 2023: 1–10). In the "Ancient Letters", we see for example *'nw'yz* 'to bring together, emass' < **ham-vāzaya-* and *'pr'yz* 'to flee, run away' < *apa-rāzaya-*; on the other hand, the name *ct'ysh* corresponds to *Catisa* of Niya documents, *myn-* 'to remain' comes from**(d)mānaya-, np'ys-, np'yns-, npys-* 'to write' from **ni-paisa-* or **ni-pinsa-*, etc. So, even in the early texts, the spelling *'y* was not an exclusive feature of secondary diphthongs.

Summing up, the monophthongization of *y*-umlaut diphthongs in Sogdian happened at least before the 7th century based on Chinese and Tocharian *Nebenüberlieferungen*, and such diphthongs cannot be distinguished from other *I* vowels in Sogdian texts since an early date.[40] The spellings like *'y* and *'w* could initially have been used for diphthongs in the first centuries BCE, when the writing tradition of the National Sogdian script emerged.

3f. The first plural simple past ending *-om(išt)* in the Western Yaghnobi dialect

The opposition of Eastern Yaghnobi 1pl.pst *-īm(išt)* and Western Yaghnobi *-ōm(išt*) was first noticed by Khromov (Khromov 1972: 103–104), and its importance is pointed out by Novák (Novák 2013: 27, 148, 218). According to Novák, the latter ending comes from the Old Iranian imperfect ending **-āma*. The Eastern Yaghnobi form *-īm(išt)*, identical in present and simple past (imperfect), comes from the optative *-aima* or the *-aya*-stem present **-ayamah*. The 1sg ending is *-om(išt)* in both dialects (from **-āma*).

In Sogdian, the 1sg ending is *ām* in the present, *-u* in imperfect, *-ān* in the optative; 1pl is *-ēm* in all tenses and moods, sometimes also *-ēman* (Gershevitch 1954: 111–112). The copula is 1sg *'ym* and 1pl *'ym* or *'ymn*. Even if they were pronounced differently (*im, ēm(an)*), the two are quite similar and both were used to form intransitive preterit.

Note also the synchronic inversion of the copula endings in Yaghnobi: present 1sg *īm*, 1pl *om*,[41] simple past 1sg *óyim* 1pl *iyóm* in all dialects (Khromov 1972: 684). The perfect 1sg is *wófta īm*, 1pl *woft-om*; pluperfect is *wóft(a) oyim, wóft(a) iyom*. Given all this complexity of similarly sounding forms, I wonder if 1pl simple past *-óm(išt)* in Western Yaghnobi could not be an analogical development.

[39] Nicholas Sims-Williams reminded me of a counterexample: *mδy''n* 'middle' (parallel to *myδ'n*) from **madyāna-*.

[40] At a first glance, the Christian Sogdian orthography (with its usage of diacritic signs) cannot help much here too. Cf., however, the spelling *npy.d-/np'.d-* for *ni-pādaya-* (Sims-Williams 2016a: 119). Perhaps, more investigation on this subject is needed.

[41] These endings render Old Iranian 1sg.prs **ahmi* and conj **ahāma* respectively.

Let me summarize the discussion in the following table:

No.	Feature	Date of separation
(a)	The Yaghnobi stress patterns and the "Rhythmic Law"	Before the 4th century (AL)
(b)	The position of the augment	Before the 5th century (Karakhoja fragment)
(c)	The third plural ending *-or/-ant*	Before the 4th century (AL)
(d)	The shift of **δr* > *ž* and **θr* > *š* in Sogdian (vs. the preservation of the clusters in Yaghnobi)	After the 2nd century BCE (historical spelling in Sogdian, name of Demeter), before the 7th century (*Žēmatvande* in Chinese rendering, *δraxm*), perhaps before the 4th century (AL, *'tδrmyk*)
(e)	The survival of *ā*+*i*-umlaut diphthongs in Eastern Yaghnobi vs. their monophthongization in Sogdian	Perhaps after the 3rd century BCE (spellings *'y, 'w* in the National script), before 7th century (*Nebenüberlieferung*)
(f)	The first plural simple past ending *-om(išt)* in the Western Yaghnobi dialect.	Unclear, perhaps an analogical development.

All in all, I would date the separation of proto-Yaghnobi from an early predecessor of the literary Sogdian to the three centuries before or three centuries after the Common Era. We will return to the dialectal isoglosses of Sogdian in chapter 5.

4. Shared innovations

Having separated once upon a time around the turn of the common era, Sogdian and proto-Yaghnobi did not go in two completely different directions, and many features that appeared later are shared by the two languages, as it was often dictated by some more general trends.

4a. As has been shown recently, Sogdian "Ancient Letters" still possessed the historic plural endings, and the new plural in *-tā* was still used as a collective (Sims-Williams 2021); in the later texts it was generalized, and only few traces of the older plural survived. In Yaghnobi, we find only the plural suffix *-t*. This innovation (or a similar one based on the **-θva* collective) is shared by a number of Iranian languages: Ossetic, New Persian, some of Pamiri tongues.

4b. The converbial suffix *-ky, -ky'*, although known in standard Sogdian, became (under Turkic influence) very common in the latest texts (Sims-Williams & Hamilton 2015: 54–56), while converbs in *-ki* are also well known in Yaghnobi; sometimes the expressions are very similar: *tnk xrʾ ʾγʾt kt xwny pwsty cwn ywzy βrtʾskn* ***wʾβky*** "Tang Qara came saying 'the edict bearer was bringing the document'" (op. cit., G12-13) vs. *γóre nüt atér 'man túra ben, man túra ben'* ***wóvke*** "she came in the cave saying "look what I will (do) with you"" (Andreev & Peshchereva 1957: 353);[42] *ZKZY ZKw ʾnyw wʾtδʾr nyʾst ʾPZY šw* ***ptrʾyšky*** *xwrty* "who takes other beings and tears them up and eats them" (P2, 457-8 apud Sims-Williams & Hamilton 2015: 55); Brāhmī ***pi zki*** *hv₁[e]r* 'having crushed it, cause (the patient) to eat (it)' (Maue & Sims-Williams 2024: 67–68 5B1) vs. *tu* ***xáški*** *žaf* "drink (it) in one gulp (lit., pulling)" (Khromov 1972: 47).[43] The Tajik participle in *-agī* is related here too (Sundermann 2016: 32–33).

[42] "Она вошла в пещеру говоря «смотри что я с тобой (сделаю)»".

[43] "Ты выпей залпом (букв., втягивая)".

4c. Another feature is the enclitic pronoun used as a possession marker placed after the possessed object. In standard Sogdian, the enclitic pronouns behaved still as Wackernagel clitics and were attached to the first stressed word (usually, a conjunction or an adverb). The forms like Manichaean *frn'š* 'his glory' *tmp'r'š* 'his body' (L73+L77; So 18196, ed. Yoshida 2001: 111–114), *m'tm* 'my mother' (Durkin-Meisterernst 2006 line 1164), inscriptional *ptryš* 'his father' (?, Livshits 2015: 290 line 14); Christian *cn xwž'mntyš* 'from its demands' (E28/16 R20 ed. Sims-Williams 2017: 106) and others (Yoshida 2009b; Sims-Williams 2015: 158) share the pattern with Turkic, Arabic, Yaghnobi and New Persian forms. Similarly to Sogdian, in Middle Persian we see this construction rarely and only in late texts (Jügel 2016: 44), while in New Persian (as in Yaghnobi) it becomes the norm. In this case, the influence might have come from two sides: Arabic and Turkic.

4d. Another feature described for Yaghnobi are serial verbs. While the first verb in the phrase has the full form, the second and consequent ones follow the so-called dependent, shortened paradigm. Cf. e.g. the following examples *aǰáxim Nűmitkon amón atír* "I stood up and left for Nomitkon"; *odámt bisyór ǰām avór aǰáx šikóri ašáw* "Many people gathered, stood up and went hunting"; *…auqót kunóšt xᵒarór* "they make food and eat" (Khromov 1972: 41–42); *ax parwóz kúntišt, osmúne vek vort* "(the bird) flies, brings him in the sky" (Lurje 2015b: 156). These serial verbs are most common in the description of movement: *aǰáx ašáw* 'stand up (and) go'[44], *asán awes* 'ascend (and) descent', etc.

I noticed similar double verb expressions (without different paradigm of course) in Christian and inscriptional Sogdian: *'t srpywn ''xšt x(r)[t qw] wyšnt s'r* "And Serapion arose (and) went to them" *r](ty) sxwsďrt γryw ʾwxšt yty nysty* "[... And] he lifted himself up (and) went down and embarked [on a ship]" (E26/28 V5-6; E26/30 V17 ed. Sims-Williams 2015: 126–133); the expression *βyrt pryp* 'obtained (and) took away' twice in the documents from Hisorak (2,2; 3,2; Lurje 2012: 446–447, 444, 451) is of particular interest here because these texts come from the Zarafshan valley not so far from Yaghnob. Of course, all these serial verb expressions (and similarly Tajiki *girifta burd* "brought having taken") have been influenced by Turkic (Doerfer 1967: 52–56; Soper 1992; Perry 2006).

4e. Simplification of syllable-final clusters (and even apocope of word-final consonants) represent quite widespread features of Yaghnobi and are occasionally detectable in some late Sogdian texts. So, 'full', usually *pwrn* <**purna-*, is rendered as *pwn* in some Christian manuscripts (Sims-Williams 2016a: 159), compare Yaghnobi *pŭnn*. The reflex of **xwaipaθya-*, normally Sogdian *xypδ*, appears as *xyp* in several late Manichaean texts (Sims-Williams & Durkin-Meisterenst 2012: 225), inscriptions (Livshits 2015: 272)[45] and *he-p* in *Brāhmī* transmission (Maue & Sims-Williams 2024: 92), compare Yaghnobi *xĕp*, *xap*, *wxap*. For a parallel apocope, see Sogdian *s'* postposition 'towards' from older *s'r* vs. Yaghnobi *sa*.

Finally, I would like to outline very briefly a quite general syntactic difference between Sogdian and Yaghnobi: in the former, there is a plethora of conjunctions, introductory particles and other function words. W. B. Henning complained that on a Sogdian fragment from Turfan, the searching eye encounters only conjunctions, particles, auxiliary verbs, pronouns—a host of meaningless elements, unlike Parthian or Middle Persian (Henning 1958: 91–92). In Yaghnobi, the quantity of modal words and conjunctions is at minimum (Bogoljubov 1956: 389).[46]

[44] Note the soubriquet of Yaghnobis used by neighbouring Tajiks: *jahak*, lit. 'jumper' (Guizzo 2013: 153).

[45] Also perhaps *xp* in another inscription (Kol'chenko & Lurje, in print).

[46] This feature was outlined recently by A. P. Vydrin in the paper 'Yaghnobi dependent forms' at the conference "Indo-Iranian Languages of the Pamir, Hindu Kush and beyond" held at the Institute of Linguistic Studies in Petersburg on 6 December 2025.

I would not ascribe this difference to psycho-linguistic features of speech of eloquent traders and laconic mountaineers (so, Pamiri Shughni is quite rich in this kind of function words),[47] but rather to more profound influence of Turkic with its tendency to use converb constructions instead of complex sentences.

5. Some cultural traits and extra-linguistic data

Only a handful of words or expressions bear witness to cultural continuity between early medieval Sogdians and present-day Yaghnobis. The word *nimóč* 'namāz, prayer', Sogdian *nm'c(yw)* indicates in very broad terms some continuation of worship practice. A number of old words in the sphere of marriage (cf. *k'm — komomišt*, *r't — rot*, *xwž — xúǰgǫ́n*, *pšt — pša^{h}-namónča* in the Appendix)[48] might indicate continuity of matrimonial customs. Sogdian *xutarne* and Yaghnobi *xutánna* 'water-mill', literary 'auto-mill' indicate a very important innovation in the agriculture that was implemented around the 6th century[49] and it is also reflected in the Pamir language Wakhi as *xədorg* etc. (Steblin-Kamenskij 1999: 410). The elements of a water-mill (*prn — pann*, *kpc'k — kapčák*, *ptš'nkh — pašáng*(?)) also preserve much of Sogdian lexicon.

In several cases, the similarities are also shared by Tajiks (and beyond), and the direct relation cannot be established. Thus, the Sogdian expression *z'ry r't* 'cried pitifully' (many times in Vessantara Jātaka, the well Sogdianized version of the Buddhist popular story) is parallel to Yaghnobi *zor-zor roy* (Andreev & Peshchereva 1957: 369), but the parallel expression *zor megiryam, girya kunam zor-zor* also occurs in Tajiki.[50] The prediction of the sex of a new-born from the shape of its mother's belly in the Sogdian text So 18700 (Reck 2005) closely corresponds to the popular medicine in Yaghnob (Delaini 2013: 131), but the same practice goes far beyond, including classical Persian texts. The Yaghnobi folk-tales, as is widely known, are closely parallel to the Tajik ones.[51]

The borrowings must also be taken into consideration. Sogdians, the traders between China, Steppe, India, Iran and Byzantium, borrowed many words from their close and distant neighbours, not just religious terms (Novák 2013: 213). When referring to 'sea, ocean', they used *sumdar-* from Indian *samudra*, for 'inn' they used *tīm* from Chinese *dian* (店 EMCh *tɛm^{h}*), the terms for silver coins were borrowed (ultimately) from Greek: *stēr*, *δraxm-* (mentioned in Section 3d), while the designation of bronze or copper coins, *pn-* came more likely from Indian *paṇa-* (Thierry 1999: 325) rather than from Chinese *fen*. For 'palace', *šaykan* from Middle Persian *šāhīgān* was used, for 'dagger' — *xa^{n}γar*, a word of ultimately Hunnic origin (Dybo 2007: 83–85), etc. Even the term 'caravan', important as it was for traders, was initially called *s'rt* from Indian *sārtha*, and later *''rxyš* from Turkic *arqïš*.

[47] As noticed by O. I. Belyaev in the discussion at the same conference.

[48] The expression quoted for *soda* 'always' in Andreev & Peshchereva 1957: 323 is a benediction for the bride, but Mirzozoda (2008: 182) provides contexts from other life situations.

[49] I am not aware of mill-stones discovered in antiquity period sites in Sogdiana. They become numerous since the early medieval period. The name *xwtyy'wrn* however appears in the Upper Indus inscriptions (Lurje 2010a: 449–450), and Nisaean Parhian personal name *ḥwtrn* is older (Schmitt 2016: 112–113), maybe the innovation came from Western Asia.

[50] Checked in the National Tajik corpus, https://tajik-corpus.org/. Unfortunately I overlooked that fact in Lurje 2010b: 780 n. 3, and it led to a hasty conclusion.

[51] I take the opportunity to note that the "Tale" about Ali and Imam Mahdi, which I recorded in 2008 and published in 2015 (Lurje 2015b), is also known by Panjakent townsfolk.

Yaghnobi has, of course, a very large number of Tajik loans (including those of ultimately Arabic origin), the share of Turkic (Uzbek and likely Kyrgyz)[52] and Russian loanwords is also significant; some words, as noted above, are shared with the Pamir languages. In contrast, loanwords dating back to the pre-Islamic period are very few. They arguably include *šiwína, šawén* 'home-made cotton garment' vs. Sogdian *wšwynk* 'sort of fabric', etc., if indeed from *βύσσος*, cf. Turkic *böz*, Russian *bjaz*' (Sims-Williams & Hamilton 2015: 67–68) and *kr'nx, kr'x, qrx* 'paralyzed', Yaghnobi *karáx, kĭráx* 'numb', if from Indian *grāha-* (Henning 1946: 715; Schwartz 1982: 86; Sims-Williams 2015: 153). Noteworthy both *šavina*[53] and *karaxš* are attested in Tajiki (Rakhimi & Uspenskaja 1954: 182), Persian *karax(t)*. The correspondence of S. *ryz'kh* 'rice', related to Indian *vrīhi-* and *(γurda)režónna* 'rice, grain' (secret word) is dubious, see Appendix under *ryž-*. If *rouča, rű́ča* 'fast' indeed comes from Sogdian *rwc* 'day' (as Tajiki *rūza*, Middle Persian *rōzag* <*lwck'*> 'fasting'), it would indicate that Western Middle Iranian borrowed (in Achaemenid period?) calendar with the day-names ending in *-rwc* was used by proto-Yaghnobis before Islam (see Appendix).[54]

The safest common pre-Islamic loan is Sogdian *'rk* 'duty, corvée labour, work', Yaghnobi *ark* 'work', which had been ultimately borrowed, like Middle Persian *harg* 'duty, tax', Bactrian *ναργο*, Yidgha *horγ* etc., from Akkadian *ilku* 'land tax, duty' via Aramaic **halkā* (Henning 1958: 41; Schwartz 2004; Sims-Williams 2007: 272).[55] In this respect, the much discussed Mt. Mugh document Б-7 (Livshits 2015: 156–157) is particularly important. The letter from an unnamed ruler of Martshkat to Wishtam, the ruler of Kishtud, informs that the people of Parghar came and did not stop doing their work (*'rkh*) and [following] orders, but then many of them ran away, so that no more than seven people remained; the author asked to catch the fugitives. Antonio Panaino's attractive idea that the people from Parghar mentioned in the letter were in fact the inhabitants of Yaghnob (Panaino 2013b: 47–51) is not without difficulties.[56] It is clear though that *'rk* 'duty, work, corvée labour' was an important part in the mountaineers' economy of the early 8th century. The Proto-Yaghnobi's contacts with the rulers of Sogdiana minimally constituted of corvée service. Their distant land was hardly a part of the flourishing culture and long-distance trade of the pre-Islamic Central Asia.

The viewpoint that Yaghnobis are refugees from mainland Sogdiana who were driven away by Arab and Mongol invasions is an old (Gauthiot 1914) and still enduring viewpoint (Gunya 2002: 18–21). It is especially favoured by Tajik historiography (as discussed in Panaino 2013a) and appears very commonly in the romanticized journalist and travel-blog accounts. At the same time, this viewpoint seems to corroborate modern Western theories of the mountain-

[52] Not all of them are shared with Tajik. Yaghnobi *kátta* 'big' from Uzbek *qatta* is not common in Tajik.

[53] I incidentally heard this Tajik word in Panjakent.

[54] Alternatively, as Sims-Williams suggested to the author, the inherent reflex of Old Iranian **raučah-* might have existed in proto-Sogdian.

[55] Iranian etymologies of Sogdian *'rk* and cognates have also been proposed (Rastorgueva & Edelman 2000: 198; Cheung 2007: 131–132). The similar specialized meanings of Akkadian *ilku* (Kienast 1980) and Middle Persian *harg* are, however, noteworthy.

[56] The main rationale for Panaino's hypothesis is that the relevant people came down (*'wxšt'nt*), and Parghar is below Martshkat (modern Hisorak, opposite to Madrushkat in Matcha). Yaghnob, however, is higher than Hisorak, and in order to gain Hisorak from Yaghnob, one has to traverse a pass across the Zarafshan Range. As I am going to show elsewhere, the pre-modern *Falghar* did not correspond to the modern Aynī district of Tajikistan as a whole, but only to its Zarafshan part. The valley of Fan-darya and its tributaries (now the part of the Aynī district) was then the *Fan* or *Iskandar* district, and Yaghnob was sometimes considered its part, and sometimes a separate entity. It is very likely that *srwcyk xwβw* of the Mount Mugh documents corresponds to the Fan princedom, named after *srw-* for modern Sarvoda, its main centre (Lurje forthc.).

ous "Regions of Refuge" (Aguirre Beltran 1979) and of upland movement in Southeast Asia and beyond to escape state violence (Scott 2009).

As we have seen above, the proto-Yaghnobi language was closely related, but different from Sogdian, at least since the Hellenistic period. Yaghnobis did not share the advances of the highly developed Sogdian civilizations, at least *en masse*. Extralinguistic evidence tends to support this view. The petroglyphs indicate that the valley was inhabited (at least, as a pastoral ground) since late antiquity (Passarelli 2008: 108), and there is limited evidence for the presence of pre-Islamis strongholds in the valley (Panaino 2013b). At the same time, the mitochondrial DNA data, ambiguous as they are for historical conclusions, suggest that Yaghnobis are genetically quite distant from the neighbouring Tajik population (Cilli et al. 2013).

More can be said about the better investigated 6th–9th century archaeological sites of the Upper Zarafshan valley, which share geographical proximity and some ecological similarity to Yaghnob: Mount Mugh, Kum, Gardani Hisor in Parghar, Hisorak in Matcha and some others.[57] Although elements of their élite culture are quite similar to the lowland Sogdiana (Panjakent and further west) and lowland Ustrushana (especially the Shahristan oasis) with respect to monumental architecture, fortifications and decorations, everyday life was very much different. Money economy in the highlands was at its minimum; the buildings were almost all of one floor; the structure of the household was different, without any reception rooms, and much more stone was in use as opposed to mudbricks. The pottery of lowland Sogdiana is predominantly wheel-made and that of highlands was predominantly hand-formed, which is a very important ethnic indicator, burial customs were different, too (Lurje 2022: 260). Moreover, the more recent historical records show that the highlanders were more often moving to the lowlands for a better life (Arandarenko 1889: 160–163; Mukhtarov 1978: II, 117–119),[58] than vice versa, although this occasionally happened at the moments of danger, in search of temporary refuge.[59]

6. Isoglosses in Sogdian and conclusions

In the course of the previous discussion, we have treated Sogdian as a uniform language. It is well known that the main differences between the texts in the National, Manichaean and Christian scripts are not of a dialectal nature but reflect different stages of the same language and different distance between the written and colloquial registers. It is natural to suppose that the vernacular of Samarkand, the largest city of Sogdiana, its nominal capital, the seat of Manichaean archegus and Christian metropolitan, served as a base for its literary language (Yoshida 2016), so to say, the Sogdian *koine*.

However, spoken reality in the confederation of Sogdian city-states was more complex. The 10th century geographers noted both similarities and differences in the speech of various parts of Sogdiana (al-Muqaddasī edited by de Goeje 1906: 335–336). Through the analysis of rich toponymical data, we can distinguish some features of the Bukharan dialect of Sogdian (Lurje 2015a). Moreover, a different, more archaic script was used in the pre-Islamic Bukhara oasis (Djakonov, Kaufman, & Livshits 1954).

Even the written language was not as uniform as it appears at first glance. The recent study shows a similar use of articles in the Buddhist Vessantara Jātaka and the Christian

[57] Apart from this, there are bronze-age petroglyphs, Sarmatian-type burials and antiquity period pottery documented in the area (Isakov, Jakubov & Karimova 2020, vol. II).

[58] The common 19th – 20th century cases of labour migrations (Dynin 1914: 73) and mercenary forces (Babadzhanov 2010: 118–119) might be mentioned too.

[59] Cf. the well-studied case of highland Ustrushana (Stark 2008).

"Deeds of Pethion" (C2/E27, 1), which indicates their dialectal affinity (Wendtland 2011: 350). The translator or scribe of the Sūtra of Causes and Effects tended to write the final *-i* with *aleph* and not with *yodh*, contrary to the common practice, and this habit is probably indicative of his dialectal background (Sims-Williams 2025a); the dialect variation of usage of past verbal forms in Christian Sogdian has been recently scrutinized by Yutaka Yoshida (2022b: 605–610).

I suppose that some particular Sogdian texts include the features that are close to Yaghnobi and perhaps mirror the fact that their authors originated from the eastern extremity of Sogdiana or Ustrushana. So, Vessantara Jātaka of Dunhuang and Nirvāṇasūtra of Turfan use the durative particle *'štn* (equal to Standard Sogdian *skwn*, which is also present there), similar to Yaghnobi *-išt* (Benveniste 1966). In two or three texts we meet *km-* 'neck' and *rx'* 'mouth', which are distinct from standard Sogdian *γrδ'k* and *kwc'kh* and similar to Yaghnobi *káma* and *rax* (see Appendix), respectively. The lack of the prothetic vowel before consonant clusters in Sogdian texts in Brāhmi transmission and small proportion of forms with prothesis in Christian Sogdian could reflect the pronunciation patterns that were more similar to Yaghnobi, where epenthesis is used instead.

In some cases, forms from a dialect close to proto-Yaghnobi may have entered the standard Sogdian. As M. N. Bogoljubov noted, in Yaghnobi historical **a* before *rt* regularly realizes as *ort* (Bogoljubov 1957a). The standard Sogdian perfect passive participles *(')βy'rt-*[60] from *βyr-* 'to obtain' (Yaghnobi *ver, vyórta*) and *δβ'rt* from *δβr-* 'to give'[61] share the same isogloss.

I tried to show above that Yaghnobi and standard written Sogdian separated from each other in their main features perhaps around the beginning of the common era, not earlier than the 3rd century BCE and not later than the 3rd century CE. Having separated they, however, underwent some parallel innovations, partially inspired by Turkic adstrate. Although in Yaghnobi there are very few traces of participation in the high culture of the Sogdians, and their ancestors did not move from the lowlands in the course of the documented history, yet some dialectal forms similar to Yaghnobi (and perhaps used in the wider area of eastern, highland Sogdiana or Ustrushana) appear from time to time in Sogdian texts.

The lexical parallels between Sogdian and Yaghnoby, which are enumerated in the Appendix and Indexes, further show the profound relation of the two. More than one hundred years after Gauthiot posed the question if Sogdian and Proto-Yaghnobi were two closely related languages or dialects of the single language, it is still impossible to give a definitive answer: even the definitions of language and dialect vary significantly. It is impossible to prove, but I strongly suppose that native of early medieval Yaghnob, if happened to learn Sogdian literacy, would not consider it learning a foreign language and would regard it a script of his own, with some peculiarities of the literary language needed to be learnt. We all have spent time and efforts learning our literary mother tongues at primary school.

Addendum

After this paper had almost been finished, I got access—through the kindness of its author—to the very recent and very important paper "Yaghnobi and the Sogdian 'Rhythmic Law'" (Sims-Williams 2025b), which is an elaboration of the talk I attended in Ravenna in 2008.

[60] The consistent spelling with *aleph* in the middle makes the explanation of *plene* writing of short *a* unlikely. Note *βyrt* in Manichaean texts (Sims-Williams & Durkin-Meisterenst 2012: 61), *byrt* in Christian (Sims-Williams 2016a: 58–59).

[61] Also in 3Sg pres *δβ'rt(y)* commonly in Buddhist texts (beside *δβrty*), but it is spelt without *aleph* with other endings *δβr'nt, δβr'y* (MacKenzie 1976: 99). Similarly to the case of *βyrt/byrt*, only the forms *δβr-/δβrt* and *θbr-/θbrt* are attested in Manichaean and Christian texts. Note Yaghnobi *tĭfar-, tĭrafta, sᵃrafta* due to metathesis. Novák gives among others the participle *tᵘfőrta* but does indicate his source (Novák 2010: 164).

The author shows convincingly that "Rhythmic law" never did affect Yaghnobi, and the development of the nominal inflexion attested in C5 into the Yaghnobi one is impossible; the "Rhythmic Law" was already active in the "Ancient Letters" (as seen from prothetic vowels). Paul Tedesco's argumentation about the "Rhythmic Law" was more nuanced than what I assumed. Xavier Tremblay's explanation of *-či* ending is rejected on the same grounds as above in the present paper and the analogous formation of 2pl *-ti/-si* is similar to the one proposed above. Sims-Williams thinks that the augment *a-* was stressed in Proto-Yaghnobi.[62]

The main divergence between the opinion expressed above and that of Sims-Williams is that the latter thinks that stress never fell on the final vowel in proto-Yaghnobi, while I take seriously the modern verbal stress pattern, which prompts me to conclude that the long syllable of the ending obtained stress if there was no long syllable in the root. The conclusion of Sims-Williams is also quite similar: "Yaghnobi must have diverged from Sogdian at an extremely early date".

After that I received—again through the kindness of the author—the Japanese "Lectures in Sogdian Grammar" of Yutaka Yoshida (2022), where, on p. 461–468 there is a short hand-list of Sogdian-Yaghnobi lexical correspondences. The Yaghnobi side of the list is based on *Jagnobskie tektsy*, while Sogdian counterparts include all the known vocabulary. A large number of the new parallels in the Appendix below and Yoshida's list are virtually identical. However, I overlooked the similarity of several words, and I gladly acknowledge Yoshida's primacy;[63] very interesting is the comparison of S. *'βt'rnky* and Y. *bildínga* 'Ursa Maior', although the phonetic correspondence is highly unusual.

7. Appendix. Sogdian-Yaghnobi lexical connections after *Jagnobskie teksty*

The only systematic collection of Yaghnobi and Sogdian lexical correspondences in the dictionary to *Jagnpbskie teksty* published in 1957. Since then, many correspondences have been noticed by the scholars who published Sogdian texts and researchers of Yaghnobi; however, they were rather stray notes on individual lexemes scattered in a large number of publications; many of them were utilized in the *Sogdian Dictionary* of Badrizamon Gharib. A systematic analysis of body part terms, an important group of lexics, was published by Ela Filippone (Filippone 2013). Many until now remain unnoticed and some moreover are based on unpublished Yaghnobi materials collected by B. A. Alborov, M. N. Bogoljubov (and some by the present author).[64]

This appendix is designed to incorporate all such correspondences but I readily admit something might have been overlooked. The appendix includes a large group of forms which appear similarly in both Sogdian, Tajik and Yaghnobi. While some of them are Sogdian heritage in the present-day Transoxiana, the others are similar development from Old Iranian to Sogdian, Yaghnobi and Tajiki, so one cannot make a clear conclusion what has been inherited in Yaghnobi and what has been borrowed.

[62] As it appears in Vedic as well (MacDonnell 1910: 315).

[63] S. *pyx* 'finger-joint' – Y. *pax* 'finger'; S. *prn'y-* – Y. *parr-/párta-* 'to fly', S. *mž'* – Y. *míǰa* 'eyebrow', S. *γwš'yck* – Y. *γúšak* 'pitcher', S. *βrpyx* – Y. *farbéx* 'fat'. I leave out his comparisons of S. *nym'k* – Y. *néma* 'side' and S. *šy-* – Y. *ši-* 'to freeze' because they need profound philological analysis on the Sogdian side; Y. *čaγzlóy* has T. *loy* 'dirt, clay' in the second part and has little to do with S. *r'γ*, Y. *roγ*.

[64] Some obvious cases are lacking in the *Jagnobskie teksty*, perhaps as a slip of pen. Some others are mentioned vaguely in the Russian-Yaghnobi glossary (371–391) as these lexemes weren't documented in the texts of the volume themselves.

Apart from independent words, some suffixes were also included. As a rule, the Old Iranian etymons and cognates in related languages are not included, but I intend to discuss the rare or abnormal phonetic or morphological changes, and quite often try to explain the semantic side. I admit that some of the pairs still remain controversial.

The Sogdian form(s) are followed by Yaghnobi one(s) and the Yaghnobi-Sogdian index can be consulted in the online "Supplementary materials". For Sogdian, the common words can be consulted in the dictionaries (Gharib 1995; Sims-Williams & Durkin-Meisterenst 2012; Sims-Williams 2016a)[65] and use to limit myself to one or two most characteristic forms, the rare words are provided with full references. For the Yaghnobi words included in Novák's dictionary, I use his system of transcription, the others are rendered as they appear in primary recordings.[66]

Abbreviations

A. — Alborov 1934; AL — "Ancient Letters"; B. — Bogoljubov 1956; Br. — Sogdian in Brāhmī script; CS — Christian Sogdian; F. — Filippone 2013; Gh. — Gharib 1995; Ja. T. — Andreev & Peshchereva 1957; Kh. — Khromov 1972; L. — Lurje, records of 2008; M. — Mirzozoda 2008; MS — Manichaean Sogdian;[67] N. — Novák 2010; S. — Sogdian; SW — Sims-Williams 2016a; S.-B. — Salemann's materials in Benveniste 1955; S.-D. — Sims-Williams & Durkin-Meisterenst 2012, T. — Tajik; Y. — Yaghnobi, YY — Yoshida 2022a.

1. *ʾ k̓wc-*, *ʾʾkwγt* 'to hang' – *okuč-* 'to hang', *okučun-*, causative of the former (A., 62),[68] cf. Y. *kúča* 'earring'.
2. *'βt'r'nkty* 'Ursa Maior(?)' (Sims-Williams apud Sundermann 1990: 31 n. 71), Avestan *haptō.iringa-*, Persian *Haft awrang* — *bildínga* 'star, Ursa Maior' (in the vernacular of Kul side-valley)[69] with YY. The phonetic correspondence is quite unusual, but it would be a rush to dismiss it.
3. *'δry, šy* 'three'[70] — *s̱aráy*/*ṯiráy* 'three' (missed by accident in Ja. T.)
4. *ʾm̓nk mγδy* 'apple-fruit' (Sims-Williams & Hamilton 2015: 71, F19) — Y *man, mannak* 'apple' (B., 8, 161, 767), an archaic word, usually *seb* from T., apples (as any other fruit-trees) do not grow in the Yaghnobi populated part of the valley (Andreev 1970: 44–45).
5. *'nδ'w-*/*'nδwt-* 'to annoit, to apply' — *andőwa* 'plaster, stucco' (B., 155), more likely from T. *andova* (articulated *andowá* in surrounding dialects), as initial **ham-* realizes as *in-* in Y. (cf. *ínkir*, S. *'nk'yr* 'oven' and *insáxs* below).
6. *'ns'xs-*/*'ns'γt-* 'to be arranged, fixed, successful' or historical abstract *'ns'k* 'suitable, fitting'-*ínsáxs* (so accented in N., 72; M., 94) 'to get used to, to be in good relations'.[71]
7. *'nzβr-*,'to withdraw' (AL), *nyzβ'r-* 'to take away' — *žavár-*, *živár-*/*žaγáta* 'to take out' (Yoshida apud Bi, Sims-Williams, and Yan 2017: 317; Sims-Williams 2023: 38; YY).

[65] Note that B. Gharib did not distinguish systematically *γ* from *x*, which I do, so e. g. *'xšy'k* should be looked upon *'γšy'k* there.

[66] Alborov used the obsolete Latinization of Tajiki, and I changed it into modern Iranological one (e. g., not *ş* but *š*, not *ç* but *ǰ*, not *ƣ* but *γ*).

[67] CS and MS are indicated only when necessary.

[68] My elderly informant from Nometkon gave the correct translation of *okučón* after a minute consideration. It was deep in his passive vocabulary.

[69] B., 678, adds T. *palding* which is otherwise unknown to me.

[70] Only *(')δry* in Manichaean, secular and Buddhist texts, while *šy, š'* in Christian (unlike *(')štyk* 'third' in the majority of texts). Perhaps, in a numeral of the first decade the old articulation */əθrē/* survived in many dialects. Cf. in this case Y. *avt*/*aft* 'seven', while in all other cases **vd* > *ft*. Nicholas Sims-Williams kindly reminded me of *qθry* beside *qšy* 'now' even in C texts.

[71] 'Унс гир, муросо кун'; 'zvyknout si na něco'.

8. *'pδr'y-* 'to protect', from **apa-θrā-*(?) — *matratá* 'bolt (on the door of a mosque)' (Showeta; Lurje 2015b: 161–162), from *hama-θrăta-* 'co-protected' (Cheung 2007: 394). Doubtful.
9. *'pγ'wnt-, pγwnt* 'to reveal, to uncover' — *buγŭnt(a)* 'pocket' (also YY)?[72]
10. *'psm'k* 'chowry, yak-tail, fly-whisk, ornament, Skt *cāmara*', from **apa-samaka-* 'wiper away' (MacKenzie 1976: 60) — *pisám-* 'sweep out' (Kuhn's materials), *pⁱsáma* 'rubbish, garbage' (B., 804, etymology without S. word; Kh., 177).
11. *'stm* 'al last' (Sims-Williams apud Sundermann 1996: 105; note 21, 2001:II, 731), Avestan *ustəma* — *sⁱtám* 'back, backside', *sⁱtamakő* 'backwards'.[73]
12. *-'str*, superlative and comparative suffix of heavy stems (Gershevitch 1954: 192) — *-istar* old suffix in the words *ŕítistar* 'before', *sarḥád(d)istar* 'higher' (B., 185, 212), cf. *mástar(tar)* 'close, nearby' (N., 104).[74]
13. *'wpt-/'wpst-* 'to fall' — *pit-* 'to fall' (recorded in Showeta; Lurje 2015b: 162), from Old Iranian **pat-*, here without preverb, cf. Khotanese *pat-* (Cheung 2007: 300–301).
14. *'xšy* 'to gnaw', *xšāya[mante]* 'itching' (Br., Maue & Sims-Williams 2024: 42) — *xⁱšóy-* (MacKenzie 1970: 44; Gh., 28; YY)
15. *'ycyn* '?' (Sims-Williams & Hamilton 2015: 53; E5) — *íčin* 'thread', Rushani *incav-* 'to sew', see more under *syncyn*.
16. *'yšty* (inf., Sims-Williams apud Sundermann 1990: 32 n. 82; Gh., 88) 'to boil' — *eš, éšta* 'to boil'; the prefixed verb *'βy'yš-* is given in Ja. T., 251.
17. *'zm(w), zmy* 'brushwood' (Livshits 2015: 219, 225) *–ízim* 'brushwood' (MacKenzie 1976: 77).
18. *bn* 'long' (CS, SW., 53) — *van(n)* 'long', <**baržna* (Sims-Williams 1985: 55; cf. B., 863; YY).
19. *βnt* 'bond, fetter, bundle, parcel' — Y *vânt* 'hay-sheaf' (MacKenzie 1976: 93; Gh., 105; YY).
20. *βr'n* 'breath' — *firón* 'smell' (Benveniste 1959: 80; B., 715; Gh., 106).
21. *βr'yc, βr(')ytch* 'sheepskin', lit. 'shorn' (?; Mt. Mugh documents; Livshits 2015: 71, 135)[75] — *viríč-/viríčta, virín-/viríta* 'to cut, sheave' (Livshits, loc. cit.)
22. **βrγ* 'dam' (widely documented in toponymy Lurje 2016), from **bag-ra-* 'dividing one' — *varγ* 'upper edge of a field' (where water is collected for irrigation).
23. *βrγ'm'k, prγ'm* 'yearling, one year child' (Sims-Williams apud Sundermann 1981: 196; line 503; Reck 2005, 2V8) — *farγúmč* 'heifer, fresh cow' (noticed also by YY).
24. *βrpyx* 'fat (?, noun)' (hapax in Mt. Mugh documents; Bogoljubov & Smirnova 1963: 32, n. 20; Begmatov 2019: 153; Sims-Williams 1983: 44: a Middle Persian borrowing) — *farbíʰ, farbẹ́x* 'fat (adjective)', T. *farbeh* (with YY).
25. *βstγwnp, fstxwmp* 'to propel, repel' — *čumf-/čúmfta* 'to push' (Gershevitch apud Sims-Williams 1983: 49; Novák 2013: 192; YY), if from **pcxwmp-*.
26. *βwr-* 'brown, grey' (Mt. Mugh, Bogoljubov & Smirnova 1963: 39), commonly in PNs (Lurje 2010a: 150–151) — *vur* 'colour-term'[76] (also YY), T. *būr* is farther phonetically.
27. *βykp'r* 'outside' — *vēkpōr(a)* 'outside, outer' (B., 863, also *vēkpár*).

---

[72] Even though 'pocket' (Tajik *kisa*) was historically not a pouch in the dress but a sack attached to the belt, the semantics remain unclear. Maybe a 'secret language' formation from the verb *bᵘγúnǰ/bᵘγúšta* 'to steal' (**buγúnǰna* > **buγútna* > *buγúnta*?)

[73] Acc. to B., 185, 839 from **pastáma-*; alternatively, S. *'stm* is a variant of *stmb* 'stern, severe, coarse' (S-D., 179).

[74] Novák refers to Khromov and Mirzozoda, but I could not find this word in those dictionaries. The first part of the last word could be a reflex of first person deictic **ima-* (cf. *máskě* < **ima-arda-sara-*; *ín(n)ūr* 'today'?), or interjection *ma, maʰ* 'here (it is), *voilá*', which is borrowed from Turkic via T. (note that the root of *sarḥáddistar* is Arabic).

[75] The same explanation is advocated by Begmatov (2019: 236); Bogoljubov & Smirnova (1963: 40) refrain from translating it; cf. Middle Persian *'kblyt* 'lamb', lit. 'once-sheared' (Huyse 1999: II, 111–112).

[76] 'Серый, бурый' in Ja. T. 349; 'фиолетовый' in B., *vūrγúrda* 'голубоглазый' and *vūrmiǰá* 'с выцветшими бровями' (Kh., 184), 'хокистарранг, бӯр' (M., 45); 'hnědý, šedohnědý' (N., 182).

28. *βykʾ, fyq* ‘shoulder’ — *fūk* ‘shoulder’ (B., 134, 713; Benveniste 1959: 80; YY).
29. *bz-* , Pl. *bzyšt* (CS) ‘goat’ (Müller 1912: 22–24; SW., 59) — *vuz* ‘goat’ (also YY).
30. *cʾf cʾf* ‘all around’, *prcʾβt* ‘surrounding, enclosure’ from **kap-*/**cap-* ‘to hold’ — *čáfa* ‘pasture for horses’ (Kh., 164), suggested by A. L. Khromov and M. Schwartz in Sims-Williams 2017: 239; the semantics need further elucidation.
31. *cšm-* ‘eye’ — *tišm* ‘eyelid’ in Kuhn’s records (B., 860). The first *t* is a result of dissimilation, as Parachi *teč, teč*‘, *čeč* (Jefimov 2009: 241) < **čašman-*.
32. *cγwt-* ‘owl’ — Y. *čŭγű́z* ‘woodpecker (Ja. T., 241); owl (M., 239)’; in Ja. T., it is compared to S. *cγz-*, which is in fact ‘frog, toad’; YY provides *cγwt-* ‘owl’ which is difficult in view of the final consonant; the explanation as a T. loan *čuγz*/*čuγd* ‘owl’ (in itself from East Iranian, see Henning 1939: 96), is preferable.
33. *cym-* ‘eye’ — *čimm-* in *čimmakí kun-* ‘to wink’, T. *čašmak zadan* (Kh., 165; Rastorgueva & Edelman 2003: 240), also *čom, cēm* in Pamiri tongues, Khotanese *tsei̯mä-*.
34. M *δrβ-* ‘to tremble’ (S.-D., 73) — *darf* ‘am abhang gelegenes land’, ‘mountain ridge’ (S.-B., 154; N., 34), cf. the village *Xisóki darv*. Semantic shift remains unclear.[77]
35. *drs* (CS) ‘skin’ (SW, 71), Brahmi *δarsne* ‘made of skin’ (Maue & Sims-Williams 2024: 89) — *dorsẹ́na wíta* ‘woolen rope’ (Ja. T., 248; Benveniste 1959: 80).
36. *δxš-* ‘to hurt’[78] — Y *daxš-* ‘to hurt’ (Gershevitch 1959: 181; YY).
37. *frʾwyšcy* (+ spelling variants) ‘forgetfullness’ — *firómič* ‘forgetfulness, forgetting’ (MacKenzie 1976: 101; YY); partial contamination to T *faromūš* according to (Schwartz 2024: 424).
38. *γδ-* ‘log, wooden beam’ (Lurje 2008: 223–224), Avestan *gaδā* ‘club, mace’, Khotanese *gei̯hā* ‘wood’, Ossetic *qæd*/*γædæ* ‘wood, log’ — *γáda* ‘a log, part of bridge’ (Kh., 168), also T. *γada* in Matcha (Khromov 1962: 194), Panjakent.[79]
39. *γmy* ‘humble’ — *γam* ‘sorrow’; T. also *γam* ‘sorrow’, so more likely a loan (Persian — Tajik word itself is an East Iranian borrowing).
40. *γnʾw-* (CS) ‘to slumber’ (SW., 84) — *γinő̯u̯-* ‘to nap’, cf. Persian, T. *γanaw-*/*γunudan*, an East Iranian loan (Schwartz 2024: 431).
41. *γδβk* (MS), *γdbyc* (CS) ‘bitch’ — *kutγáfča(k)* ‘puppy’ with *kut* ‘dog’ in the beginning.
42. *γrβ* ‘many, much’ — *γaftar* ‘much, many, more’ (S.-B., 152; B., 723; M., 51), from comparative **γarf-tar* with dissimilatory loss.
43. *γrw-* ‘cane, bamboo’ (Yoshida 1996: 169), Avestan *grauua,* Parthian *grww* — *γaráu̯* ‘bamboo’ (Kh., 168).[80]
44. *γwβ-* ‘to praise, to glorify’, *γwβcyk* “praised” (Ragoza 1980: 34; fr. 40, 11) — *γúči* ‘singing bird’[81], from **γúvči*; for semantics cf. Persian *parastū* ‘swallow’, from the verb *parastīdan* ‘to praise’; the suffix might be affected by *síča* ‘sparrow’.
45. *(-)γwδn* ‘covering’ (from **gaudana-*, in *nγwδn* ‘dress’, *crpγwδn* ‘ointment’, *βʾrγwδn* ‘perfume?’, *ryṭγw(w)δδ* ‘face cover, veil’), Brāhmī script *g₁o d₁āṃ* (Maue & Sims-Williams 2024: 83) — *γūdán, γűdʲn-* ‘mountain sagebrush, which is used as mat’ (Ja. T., 259).
46. *γwrdtq* (CS) ‘kidneys’ *–γúrkak, gurkák* (Sims-Williams 1985: 178; F., 644); the second form influenced by T. *gurda*.

[77] Moreover, as Nicholas Sims-Williams suggested, the first consonant in S was probably *θ*, not *δ*, as evident from CS *wyθrbnync*, Manichaean Bactrian *pydhrβ*. Perhaps this connection should be disapproved.

[78] All contexts are broken; S.-D. 76, s. v. *dxšny* and SW, 73 s. v. *dxščyk* consider the meaning unclear; the relation to Y. word is considered as an option by Sims-Williams (2017: 231).

[79] In the document from Muhammad Bashshara mausoleum near Panjakent of 1213/1798-99 (Jegani & Chekhovich 1984 Doc. 92, 4094): *čūbi γadai taxtai pul* ‘logs of wood of the bridge’.

[80] Bamboos do not grow in the cold arid zones such as Yaghnob. Perhaps, a kind of tall cane of riverbed is meant.

[81] ‘Lerche’ (S-B., 153); ‘ласточка (swallow)’ (B., 722); ‘skřivánek (lark)’ (N., 60)

47. *γwš'yck* ‘jar, pitcher’ (lit., ‘the one with ear > handle’?) — *γušák* ‘small earthenware pitcher’ (MacKenzie 1970: 56; YY)
48. *γẓt'(kyn)* ‘debauched, licentious’ — *γažd kun-* ‘to stain’ (Kh., 168).[82]
49. *k'β, kβt* ‘to split’ — *kaf-/kafta* ‘to crack’, T. *kafidan* is closer in meaning and vowel quantity.
50. *k'm-* ‘to wish’ — *kom-* in *inč kāmāmišt, kāmām* ‘ein weib nehmen’ (S.-B., 161; B., 746).
51. *k't* ‘dumb, deaf’ — *kṓta* ‘stupid, crazy’ (M., 108; N., 87)[83], cf. English *dumb* and *dumb bunny*, German *dumm*.
52. *kʾzʾkh* ‘hut, little house’ (Henning 1945: 158, n.) — *kóza, qozá* ‘hut; dwelling on summer-camp’, T. *koza*.
53. *kβn-* ‘few, a little’ — *kávⁱn* ‘few’ (S.-B., 160; Benveniste 1979: 184; B., 86, 758; Ja. T., 379).
54. *kmʾ, qm'y* ‘neck’ (fem.), the word attested in one Manichaean (M112, 21)[84] and one Christian (E28/10 R12) context.[85] — *káma* ‘neck’ (cf. Yoshida & Lurje apud S.-D., 97; F., 643; YY).
55. *kpck, kpc'kk* ‘unit of volume’, perhaps 5 liters or less (Bogoljubov & Smirnova 1963: 30–31) — *kapčák* ‘a utensil in water-mill, a trowel used to pour grain from chest to millstones’ (B., 750; Kh., 171; Khromov 1966a: 130);[86] the word *kafčok* is used in this sense in Zarafshan T. as well (Steblin-Kamenskij 2008); *beli kafčok* means ‘scoop spade’ to distinguish from *bel* ‘bayonet shovel’.
56. *kšph* ‘turtle’ — *kī́spa, kḗspa* ‘large basket’; dubious: the variation *šp/sp* needs explanation, and land turtles in the area are quite small.
57. *kwβ-* ‘peak of mountain’ (hapax *kwβty* in SSeele, 162, Obl.pl.); as the text is a translation from Parthian, the borrowing from *kōf* can be envisaged (Sundermann 1997: 157),[87] or rather *kwf* ‘wave’ (Sims-Williams 2017: 68) — *kúva* ‘hunchback’ (B., 758), cf. Khotanese *kuvaa-* ‘heap’(Bailey 1979: 63).[88]
58. *kwn* ‘anus’ (Sims-Williams 1976: 66) — *kun* ‘anus’,[89] T. *kun*.
59. *qwšy* (CS) ‘side, flank’ — *kuš* ‘furrow’ (B., 755), semantics need further elucidation.
60. *kyrm-* ‘snake, worm’ — Y *kírⁱm* ‘worm, dragon’ (Ja. T., 273), but also ‘snake’ (A., B., 732, 397, 752, 768; Gauthiot 1914: 93; N., 86; YY);[90] in T., *kirm* is ‘worm’ only.
61. *mʾnʾkH* ‘part of head’ (Sundermann 2002: 140, 22; S.-D., 108) — Y *manáʰ* (and variants) ‘chin’, Persian (glossed as Mawarannahr word), T. *manah* (F., 640–641).
62. *m'n'nt'k* ‘resembling’ — *-mōnta* postposition ‘alike’ (B., 158; YY).[91]

[82] ‘Грязнить, пачкать’.

[83] ‘Гӯл, гаранг’; ‘hloupý, ospalý’.

[84] The reproached Manichaean hierarch is accused of washing *km* in the water (*ʾpyy ʾpryw kmʾ snʾyyskwn*). Werner Sundermann initially compared it to Persian *kamā* ‘a sleeve; old cloth sewd together, whereon bread is laid to be baked in the oven’ (Sundermann 1980: 294) and later preferred explaining it as knife, a loan from Turkic *qamγa* (Sundermann 2007: 407, 2016: 31). This explanation is questionable as *qamγa* is expected to be written **x'mγ'* or alike in S. The similarity of this word to Y. *káma* and plausibility of ‘neck’ in this context was noticed independently by Y. Yoshida and myself.

[85] Sims-Williams 2017: 78, 86, the story of Mar Awgin. Here *qm'[y] (γ)d[*’ ‘the wound on his neck’ is confirmed by Syriac original.

[86] No. 20 on the scheme in Steblin-Kamenskij 1999: 465.

[87] Sogdian toponyms containing *Kūf* are few and can be questioned (Lurje 2004: 122–123).

[88] YY takes the obsolete meaning of *kwβ* ‘foam’ and compares to *xáf(a)* ‘scale, foam’ (which has good Pamiri parallels).

[89] Cf. also *tobkúna* (Ja. T., 377), *xankṹna* ‘upper part of a canal’ (L), from *tóba, xan* ‘irrigation canal’.

[90] Cf. in the tale recorded in Tagi Chanor in 2008: *Kírim wóvta man zahr kunómišt* ‘the Snake said “I make venom”’ (Lurje 2010: 14).

[91] Other sources give *mṹnta, mī́nta, mḗ(n)ta, má(i)nta* from *myn(t)* (Ja. T., 285; B. 313, 315, 348; N. 106; Gh., 206: *mayanta*)

63. *m'y, m'yš, m'yšt* (CS) 'indeed' (?) 'particle indicating that a quotation follows' which has been compared to Tocharian B *mai* 'doch, etwa, peut-être' (SW., 104; Sims-Williams 2019: 71) — *may, máy-ĭš* 'cautiously' particle 'take care, this way' (Ja. T., 285; M., 120; N., 105);[92] semantics need elaboration.[93]
64. *mrγ'* 'wood, meadow' — *marγ* 'wet place covered with grass; marshy lawn', T. *šebar*, cf. T. *marγzor* (Andreev 1970: 99; B., 769).[94]
65. *mrz-* "to rub" — *γumórz, γǫ́morz* 'an open place for overnight stay of cattle on summer camp' (the place where cows stay cramped[95])?
66. *mwry'* (MS) 'channel, pipe' S.-D., 117); *mwry'-* (CS) 'tunnel' (Sims-Williams 2015: 145; SW., 109) — *mur* 'kitchen, oven',[96] cf. Matcha T. *mūr, mūro* 'hollow, tight gorge in the mountains' (Khromov 1962: 177).
67. *myn-* 'to think, consider, imagine' *myn'y* 'thought, decision, apperance' — *mayna* 'memory' (B., 775, with etymology).
68. *myš* (CS) 'sheep, ewe' (hapax E27/12r12, SW., 112) — *mẹš* 'sheep' (also YY).
69. *mz'* 'eyelash' — *míǰ(ǰ)a* 'eyelash', T. *mijja*.
70. *n'wn* 'river (?)' [97] — *noų́* 'valley' (MacKenzie 1976: 114).
71. *nγ'm* 'time, moment'[98] — *γum, γǫm* 'time, *fois*'.
72. *nγn-* 'bread' — *naγama, niγama* 'time to eat (usually lunch)' (M., 129),[99] cf. T. *noništa*. The ultimate *m* remains difficult to explain.
73. *nnsβwn'kh* 'nostril' (Sundermann 2002: 138, 11) — *naysvúna* 'nostril' (F., 641–642; YY).
74. *nnšk-, nšq-* 'palm (of the head)' (Yoshida 1984: 145–146; Sims-Williams 1995: 60)[100] — Y *nišk* 'palm' (Novák 2013: 186; Sims-Williams 2015: 150; F., 633; YY).
75. *nyʾs-/nyʾt-* 'to take' — *nos-, nóta* 'to take' (MacKenzie 1970: 62; Gh. 249).
76. *nynʾkkh, nynʾk-* 'pupil (of an eye)' (Sundermann 2002: 139; Yoshida 2019: 96) — S *nínak* 'pupil', T. *ninak* (cf. YY).
77. *[nyp]š* 'honeycomb'(?) (Schwartz 1982: 188; Schwartz apud Sims-Williams 2020: 41) — *nípša* 'comb', *nĭ́póš-/nĭ́pošta* 'to comb'; as one can see, the Sogdian word (hapax in So 15490 = Ps. 19, 11) is fragmentary and this restoration is a learned guess based on Y. word.
78. *p'rδwnp* 'crupper' — *pordúm* 'undertail strap' (B., 800), Persian *pārdum*, T. *pordum* (Gershevitch 1946: 147).
79. *pc'w* 'to agitate, uproar, quarrel, disturb' (Henning 1936: 68; Sims-Williams, Schwartz, and Pittard 2014: 51) — *čow-* 'to press'. The two verbs are different in meaning, so the Sogdian one perhaps with the preverb **apa-* or **upa-*, 'to release from press'?

---

[92] 'Осторожно'; 'боэҳтиёт, эҳтиётқорона, чунон, ҳамчунон'; 'opatrný, opatrně'

[93] From Old Iranian **(H)may-* 'to fix(ate)' > 'concentrate oneself' (Imp.)? Or rather onomapoetic?

[94] Other sources provide the meaning 'the grass best consumed by cattle', apparently herbs from such lawns.

[95] Cf. in Tienshan realia: "One night there was a blizzard in the mountains. Sharp, stinging snow clung on the horses' manes, it made their tails heavy and hurt their eyes. The herd was restless. The horses huddled together, trembling. The old mares snorted nervously and drove the foils into the middle of the herd…" Chingiz Aitmatov, "Farewell, Gyulsary", tr. F. Glagoleva, 1969, II, 11.

[96] On Yaghnobi kitchens which are hardly more than some place around an oven under a central chimney, see Andreev 1970: 103–105.

[97] Alternatively, *n'wznth*, plural of *n'wh* 'boat' + *-wzn* 'course' (cf. *'nxrwzn*), as (Sims-Williams 1978: 260)

[98] *γʾm* "pace" (Henning 1940:Y16) is purely a restoration.

[99] 'Вақти хурок (бештар дар мавриди хуроки нисфирузӣ)'; N., 111, gives with optional second *n*: *naγám(n)a* / *nʲγám(n)a*. Cf. *niγamma* '(time of) haying' (secret, Kh., 175; Khromov 1966a: 131). Ilya Yakubovich suggests here the reflex of *nγ'm* 'time, moment'.

[100] Likely also *n(n)šk'* and *p(š)'nnšky* 'back of the hand' in the "Book of Limbs" (differently Sundermann 2002: 144).

80. *pδ-* ‘foot, step, *pied(mont)*’ — *pad* ‘sole, footstep’ (more distant *pδʾk*, compared in Ja. T., cf. the following; also YY).
81. *pδʾk* ‘on foot’ — *dog* ‘footstep’ (S.-B., 153); for the loss of *p* cf. Y. *dara* ‘stomach’, S. *kδʾrʾk*, Y. *tuxóy-* ‘to kill’, S. *ptxwʾy-*. The suffix vowel quantity, however, seems to differ.
82. *pδn’y* ‘wide’, Old Iranian **paθana-* — *patan/m* ‘wide’ (?, A.),[101] *pátaxan* ‘large irrigational canal’ (L., in Nomitkon),[102] from **patan-xan*, cf. *bidóvanka* ‘belt’ from **bidon-vanka*.
83. *pr* preposition ‘to, on, at, in, for, upon’, ‘in exchange for’ (AL and Turko-Sogdian) — *par* ‘for, at’ (Khromov 1970; Sims-Williams 2023: 58; YY).
84. *prn’y* ‘to fly’ — *parr-/párta* ‘to fly’ (as noted by YY), T. *par(r)idan*.
85. *p(r)št* ‘lips (pl.)’ — *pšaʰnamónča, pašmű́nča, pišahmónča* ‘old woman who opens (T. *namondan*) the face of the bride during wedding’, T. *rūynamō*; maybe also in *puškúnča* ‘oven cover’ (M., 163). For the loss of historical **r̥* before *š* cf. *tišpa* ‘bitter’ from **tr̥špaka-* (Khromov 1987: 653), *fišóm-* ‘to send’ from *fra-šāma-*, S. *’βš’m*.
86. *pst’* (CS) in *pst’ mrx’ z’y* ‘low-laying flat land’ (SW., 146) — *past* ‘low-lying’, T. *past*.
87. *pšm-* ‘wool’, *’pšm’knty* ‘wool-stuffed’ (MacKenzie 1976: 60; So. 10200(8) in Yoshida 2007: 470 n. 27; Begmatov 2019: 40–41)[103] — *páši̯m*, T. *pašm*.
88. *ptγwnt-/ptγwst-* (CS) ‘to cover, to hide’ — *tuγunt-/tuγúntta* ‘to cover freshly sown seeds with soil’ (M., 197; N., 173).
89. *ptš’nkh, pcnk* ‘torture device, cross’ — *pašáng* ‘wooden part of a water-mill some 1,5 — 2 meters long’ (M., 153; N., 130).[104]
90. *ptšmʾr* ‘number, count’, *ptšm(’)r-* ‘to count, to think’ — *pušmára* (A.),[105] verb *pᵘšmar-* (Junker 1914: 11–16); S. *pt-* normally corresponds to Y *tu-/du-* (*ptxw’y-* ‘to kill’ — *tuxóy-*, *ptγwš-* ‘to listen’ — *duγúš-*; *ptr’yš-* ‘to tear apart’ — *dⁱríš-/dĭráyš-* and *ptγwnt-* — *tuγunt-* above; unlikely *ptr’yz-* — *tiráš-*, see under *š’š-*), but there are a handful of cases of the loss of *t*: *ptz’n-* and *bizón-* ‘to know’,[106] *ptš’nkh* — *pašang* and *pty’p* — *biyop* in the list; one can follow Junker and derive from **upa-(h/š)mar-* (and **apa-zān-*, B., 235) as well.
91. *ptxšnyk* ‘?’ (a hapax in P22, 8; ed. Benveniste 1940: 156) — *pĭtaxš-* ‘to cover’ (archaic). The words of S prediction *kδ xwr βγy krmyr-β’tk ptxšnyk β’t ZY kyw’n krm’yr-γwn’k wy[n’t* should be supposedly translated as ‘if the Lord Sun is *covered by red glove, and Saturn is visible in red colour’.[107]
92. *pty’p-* (MS) ‘to reach’ — *biyóp-/biyópta* ‘to catch up’, cf. under *ptšm’r* — *pušmára* on the preverb.[108]
93. *pyδ’r* postposition ‘for’ — *piyóra* postposition ‘for’, an archaic word, now *báxša, xotír*, noticed several times (B., 42, 101, 314-15, 881; Kh. 60, Ja. T., 375). The change of *δ* into *y* is irregular, but cf. *’’δc* — *oč* ‘something’, M. *kδ’m* — *kom* ‘which one’.

[101] Indistinctly written, confirmed by one informant in Kashi, but in an ‘unsure’ way.

[102] Definitely not ***kátaxan*.

[103] Begmatov follows Livshits (1962b: 155 n. 14) in reading *pšm* in Б-1, but it is hindered by the fact that a light-stem noun is very much expected. It is noteworthy that this reading option is absent in the reedition (Livshits 2008, 2015).

[104] The shaft of the wheel with its wedge, thus a cross-shaped device?

[105] Confirmed in 2008 by an elderly informant from Gharmen and, in an unsure way, the informant from Nomitkon.

[106] The variation of voiced and voiceless initial consonant follows the voicing of the first root consonant, see Bogoljubov 1957a.

[107] From **apa-tač-*, inch. **apa-taxš-*, cf. T. *andoxtan* ‘to throw’, Ossetic Digor *taxun* ‘to equip, dress up’ (Cheung 2007: 374–375)?

[108] The comparison with more common *’βy’p-* meets phonetic difficulties, and derivation from **api-āp-* (B., 101, 683) is possible (cf. Chorasmian *byfs-*), though such a verb is not attested in S.

94. *pyrwt* ‘moustache’ (Sundermann 2002: 141) — Y *burút* ‘moustache’ (F., 634), T. *burūt*, from Turkic (and ultimately Tocharian?) In view of initial *p-/b-* difference the direct continuation of S. word in Y. is unlikely.
95. **pys* ‘spotted, exemic(?)’ in the common personal name *pysk* (Lurje 2010a: 320–21; Rossi 2013) — *pes, peⁱs* ‘false lepra, eczema’(A.), T. *pes*, Middle Persian *pēs*.
96. *pyx* ‘knot, vertebra, node, joint’ — Y *pĭx, pĭnx* ‘splinter’, *pinx* ‘bird’s paw’ (Khromov 1966a: 134), also in Matcha T. *pix* ‘splinter, thorn’; semantics unclear. YY compares *pyx* ‘finger-joint’ with *pax* ‘finger’ (a rare word, surely related to *páx(x)a* ‘bunch of fives’).
97. *r’γ* ‘plain’ — *roγ* ‘piedmont’ (N., 143), T. *roγ*.
98. *rʾnʾ(kh)* ‘belt’ — *ron* ‘thigh’; T. *ron* (also YY).
99. *rʾt* ‘gift’ — *rot* ‘marriage gift’ (Khromov 2007; M., 170).
100. *r’w’th* ‘plain’ — *rówut, roṷt* ‘side mountain valley’ (Schwartz 1970: 301 with reference to Khromov); likely a *vr̥ddhi* formation of **rautah-* ‘river’ (Sims-Wiliams and Bi 2022: 495–96).
101. *rγzy* ‘fabric made of wool’ (Sims-Williams & Hamilton 2015: 30; mount Mugh and late documents) — *ráγza* ‘trousers made of wool’, ‘coarse home-made wool cloth’ (B., 810; Livshits 1962b: 155; Gh., 342; YY); T. *raγza* (glossed as Badakhshan and Kashmir word in medieval Persian dictionaries).
102. *rwc* ‘day’ (from Western Middle Iranian) or rather *rwcyk, rwtck* ‘provisions’ — *rouča, rűča* ‘fast’ (B., 49, 817), T. *rūza*.
103. *ro-m* ‘hair’ (Br., Maue & Sims-Williams 2024: 55–56) — *rőma* ‘pubic hair’ (hypercorrection from **rűma*?), *rűmoč* ‘id.’ (Kh., 179; also YY).
104. *rwps* ‘fox’ — *rűpas* ‘fox’ (Khromov 1987: 653; N. 145).
105. *rwtk* “entrails” (So14700(16)b/TiiT N, 11) Gershevitch, Frog-plain, p. 209) — *rúta* ‘intestine’ (Gershevitch 1975: 209; YY), T. *rūda*.
106. *rxʾ(h)* ‘face, mouth’ (L40, 2; Afr, 2) — Y *rax* ‘mouth, chaps, hole’ (Livshits 2015: 240); cf. also (Kh., 135; Novák 2013: 214 F., 642), T. *rax* ‘crack’.
107. *ryrʾkH* ‘part of mouth, saliva’ (Henning 1940: 5, fr. IV, 10; Sundermann 2002: 140) — Y *réra/rira* ‘gums’(Khromov 2007).
108. *rys-* ‘to lick’ (VJ, 990 ed. Benveniste 1946) — Y *les-* ‘to lick’, T. *lesidan* (noted by B., 762).
109. *ryž-* ‘to desire’ or *ryz’kh* ‘rice’ — *(γurda)rižóna* ‘rice’ (secret word, B., 717, 827; N., 62); ‘healer who uses to cure eyes with grain or rice; grain’ (M., 61; N., 62). One would tend to compare it to S. *ryz’kh* ‘rice’ (Bailey 1979: 364; Gh., 349; YY); however, despite Pašto *(w)ríža*, the Manichaean form glossing Middle Persian *(b)rync* is *ry(z)[’*, not ***ry(j)[’* (M 4164, 2 in Morano 2005: 221), and in Brāhmī script we see *rzā* (Maue & Sims-Williams 2024: 66–68) so Yaghnobi *ž* is unexpected. Moreover, until recently rice was a very expensive and rare commodity in the mountain region (Dynin 1914: 84), and in pre-Islamic layers of Sogdiana one has not met rice (*oryza sativa*) during archaeobotanical analysis (Mir-Makhamad et al. 2024: 19); perhaps it was limited to élite consumption. I suppose that here the Sogdian root *rēž-* ‘to desire’ was frozen to indicate a (mostly) luxury foodstuff.
110. *s’γr’, swγr’* ‘saddle’ (?)[109] — *xüšsaγrī́* ‘horse (with beautiful crupper)’ (B., 169), T. *saγrī* ‘crupper, *la peau de chagrin*’, Persian *săγrī*, usually considered a Turkic loan (Doerfer 1965: III, 222–225).

[109] Mugh, Б-19, 6; Б-12, 3 (Livshits 2015: 132–33); in the first example, Bogoljubov and Smirnova read *γnγr’* (for *xnγr* ‘dagger’?; final aleph might indicate a position after numeral), Begmatov (2019: 185) follows Livshits, who adds Ossetic, Pashto *sarγ*. In the second case, the reading *swx’n* as a PN is preferred (Bogoljubov & Smirnova 1963: 48; Lurje 2010a: 358; Begmatov 2019: 117–118).

111. *s'ṭ* 'all' — *sóda* 'forever' (Ja. T., 323; M., 182; N., 155); the correspondence is problematic due to unexpected vocalization and change of meaning.
112. *swnp-, swmb-/swβt* 'to bore' — possible nominal derivates in *sūf* 'aperture in the millstone' (B., 850), *sum* 'a hole in a shoe' (N., 156), *súmba* 'a piston in a hollow piece of wood used as a kind of blow pipe game'[110].
113. *syncyn* '?', a hapax in the late Dunhuang document E5 (Sims-Williams & Hamilton 2015: 49–53), cf. *'ycyn, íčin*[111] — *sínčin* 'needle'.[112]
114. *-šwβt-* 'dense' in *w'šwβtt* 'aussi serré' (P2, 1016, ed. Benveniste 1940: 48)[113] — *šuft* 'dense, tight' (M., 251;[114] N., 162)
115. *šwt'* (CS) 'furrow, field' (E26/1 V12 ed. Sims-Williams 2015: 20) — *šudgór* 'arable land, tillage' (Ya. T., 328), T. *šudgor*.[115]
116. *šx-* 'branch' + *'stk* 'bone' — *šĭxastáka, šĭkastáka* 'armpit'. YY compares *š(n)x-* to *šox*, (which corresponds to S. *š'x,* T. *šox* 'branch' better in its vocalism).
117. *šyš-/šyštk* 'to be scattered, dispersed', *š'š-, šyš-, δr''š-* 'to scatter, to disperse' — *tĭrá(y)š-/ tĭrašta* ''to fall' in the western dialect (Cheung 2007: 397).[116]
118. *-tʾr-* 'string' in *cmbt̲ʾryh̲* 'string of a lute' (Henning 1940: 16, a 1); **δst'rch* 'ten-stringed' (SW., 71) — *tor* 'thread, string', T. *tor*.
119. *tʾrʾc* (MS) 'plunder, booty', 'the final world period' (M767iiR5 in Haloun and Henning 1952: 212; S.-D., 190) — *toroǰ* 'plunder', T. *toroj*, from Turkic (Doerfer 1965: II, 435–436).
120. *tbr-/*tγt-* (CS) 'to insert' — *dĭvár-, dĭγáta* 'to bring in, lead in' (Sims-Williams et al. 2014: 40; YY).
121. *tβs-/tβt-* 'to get hot' — *tasp* 'heat, sultry weather', T. *tafs*, *tasp,* verb *tafsidan* (a S. loan?).
122. *tγʾrʾ* 'khum, pythos' (Livshits 2015: 274–278) — *taγorá* 'wooden dish', T. *taγora* 'basin'.
123. *tpy* (CS) 'jar, pitcher', *rwγny tpy* 'jar of the oil' (E27/94v21 ed. Sims-Williams 1985: 170, 177; SW., 194)[117] — *tĭbángča* 'wooden dish for oil', T. dialectal *tabanč*, standard *tabaq*. As reminded by Nicholas Sims-Williams, S *tpn'* 'coffin', Persian *tabangōy*, Chorasmian *tpnkwk* might also belong here (Henning 1965: 177).
124. *tr-, tyr* (CS) 'to go' (SW., 194) — *tir-/tórta* 'to go (out)' (also YY).
125. *tr'nk(')* 'torture' — *taráñg kun-* 'to pull, tighten', T. *tarang* (comparison of B., 857).

---

[110] Long and partially contradictory explanations in M., 184; N., 156.

[111] The expression *nm''y βykp'r trny 'ycyn šxy syncyn βyrty L' (x)cy* is translated as 'In this (person), (who is) outwardly soft (but) **inwardly** hard, no **faith** (?) is (to be) found' (semibold preserved as indicating hypothetic translation); James Hamilton explains *'ycyn as* Turkic *ičin*, instrumental of *ič* 'interior', and in *syncyn* an instrumental of *sinči* 'general appearance, dress etc.', which appears only in late Anatolian Turkish and Mongolic *šinǰi* in the sense 'omen, notice' (the common Uighur antonym to *ič* 'inside' is however *taš* 'outside'). The general meaning is 'outwardly soft, inwardly hard', an opposite to the Chinese auspicious concept 'inside softness, outside hardness' of the "Book of Changes", or 'outwardly mild, inwardly hard' in Tangshu; this explanation presupposes the knowledge of Chinese classics by a Turko-Sogdian textile trader.

I plead for a totally different option: 'it seems (BS *nm'y*) outside one cannot find (either) soft thread (or) hard needle'; the author seemed to search for higher quality goods.

[112] Also in a series of expressions: *i sinčini sari* 'on the top of the needle, on the edge of extinction' (Basello, Guizzo, and Ognibene 2008); *anáx xápi sínčin nóstišt* 'he/she fancies, imagines (takes his/her needle)' (Kh., 180); in the secret language it denotes 'child', 'person', 'official' (M., 179; N., 153).

[113] Differently, to Persian *sift* 'tight, thick' (Henning 1946: 724).

[114] 'Зич, чафс'.

[115] In Persian *šudgār* is a rare word; cf. early spellings *šuδyār/šulyār/šuδkār* (Lazard 1963: 146).

[116] The comparison of *tĭraš-* to *ptr'yz-* 'to raise up, to erect' (Ja. T., 336) is hindered by semantics and irregular correspondence of *ž* and *š*. The relation to S. *ptryš-* 'to tear apart' is also questionable in view of regular Y *dʲríš-, dĭráyš-*.

[117] Perhaps also *tpc* 'measure of barley' in Hisorak document 3, if not to be read as *tsw* (Lurje 2012: 472).

126. *trp'r'k* 'acid' — *tíšpa* 'sour' (B., 147; Novák 2013: 95); different developments of Old Iranian **tr̥p-ša/ra-*.
127. *tš-, tšycq* 'kind of axe' — Y *tešá* 'small hoe', T. *teša*; in view of vowel and accent, a T. loan.
128. *wʾz* 'wing' — *vóza, wóza* "open oven made of stones; half of close oven tandur', semantics remain unclear.
129. *wγnh* 'famine, starvation' — *wáγnič* 'small plant with edible berries' (the one consumed because of big hunger?)
130. *wr'n* (CS) 'lamb' — *rōn, run, ruwón* 'lamb' (B., 57, 63, 824; Schwartz 1982: 63; Tremblay 2006: 277; YY).
131. *wrnh, wrn*' 'wool' (P19, 20 in Benveniste 1940: 232; So 14822 in Reck 2016 No. 724) — *vánna, wánna* 'fallen snake-skin'?
132. *wryc-/wrγ(t)-* 'to pour' — *rīxs-/rǐxta-* 'to crumble', as noted by YY, Y. form is built with inchoative suffix *-s* and does not necessarily contain a preverb. T. *rextan* differs in its present stem *rez*.
133. *wšyny, wšwynʾk, šwyːnny* etc. 'cloth made of linen or cotton' (Sims-Williams & Hamilton 2015: 67–68), Bactrian *οιβοηγγο* (Sims-Williams 2002: 228) — *šiwína, šawén* 'home-made cotton garment', T. *šavina*.[118]
134. *wyt̲ δʾrwk* 'willow-tree?' (M2006 R4 in Boyce 1960: 85; S.-D., 211; Sims-Wiliams & Bi 2022: 495) — *wet* 'willow'.
135. *wyt, wyt'k* 'cord' (several times in Mugh documents, Livshits 2015: 136; Begmatov 2019: 241; Sims-Wiliams and Bi 2022: 495) — *wíta* 'cord, rope' (also YY).
136. *wyx* 'root' — *wexčák* 'a kind of plant which is used as fuel when dry' (M.,[119] N., 184)
137. *xʾm(ʾk)* 'raw' — *xom* 'raw', T. *xom*.[120]
138. *xnt-* 'to laugh' (VJ, 1355, CS E30/6v16) — Y *xant- /xánta-* 'to laugh', T. *xandidan* (noted by B., 883); YY, adds S. *'sx'nt, qtsx'ncyk* 'joke'.
139. *xrwm, xwrm, xwrwm* 'earth' — *xŭrmá* 'earthenware bowl', T. *xurma* (semantics unclear).
140. *xwc-* 'nice, pleasant' — *xwač* 'nice' (Old Yaghnobi, B., 42, 892, quoting Kuhn and noticing Sogdian counterpart).
141. *xwž-/xwšt-* 'to ask, request, beg' — *xúǰgo̯n, xŭžgo̯n, xŭšgo̯n, xŭšγón*, B. *xwžgún* 'matchmaker' (in wedding),[121] T. *xostgor* (from *xostan* 'to demand').
142. *xyp* 'own', the late form of *xypδ* (see 4e above) — *xep, xap* 'own, private'.
143. *xyr'k* 'dim, stupid' — *xíra* 'dark, dim, dirty', T. *xira* (MacKenzie 1976: 105; Gh., 181).
144. *-y'w* suffix of language 'in X-ian' — *-yow,* suffix of similarity (Gershevitch apud Khromov 1970: 230 n. 8).
145. *ywc-/ywxt-* 'to teach' — *yučón-* 'to teach, admonish, moralize' (Kh., 186, with etymology).
146. *yw̯γ/yγw* (CS) 'yoke' (E28/18r27 Sims-Williams 2017: 114) *ywγk*' 'yoke-bone' (Ch/So 16201/iii/41 Sundermann 2002: 142), *ywγ* 'name of a lunar mansion' (M549i+R1) — *yuγ* 'yoke' (F., 634), T. *yuγ* (also YY).
147. *yxn-* 'ice' (Sims-Williams apud Sundermann 1990: 26, n. 33) *–ix, ex* 'ice', *īxn* 'icy' (B., 157) and likely *Yaγnób* itself (Khromov 1987: 644), a very cold river.
148. *z'ry* 'pitifully' — *zor-zor, zorí* 'pityfully', T. *zor, zor-zor*.
149. *z̨yβ-, jβ-/z̨yβ't-* 'bite, chew' — Y *žav-/žafta* 'eat, slurp'

[118] My records in Panjakent; apparently not in dictionaries.

[119] In a contribution that I was unable to access.

[120] B., 881, provides *xomak* 'rawhide, bastard, miscarriage'.

[121] From Tajiki (with question mark) in Ja. T., 362, but I have not found such a word in T. The suffix is unclear, S. *-k'n, -k'ny* is denominal (Gershevitch 1954: 162).

150. *z-βr'yt'k* 'burnt, roasted' (So 14185(2)r, 123; Yoshida apud Reck 2009: 389; S.-D., 232) — *viráy-/viráyta* 'to roast'.
151. *zrγwn'k* 'green', *zrγwn* 'vegetable' — *zarγúna* 'yellow' (B., 186; N., 203).
152. *(')zr'ync-* 'to save, to release' — *zirínč* 'wooden gate' (Kh., 186), the similarity noticed by Schwartz (1982: 125), but the semantics remain unclear (both from 'asylum'?).

8. Indexes

Index A (Sogdian forms quoted in the dictionary of *Jagnobskie teksty*), Index B (Yaghnobi forms quoted in the 'Sogdian Dictionary' of Badrizamon Gharib) and Index C to this paper are included as online supplementary materials at https://jolr.ru/jlr24/lurje.pdf.

Literature

Aguirre Beltran, Gonzalo. 1979. *Regions of Refuge. The Society for Applied Anthropology Monograph Series Vol. 12*. Washington DC: The Society for Applied Anthropology.

Akimbetov, Sh. 1881. Ocherki Kogistana. *Turkestanskie vedomosti* 2: 7–8, 3: 10–11.

Alborov, B. A. 1934. *Jagnobskii jazyk. Chast' III. Slovar'*. Stalinabad.

Andreas, F. C. 1910. Zwei soghdische Exkurse zu Vilhelm Thomsens: Ein Blatt in türkischer Runenschrift. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* 1910: 307–314.

Andreev, M. S. 1970. *Materialy po etnografii Jagnoba*. (Podgotovila k pechati A. K. Pisarchik.) Dushanbe: Donish.

Andreev, M. S., E. M. Peshchereva. 1957. *Jagnobskie teksty. S prilozheniem jagnobsko-russkogo slovarya, sostavlennogo M. S. Andreevym, V. A. Livshitsem i A. K. Pisarchik*. Moskva / Leningrad: Nauka.

Arandarenko, G. A. 1889. *Dosugi v Turkestane 1874–1889*. Sankt-Peterburg: Stasyulevich.

Babadzhanov, B. B. 2010. *Kokandskoe khanstvo — vlast', politika, religija*. Tashkent / Tokio: Yangi Nashr.

Bailey, Harold W. 1979. *Dictionary of Khotan Saka*. Cambridge University Press.

Bartholomae, Christian. 1908. Zu den arischen Wörtern für 'der erste'und 'der zweite'. *Indogermanische Forschungen* 23: 43–92.

Basello, Gian Pietro, Daniele Guizzo, Paolo Ognibene (eds.). 2008. *Sulla punta di uno spillo. Rapporto scientifico sulla Prima missione etnolinguistica nella valle della Yaghnob*. Ravenna / Milano: Mimesis.

Begmatov, Alisher. 2019. *Sogdian Textual Materials from Central Asia: A Critical Re-Edition of the Documents from Mount Mugh. Unpublished PhD thesis*. Kyoto University.

Benkato, Adam. 2020. Multiple Copularization in Sogdian and Yaghnobi. *Transactions of the Philological Society* 118(1): 172–91. doi: 10.1111/1467-968X.12181.

Benveniste, Benveniste, Émile. 1979. *Études Sogdiennes*. *Beiträge zur Iranistik Vol. 9*. Wiesbaden: Reichert.

Benveniste, Émile. 1929. *Essai de grammaire Sogdienne. Deuxieme partie*. *Mission Pelliot en Asie Centrale, série petit in-octavo. Vol. III*. Paris: Librairie Orientaliste Paul Geuthner.

Benveniste, Émile. 1940. *Textes Sogdiens. Édités, traduits et commentés par E. Benveniste*. Paris: Librairie Orientaliste Paul Geuthner.

Benveniste, Émile. 1946. *Vessantara Jātaka. Texte sogdien. Édité, traduit et commenté par Émile Benveniste*. Mission Pelliot en Asie Centrale, série In-Quatro. Paris: Librairie Orientaliste Paul Geuthner.

Benveniste, Émile. 1955. Un lexique du Yagnobi. *Journal Asiatique* CCXLIII(2): 139–162.

Benveniste, Émile. 1959. (Compte-rendu) M. S. Andreev — E. M. Peščerova. Yagnobskie Teksty. *Bulletin de la Société de linguistique de Paris* 54: 79–81.

Benveniste, Émile. 1966. Le verbe stā- comme auxiliaire en iranien. *Acta Orientalia* 30: 45–49.

Bi, Bo, Nicholas Sims-Williams, Yan Yan. 2017. Another Sogdian–Chinese Bilingual Epitaph. *Bulletin of the School of Oriental and African Studies* 80(2): 305–318.

Bielmeier, Roland. 1989. Yaghnōbī. In: R. Schmitt (ed.). *Compendium linguarum iranicarum*: 480–488. Wiesbaden: Harrassowitz.

Blois, François de. 2024. From Choresmia to Cathay: A New Look at al-Bayrūnī's Chronology. *The Medieval History Journal* 27(1): 173–193. doi: 10.1177/09719458241248580.

Bogoljubov, M. N. 1956. *Jagnobskij (novosogdiiskij) jazyk. Issledovanija i materialy*. D. filol. nauk. Leningrad: LGU.

Bogoljubov, M. N. 1957a. Jagnobo-sogdijskie dialektologicheskie otnoshenija. *Vestnik Leningradskogo Universiteta. Serija istorii, jazyka i literatury* 8(2): 111–113.

Bogoljubov, M. N. 1957b. *Jagnobskii (novosogdijskij) jazyk. Issledovanija i materialy*. Avtoreferat dissertatsii na soiskanie uchenoi stepeni doktora filologicheskikh nauk. Leningrad: LGU.

Bogoljubov, M. N., O. I. Smirnova. 1963. *Khoz'ajstvennye Dokumenty. Sogdijskie Dokumenty s Gory Mug*. Moskva: Vostochnaja literatura.

Boyce, Mary. 1960. *A Catalogue of the Iranian Manuscripts in Manichean Script in the German Turfan Collection*. Berlin: Akademie Verlag.

Cheung, Johnny. 2007. *Etymological Dictionary of the Iranian Verb*. *Leiden Indo-European Etymological Dictionary Series Vol. 2*. Leiden / Boston: Brill.

Cilli, Elisabetta, Sara De Fanti, Paolo Delaini, Antonio Panaino, Giorgio Gruppioni. 2013. A Contribution to the Genetic History of the Yaghnobis of Tajikistan through Mitochondrial DNA Analysis. In: A. Panaino, A. Gariboldi, P. Ognibene (eds.). *Yaghnob Studies I. Papers from the Italian Missions in Tajikistan*. *Indo-Iranica et Orientalia. Series Lazur, Vol. 12*: 141–150. Milano / Udine / Paris: Mimesis.

Delaini, Paolo. 2013. Traditional Medicine in the Yaghnob Valley. Notes on Body Therapy, Pregnancy and Birth Care. In: A. Panaino, A. Gariboldi, P. Ognibene (eds.). *Yaghnob Studies I. Papers from the Italian Missions in Tajikistan*. *Indo-Iranica et Orientalia. Series Lazur, Vol. 12*: 119–140. Milano / Udine / Paris: Mimesis.

Djakonov, I. M., K. V. Kaufman, V. A. Livshits. 1954. O drevnej sogdijskoj pis'mennosti Bukhary. *Vestnik drevnej istorii* 1954(1): 150–163.

Doerfer, Gerhard. 1965. *Türkische und mongolische Elemente in neupersischen*. *Vol. II*. Wiesbaden: Steiner.

Doerfer, Gerhard. 1967. *Türkische Lehnwörter im Tadschikischen*. *Abhandlungen für die Kunde des Morgenlandes Vol. XXXVII, 3*. Wiesbaden: Steiner.

Durkin-Meisterenst, Desmond. 2008. Yaghnobi. In: B. Huber, M. Volkart (eds.). *Chomolangma, Demawend, Kasbek. Festschrift Ronald Bielmeier*: 413–422. Halle (Salle): IITBS.

Durkin-Meisterernst, Desmond. 2006. *The Hymns to the Living Soul: Middle Persian and Parthian Texts in the Turfan Collection*. *Berliner Turfantexte*. Turnhout: Brepols.

Dybo, A. V. 2007. *Lingvisticheskie kontakty rannikh t'urkov*. Moskva: Vostochnaja literatura.

Dynin, V. V. 1914. Ocherk byta gortsev verkhov'ev Zeravshana (opyt ekonomicheskogo issledovanija). *Izvestija Turekskanskogo otdela imperatorskogo russkogo geograficheskogo obshchestva* X(1): 37–124.

Edel'man, D. I. 1986. *Sravnitel'naja grammatika vostochnoiranskikh jazykov. Fonologija*. Moskva: Nauka.

Egani, A. A., O. D. Chekhovich. 1984. Regesty sredneaziatskikh aktov III. Pis'mennye pam'atniki Vostoka. *Istoriko-filologicheskie issledovanija. Ezhegodnik*. 1976–1977: 105–10, 321–362.

Efimov, V. A. 2009. *Jazyk parachi. Grammaticheskij ocherk. Teksty. Slovar'*. Moskva: Vostochnaja literatura.

Filippone, Ela. 2013. Yaghnobi Body Part Terms: Some Introductory Notes. In: S. R. Tokhtasyev, P. B. Lurje (eds.). *Commentationes Iranicae. Sbornik statej k 9-letiju V. A. Livshitsa*: 631–648. Sankt-Peterburg: Nestor — Istorija.

Gauthiot, Robert. 1911. À propos des dix premiers noms de nombre en sogdien bouddhique. *Mémoirs de la Société de Linguistique de Paris* XVII(3): 137–161.

Gauthiot, Robert. 1913. [Compte rendu de la mission au Turkestan russe]. *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 57(9): 671–680.

Gauthiot, Robert. 1914. *Essai de Grammaire Sogdienne. Première partie*. Paris: Geuthner.

Geiger, Wilhelm. 1898. Über das Yaghnōbī. In: W. Geiger, E. Kuhn (eds.). *Grundriss der iranischen Philologie*. Vol. I/2: 334–344. Strassburg: Trübner.

Gershevitch, Ilya. 1946. Sogdian Compounds. *Transactions of the Philological Society* 46: 137–149.

Gershevitch, Ilya. 1954. *A Grammar of Manichaean Sogdian*. Oxford: Blackwell.

Gershevitch, Ilya. 1959. *The Avestan Hymn to Mithra*. *University of Cambridge Oriental Publications*. Cambridge University Press.

Gershevitch, Ilya. 1975. Sogdians on a Frogplain. In: M. Dj. Moïnfar (ed.). *Mélanges linguistiques offerts à Émile Benveniste*. *Collection linguistique publiée par la Société de linguistique de Paris, Vol. LXX*: 195–211. Paris: Peeters.

Gharib, Badrezzaman. 1995. *Sogdian Dictionary. Sogdian, Persian, English*. Tehran: Farhangan.

de Goeje, M. J. (ed.). 1906. *Descriptio Imperii Moslemici Auctore Al-Mukaddasi. Editio Secunda*. *Vol. III. Bibliotheca Geographorum Arabicorum*. Lugdini Batavorum: Brill.

Guizzo, Daniele. 2013. Who Are the Yaghnobis? Past, Present and Future of a Minority. In: A. Panaino, A. Gariboldi, P. Ognibene (eds.). *Yaghnob Studies I. Papers from the Italian Missions in Tajikistan*. *Indo-Iranica et Orientalia. Series Lazur, Vol. 12*: 151–166. Milano / Udine / Paris: Mimesis.

Gunya, A. N. 2002. *Yaγnob Valley. Nature, History and Changes of a Mountain Community Development in Tadjikistan. Institute of Geography*. Moscow: KMK.

Haloun, G., Walter B. Henning. 1952. The Compendium of the Doctrines and Styles of the Teaching of Mani, the Buddha of Light. *Asia Maior. New Series* III(2): 184–212.

Henning, Walter B. 1936. *Ein Manichäisches Bet- und Beichtbuch*. *Abhandlungen der preussischen Akademie der Wissenschaften, Vol. 10*. Berlin: Walter de Gruyter.

Henning, Walter B. 1939. Sogdian Loan-Words in New Persian. *Bulletin of the School of Oriental and African Studies* 10(1): 93–106.

Henning, Walter B. 1940. *Sogdica*. *Vol. II*. London: Forlong.

Henning, Walter B. 1945. Two Central Asian Words. *Transactions of the Philological Society* 44: 150–162.

Henning, Walter B. 1946. The Sogdian Texts of Paris. *Bulletin of the School of Oriental and African Studies* XI(4): 713–740.

Henning, Walter B. 1958. Mitteliranisch. In: *Iranistik. Linguistik*. *Handbuch der Orientalistik, Vol. IV*: 20–130. Leiden / Köln: Brill.

Henning, Walter B. 1965. The Choresmian Documents. *Asia Maior. New Series II*: 166–179.

Huyse, Philip. 1999. *Die dreisprachige Inschrift Šābuhrs I. an der Ka'ba-i Zardušt*. *Corpus Inscriptionum Iranicarum Vols. I–II*. London: School of Oriental and African Studies.

Isakov, A. I., Yu. Ya. Yakubov, G. R. Karimova. 2020. *Verkhov'ja doliny Zarafshana. T. I–II. Arkheologicheskaja karta Tadzhikistana*. Dushanbe: Donish.

Iskander-t'ur'a [A. L. Kun]. 1881. Svedenija o jagnauskom narode. *Turkestanskie vedomosti* 3–4.

Jügel, Thomas. 2016. Enclitic Pronouns in Middle Persian and the Placeholder Construction. *Quarterly Journal of Language and Inscription. Faslnâme-Ye Takhassosi-Ye Zabân-o Katibe* I(1): 41–63.

Junker, Heinrich F. J. 1914. *Drei Erzählungen auf Yaγnābī*. *Sitzungsberichte der Heidelberger Akademie der Wissenschaften*. Heidelberg: Winter.

Khromov, A. L. 1962. Govory tadzhikov matchinskogo rajona. *Trudy AN TadzhSSR*. Tom CVVII. Dushanbe: AN Tadzh SSR.

Khromov, A. L. 1966a. Novye materialy po leksike jazyka jagnobtsev. *Acta Orientalia* XXX: 129–135.

Khromov, A. L. 1966b. Obshchaja lingvisticheskaja kharakteristika toponimii i mikrotoponimii Jagnoba. *Izvestija Akademii nauk Tadzhikskoi SSR, otdelenie obshchestvennykh nauk* 45(3): 83–87.

Khromov, A. L. 1970. Die Präposition par und Postposition *-yow* im Yaghnobi. In: M. Boyce, I. Gershevitch (eds.). *Henning Memorial Volume*, *Asia Major Library*: 228–230. London: Lund Humphries.

Khromov, A. L. 1972. *Jagnobskij jazyk*. Moskva: Nauka.

Khromov, A. L. 1987. Jagnobskij jazyk. In: V. S. Rastorgueva (ed.). *Osnovy iranskogo jazykoznanija. Novoiranskie-jyazyki: vostochnaja gruppa*: 644–701. Moskva: Nauka.

Khromov, A. L. 2007. Rot — svadebnyj podarok u jagnobtsev (etimologicheskie zametki). In: V. A. Efimov, E. K. Molchanova, V. B. Ivanov, L. R. Dodykhudoeva (eds.). *Pam'ati V. S. Rastorguevoi. Sbornik statei*: 199–203. Moskva: Kl'uch-S.

Kienast, Burkhard. 1980. Ilku' edited by D. O. Edzard. *Reallexikon Der Assyriologie Und Vorderasiatischen Archäologie* 5: 52–59.

Klimchitskii, S. I. 1937. Jagnobsko-sogdijskie sootvetstvija. Zapiski instituta vostokovedenija AN SSSR VI: 15–25.

Klimchitskii, S. I. 1938. Sekretnyj jazyk u jagnobtsev i jazgulemtsev. Trudy Tadzhikistanskoj bazy Akademii nauk SSSR IX: 104–117.

Kol'chenko, V. A., P. B. Lurje. V pechati. Nadpis' Na Khume Iz Gorodishcha Ken-Bulun. *Arkheologiya Evraziiskikh Stepei*.

Korn, Agnes. 2016. A partial tree of Central Iranian. *Indogermanische Forschungen* 121: 401–434.

Korn, Agnes. 2019. Isoglosses and Subdivision of Iranian. *Journal of Historical Linguistics* 9(2): 239–281.

Kun, A. L. 1871. *Turkestanskij al'bom. Chast' etnograficheskaja. Tuzemnoe naselenie*. T. I. Tashkent: Lit. Voenno-topogr. Otdela Turkest. Voen. Okruga.

Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.

Livshits, V. A. 1962a. Iranskie jazyki narodov Srednei Azii. In: P. Tolstov, T. A. Zhdanko, S. M. Abramzon, N. A. Kisl'akov (eds.). *Narody Srednej Azii i Kazakhstana. Narody Mira. Etnograficheskie ocherki*: 131–158. Moskva: Izdatel'stvo Akademii nauk SSSR.

Livshits, V. A. 1962b. *Juridicheskie dokumenty i pis'ma. Sogdijskie dokumenty s gory Mug*. Moskva: Vostochnaja literatura.

Livshits, V. A. 2008. *Sogdijskaya epigrafika Srednej Azii i Semirech'ja*. Sankt-Peterburg: Filologicheskij fakul'tet SPBGU.

Livshitz, Vladimir A. 1970. A Sogdian Alphabet from Panjikant. In: M. Boyce, I. Gershevitch (eds.). *Henning Memorial Volume*, *Asia Major Library*: 256–263. London: Lund Humphries.

Livshits, Vladimir A. 2015. *Sogdian epigraphy of Central Asia and Semirech'e*. *Vols. II–III (ed. by N. Sims-Williams)*. *Corpus Inscriptionum Iranicarum*. London: School of Oriental and African Studies.

Livshits, V. A., A. L. Khromov. 1981. Sogdiiskii yazyk. In: V. S. Rastorgueva (ed.). *Sredneiranskie yazyki. Osnovy iranskogo yazykoznaniya*, t. 2: 347–514. Moskva: Nauka.

Loy, Thomas. 2005. *Jaghnob 1970. Erinnerungen an eine Zwangsumsiedlung in der Tadschikischen SSR*. Wiesbaden: Reicher.

Lurje, P. B. 2004. *Istoriko-lingvisticheskij analiz sogdijskoi toponimii*. Dissertatsiya... kandidata filologicheskikh nauk. Sankt-Peterburg.

Lurje, P. B. 2008. K chtenijam sogdijskikh nadpisej, otkrytykh v Pendzhikente. In: M. E. Rezvan (ed.). *Rakhmatname. Sbornik statei k 70-letiyu Rakhmata Rakhimovicha Rakhimova*: 213–225. Sankt-Peterburg: MAE RAN.

Lurje, P. B. 2010. Otchet o poezdke v Jagnob v ijune 2008 g. *Materialy Pendzhikentskoi arkheologicheskoi ekspeditsii XII*: 10–19.

Lurje, P. B. 2012. Sogdijskie dokumenty iz raskopok rannesrednevekovogo Martshkata (predvaritel'noe soobshchenie). *Indoevropejskoe yazykoznanie i klassicheskaja filologija. Materialy chtenii pam'ati I. M. Tronskogo* XVI: 433–460.

Lurje, P. B. 2015a. Toponimicheskie dannye o bukharskom dialekte sogdijskogo jazyka. In: A. V. Omel'chenko (ed.). *Bukharskij oazis i ego sosedi v drevnosti i srednevekov'e. Na osnove materialov nauchnykh konferentsij 2010 i 2011 gg.*: 127–142. Sankt-Peterburg: Gosudarstvennyj Ermitazh.

Lurje, P. B. 2015b. Jagnobskaya skazka ob Ali i Imame Makhdi. In: M. S. Pelevin (ed.). *Na pastbishche mysli blagoj. Sbornik statej k jubileju I. M. Steblina-Kamenskogo*: 154–163. Sankt-Peterburg: Kontrakt.

Lurje, P. B. Forthc. K administrativnoj geografii dokumentov s gory Mug. *Vestnik MITsAI*.

Lurje, Pavel. 2022. Review of Payne, Richard E. / King Rhyne (Hg.): The Limits of Empire in Ancient Afghanistan. *Orientalistische Literaturzeitung* 117(3): 257–261. doi: 10.1515/olzg-2022-0089.

Lurje, Pavel B. 2010a. *Personal Names in Sogdian Texts*. *Iranisches Personennamenbuch.* Vienna: Österreichische Akademie der Wissenschaften.

Lurje, Pavel B. 2010b. Rev. Literarische Stoffe und ihre Gestaltung in Mitelliranischer Zeit, Ehrenkolloquium anlässlich 70 Geburtstag Werners Sundermanns. *Orientalistische Literaturzeitung* 105: 779–784.

Lurje, Pavel B. 2016. Le mot sogdien pour "digue". In: M. Espagne, S. Gorshenina, F. Grenet, S. Mustafaev, C. Rapin (eds.). *Asie Centrale : transferts culturels le long de la Route de la soie*: 239–246. Paris: Vendémiaire.

MacDonnell, A. A. 1910. *Vedic Grammar*. *Vol. I, 4. Grundriss Der Indo-Arischen Philologie Und Altertumskunde*. Strassburg: Trübner.

MacKenzie, David Neil. 1970. *The Sutra of The Causes and Effects of Action in Sogdian*. Oxford University Press.

MacKenzie, David Neil. 1976. *The Buddhist Sogdian Texts of the British Library*. *Acta Iranica*. Téhéran-Liège: Bibliothèque Pahlavi.

Maue, Dieter, Nicholas Sims-Williams. 2024. *Sogdian Documents in Brāhmī Script*. *Corpus Inscriptionum Iranicarum*. London: School of Oriental and African Studies.

Mir-Makhamad, Basira, Pavel Lurje, Vikentiy Parshuto, Abdurahmon Pulotov, Firuz Aminov, Michael Shenkar, Muminkhon Saidov, Nikita Semenov, Sharof Kurbanov, Sirojiddin Mirzaakhmedov, Khusniddin Rakhmanov, Rita Dal Martello, Robert Spengler. 2024. Agriculture along the Upper Part of the Middle Zarafshan River during the First Millennium AD: A Multi-Site Archaeobotanical Analysis. *PLOS ONE* 19(3):e0297896. doi: 10.1371/journal.pone.0297896.

Mirzozoda, Saifiddin. 2008. *Farhangi Yaghnobī-Toɣikī*. Dushanbe: Anjumani Devashtich.

Morano, Enrico. 2005. Preparing the Corpus of the Sogdian Texts for the SOAS Manichaean Dictionary Project. In: C. G. Cereti, M. Maggi. *Middle Iranian lexicography, Proceedings of the Conference held in Rome, 9-11 April 2001*, *Serie Orientale Roma*: 215–226. Roma: Accademia nazionale dei lincei.

Mukhtarov, Akhror. 1978. *Epigraficheskie pam'atniki Kukhistana*. Tt. I, II. Dushanbe: Donish.

Müller, F. W. K. 1904. *Handschriften-Reste in Estrangelo-Schrift aus Turfan, Chinesisch-Turkistan*. *Abhandlungen der königlich Preußischen Akademie der Wissenschaften. Philologisch-historische Klasse*. Berlin: Georg Reimer.

Müller, F. W. K. 1908. *Uigurica [I]. Abhandlungen der königlich Preußischen Akademie der Wissenschaften*. Berlin: Georg Reimer.

Müller, F. W. K. 1912. *Soghdische Texte. I*. In: *Abhandlungen der königlich Preußischen Akademie der Wissenschaften*: 199–309. Berlin: Georg Reimer.

Novák, Ľubomír. 2010. *Jaghnóbsko-český slovník: s přehledem jaghnóbské gramatiky. Vyd. 1*. Praha: Univ. Karlova v Praze, Filozofická Fak.

Novák, L'ubomír. 2013. *Problem of Archaism and Innovation in the Eastern Iranian Languages. PhD Thesis*. Charles University, Prague.

Novák, Ľubomír. 2014. *Historical Phonology of Yaghnōbī and Sogdian*. Ms. (draft).

Novák, L'ubomír. 2018. Yaghnobi: An Example of a Language in Contact. *Chatreššar International Journal for Indo-European, Semitic, and Cuneiform Languages* 2: 26–50.

Ogihara, Hirotoshi. 2019. Tumshuqese Imperfect and Its Related Forms. In: P. Lurje (ed). *Proceedings of the Eighth European Conference of Iranian Studies*: 297–310. Saint Petersburg: State Hermitage Museum.

Panaino, Antonio. 2008. Attraversando la Valle dello Yaynob: passato, presente e (possibile) futuro di una minoranza etno-linguistica. In: G. P. Basello, D. Guizzo, P. Ognibene (eds.). *Sulla punta di uno spillo. Rapporto scientifico sulla Prima missione etnolinguistica nella valle della Yaghnob*: 1–78. Ravenna / Milano: Mimesis.

Panaino, Antonio. 2013a. The Italian Scientific Mission to Tajikistan. The Case of Yaghnob Valley. In: P. B. Lurje, A. I. Torgoev (eds.). *Sogdians, their precursors, contemporaries and heirs. Based on proceedings of conference "Sogdians at Home and Abroad" held in memory of Boris Il'ich Marshak (1933–2006), Transactions of the State Hermitage*: 461–476. Saint Petersburg: State Hermitage Museum.

Panaino, Antonio. 2013b. The Yaγnob Valley in the Framework of the History of the Sogdian Upper Zarafšān: A Political and Economical Point of View. In: A. Panaino, A. Gariboldi, P. Ognibene (eds.). *Yaghnob Studies I. Papers from the Italian Missions in Tajikistan. Indo-Iranica et Orientalia, Vol. 12. Series Lazur*: 31–70. Milano / Udine / Paris: Mimesis.

Panaino, Antonio, Andrea Gariboldi, Paolo Ognibene (eds.). 2013. *Yaghnob Studies I. Papers from the Italian Missions in Tajikistan. Indo-Iranica et Orientalia, Vol. 12. Series Lazur*. Milano / Udine / Paris: Mimesis.

Passarelli, Marta. 2008. Il ritrovamento di incisioni rupestri nel Tang-i Dahana. In: G. P. Basello, D. Guizzo, P. Ognibene (eds.). *Sulla punta di uno spillo. Rapporto scientifico sulla Prima missione etnolinguistica nella valle della Yaghnob*: 95–110. Ravenna / Milano: Mimesis.

Perry, John R. 2006. Turkic-Iranian Contacts I. Linguistic Contacts. In: *Encyclopaedia Iranica. Online Edition*. Available online at: www.iranicaonline.org/articles/turkic-iranian-contacts-i-linguistic.

Pulleyblank, Edwin. 1991. *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin*. Vancouver: UBC Press.

Ragoza, A. N. 1980. *Sogdijskie fragmenty Tsentral'noaziatskogo sobranija Instituta Vostokovedenija*. Moskva: Nauka.

Rastorgueva, V. S., D. I. Edel'man. 2000. *Etimologicheskij slovar' iranskikh jazykov. Tom I: A-Ā*. Moskva: Vostochnaja literatura.

Rastorgueva, V. S., D. I. Edel'man. 2003. *Etimologicheskij slovar' iranskikh jazykov. Tom II: b-d*. Moskva: Vostochnaja literatura.

Rakhimi, M. V., L. V. Uspenskaya (eds.). 1954. *Tadzhiksko-russkij slovar'*. Moskva: GIz inostrannykh i natsional'nykh slovarej.

Reck, Christiane. 2005. The Tribulations of Human Existence: A Sogdian Fragment Corresponding to a Passage of Burzōy's Preface to Kalīlah Wa Dimnah?' In: A. van Tongerloo, L. Cirillo (eds.). *Il Manicheismo. Nuove prospettive della richerca. Manichaean Studies Vol. V*: 331–341. Lovanii-Neapoli: Brepols.

Reck, Christiane. 2009. The Ascension of the Light Elements and the Imprisonment of Ahriman. The Cosmological and Eschatological Part of a Sogdian 'Sammelhandschrift'. In: W. Sundermann, A. Hintze, F. de Blois (eds.). *Exegisti monumenta: Festschrift in Honour of Nicholas Sims-Williams. Iranica Vol. 17*: 369–398. Wiesbaden: Harrassowitz.

Reck, Christiane. 2016. Mitteliranische Handschriften Teil 2. Berliner Turfanfragmente buddhistischen Inhalts in soghdischer Schrift. *Verzeichnis der orientalischen Handschriften in Deutschland*: XVIII(2).

Rossi, Adriano V. 2013. Perception of Unevenness and the Lexical Family of Sogdian Pēsak ~ Pese ~ Pisē. In: S. R. Tokhtasyev, P. B. Lurje (eds.). *Commentationes Iranicae. Sbornik statej k 9-letiju V. A. Livshitsa*: 367–377. Sankt-Peterburg: Nestor — Istorija.

Sachau, Edward. 1905. Litteratur-Bruchstücke aus Chinesisch-Turkistan. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* 1905: 964–978.

Salemann, C. 1907. Manichaica II. *Bulletin de l'Académie Impériale Des Sciences de St.-Pétersbourg* 1(14): 531–558.

Schmitt, Rüdiger. 2010. Bespr. Novák 2010. *Philologia Fenno-Ugrica* 16–17: 95–100.

Schmitt, Rüdiger. 2016. *Personennamen in parthischen epigraphischen Quellen*. *Iranisches Personennamenbuch*. Wien: Österreichische Akademie der Wissenschaften.

Schwartz, Martin. 1970. On the Vocabulary of the Khwarezmian "Muqaddimatu l-Adab", as edited by J. Benzing. *Zeitschrift Der Deutschen Morgenländischen Gesellschaft* 120(2): 288–304.

Schwartz, Martin. 1982. Studies in the Texts of Sogdian Christians. Revised Version. *Typescript*.

Schwartz, Martin. 2004. More on Harkā and *harkapati. *Bulletin of the Asia Institute. New Series* 18: 143–145.

Schwartz, Martin. 2021. Avestica, Sogdo-Yagnobica, and Indo-Europæica. *Rodnoy jazyk* 1: 199–205.

Schwartz, Martin. 2024. Mnemonica Iranica. In: M. Macuch, A. Zeini (eds.). *Deciphering the Illegible. Festschrift in Honour of Dieter Weber*, *Iranica*: 423–434. Wiesbaden: Harrasowitz.

Scott, James C. 2009. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. *Yale Agrarian Studies Series*. New Haven: Yale University Press.

Sims-Wiliams, Nicholas, Bo Bi. 2022. Reading the Heavens: Three Manichaean Sogdian Manuscripts. In: A. Cantera, M. Macuch, N. Sims-Williams (eds.). *The Reward of the Righteous. Festschrift in Honour of Almut Hintze*: 481–500. Wiesbaden: Harrassowitz.

Sims-Williams, Nicholas. 1976. The Sogdian Fragments of the British Library. *Indo-Iranian Journal* XVIII: 43–82.

Sims-Williams, Nicholas. 1978. Review: The Buddhist Sogdian Texts of the British Library by D. N. MacKenzie. *Indo-Iranian Journal* (20): 256–260.

Sims-Williams, Nicholas. 1981. The Sogdian Sound-System and the Origins of the Uyghur Script. *Journal Asiatique* 269(1–2): 347–360.

Sims-Williams, Nicholas. 1982. The Double System on Nominal Inflexion in Sogdian. *Transactions of the Philological Society* 80(1): 67–76.

Sims-Williams, Nicholas. 1983. Chotano-Sogdica. *Bulletin of the School of Oriental and African Studies* XLVI(1): 40–51.

Sims-Williams, Nicholas. 1984. The Sogdian "Rhythmic Law". In: W. Skalmowski, A. van Tongerloo (eds.). *Middle Iranian Studies. Proceedings of the International Symposium organized by the Katholieke Universiteit Leuven from the 17th to the 20th of May 1982*, *Orientalia Lovaniensia analecta*: 203–215. Leuven: Peeters.

Sims-Williams, Nicholas. 1985. *The Christian Sogdian manuscript C2*. *Berliner Turfantexte*. Berlin: Akademie-Verlag.

Sims-Williams, Nicholas. 1989a. Sogdian. In: R. Schmitt (ed.). *Compendium linguarum iranicarum*: 173–192. Wiesbaden: Reichert.

Sims-Williams, Nicholas. 1989b. *Sogdian and Other Iranian Inscriptions of the Upper Indus*. *Vol. 1. Corpus Inscriptionum Iranicarum*. London: School of Oriental and African Studies.

Sims-Williams, Nicholas. 1992. *Sogdian and Other Iranian Inscriptions of the Upper Indus*. *Vol. 2. Corpus Inscriptionum Iranicarum*. London: School of Oriental and African Studies.

Sims-Williams, Nicholas. 1995. Christian Sogdian Texts from the Nachlass of Olaf Hansen, I: Fragments of the Life of Serapion. *Bulletin of the School of Oriental and African Studies* 58(1): 50–68.

Sims-Williams, Nicholas. 2002. Ancient Afghanistan and Its Invaders. In: N. Sims-Williams (ed.). *Indo-Iranian Languages and Peoples*. *Proceedings of the British Academy Vol. 116*: 225–242. Oxford University Press.

Sims-Williams, Nicholas. 2007. *Bactrian Documents from Northern Afghanistan II: Letters and Buddhist Texts*. London: Nour.

Sims-Williams, Nicholas. 2008. Sogdian-Turkish Bilingualism and Linguistic Interference in 9th – 10th Century Dunhuang. In: Z. Zarshenas (ed.). *Ǧašn-nāme-ye doktor Badr-al-zamān Qarīb = Papers in Honour of Professor B. Gharib*: 41–51. Tehran: Farhangestan.

Sims-Williams, Nicholas. 2015. *The Life of Serapion and Other Christian Sogdian Texts from the Manuscripts E25 and E26*. *Berliner Turfantexte*. Turnhout: Brepols.

Sims-Williams, Nicholas. 2016a. *A Dictionary: Christian Sogdian, Syriac and English (Beiträge zur Iranistik 41. Revised*. Wiesbaden: Reichert.

Sims-Williams, Nicholas. 2016b. "Stater" and "Drachm" in Sogdian and Bactrian Weight Inscriptions. In: *Tulufan yu sichouzhilu jingjidai gaofeng luntan ji diwujie tulufanxue guoji xueshu yantaohui lunwenji* (吐鲁番与丝绸之路经济带高峰论坛暨第五届吐鲁番学国际学术研讨会论文集): 102–108. Shanghai: Shanghai guji chubanshe.

Sims-Williams, Nicholas. 2017. *An Ascetic Miscellany: The Christian Sogdian Manuscript E28*. *Berliner Turfantexte*. Turnhout: Brepols.

Sims-Williams, Nicholas. 2019. *From Liturgy to Pharmacology: Christian Sogdian Texts from the Turfan Collection*. *Berliner Turfantexte*. Turnhout: Brepols.

Sims-Williams, Nicholas. 2020. Sogdian Biblical Manuscripts from the Turfan Oasis. In: S. N. C. Lieu, G. L. Thompson (eds.). *The Church of the East in Central Asia and China*, *China and the Mediterranean World*: 13–21. Turnhout: Brepols.

Sims-Williams, Nicholas. 2021a. *A Dictionary: Christian Sogdian, Syriac, and English. 2nd edition, revised and completed*. *Beiträge Zur Iranistik*. Wiesbaden: Reichert Verlag.

Sims-Williams, Nicholas. 2021b. The Avesta in Sogdiana. *Himalayan and Central Asian Studies* 25(1–3): 54–62.

Sims-Williams, Nicholas. 2021. Plural and Collective in the Sogdian Ancient Letters. *Transactions of the Philological Society* 119(2): 174–181. doi: 10.1111/1467-968X.12211.

Sims-Williams, Nicholas. 2022. The Rediscovery of Sogdian. In: G. B. Mikkelsen, K. Parry (eds.). *Byzantium to China: Religion, History and Culture on the Silk Roads. Studies in Honour of Samuel N.C. Lieu*: 483–497. Leiden / Boston: Brill.

Sims-Williams, Nicholas. 2023. *The "Ancient Letters" and Other Early Sogdian Documents and Inscriptions*. *Corpus Inscriptionum Iranicarum, pt. II: Inscriptions of the Seleucid and Parthian periods of Eastern Iran and Central Asia, vol. III: Sogdian V*. London: School of Oriental and African Studies.

Sims-Williams, Nicholas. 2025a. Is there a "predicative Instrumental" in Sogdian? In: N. Sims-Williams. *Chotano-Sogdica. Linguistic studies on Sogdian and Khotanese*, *Grammatica Iranica 4*: 209–220. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Sims-Williams, Nicholas. 2025b. Yaghnobi and the Sogdian "Rhythmic Law". In: N. Sims-Williams. *Chotano-Sogdica. Linguistic studies on Sogdian and Khotanese*, *Grammatica Iranica 4*: 221–228. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Sims-Williams, Nicholas, François de Blois. 2005. The Bactrian Calendar: New Material and New Suggestions. In: D. Weber (ed.). *Languages of Iran: Past and Present. Iranian Studies in memonam David Neil MacKenzie*. *Iranica Vol. 8*: 185–196. Wiesbaden: Harrassowitz.

Sims-Williams, Nicholas, Desmond Durkin-Meisterenst. 2012. *Dictionary of Manichaean Sogdian and Bactrian*. *Vol. III. Part 2. Corpus Fontium Manichaeorum. Subsidia. Dictionary of Manichaean Terms*. Turnhout: Brepols.

Sims-Williams, Nicholas, Desmond Durkin-Meisterernst. 2022. *Dictionary of Manichaean Texts (Texts in Sogdian and Bactrian). 2nd ed. Corpus fontium manichaeorum*. Turnhout: Brepols.

Sims-Williams, Nicholas, James Hamilton. 2015. *Turko-Sogdian Documents from 9th — 10th Dunhuang (Corpus Inscriptionum Iranicarum, Pt. II, Vol. III*. London: School of Oriental and African Studies.

Sims-Williams, Nicholas, Martin Schwartz, William J. Pittard. 2014. *Biblical and Other Christian Sogdian Texts from the Turfan Collection*. *Berliner Turfantexte*. Turnhout: Brepols.

Soper, John. 1992. Uzbek Influence on the Tajik Verb System. *Ural-Altaische Jahrbücher* 64: 49–58.

Stark, Sören. 2008. Approaching the periphery: highland Ustrūshana in the pre-Mongol period. In: É. Vaissière (ed.). *Islamisation d'Asie Centrale*: 215–238. Paris: Association pour l'Ávancement des Études iraniennes.

Steblin-Kamenskij, I. M. 1999. Etimologicheskij slovar' vakhanskogo jazyka. Sankt-Peterburg: Peterburgskoe vostokovedenie.

Steblin-Kamenskij, I. M. 2008. Tri zeravshanskikh slova. In: M. E. Rezvan (ed.). Rakhmat-name. Sbornik statei k 70-letiju Rakhmata Rakhimovicha Rakhimova: 338–347. Sankt-Peterburg: Kunstkamera.

Sundermann, Werner. 1980. Probleme der Interpretation manichäisch-soghdischer Briefe. *Acta Antiqua Academiae Scientiarum Hungaricae* 28: 289–316.

Sundermann, Werner. 1981. *Mitteliranische manichäische Texte kirchengeschichtlichen Inhalts*. *Vol. 11. Berliner Turfantexte*. Berlin: Akademie-Verlag.

Sundermann, Werner. 1990. *The Manichaean Hymn Cycles Huyadagmān and Angad Rōšnān in Parthian and Sogdian*. *Vol. CII, Suppl. Vol. VII. Corpus Inscriptionum Iranicarum*. London: School of Oriental and African Studies.

Sundermann, Werner. 1996. Three Fragments of Sogdian Letters and Documents. In: *La Persia e l'Asia Centrale da Alessandro al X secolo*, *Atti dei convegni lincei*: 99–111. Roma: Accademia nazionale dei lincei.

Sundermann, Werner. 1997. *Der Sermon von der Seele. Eine Lehrschrift des östlichen Manichäismus. Edtiton der parthischen und soghdischen Version*. *Berliner Turfantexte.* Turnhout: Brepols.

Sundermann, Werner. 2001. *Manichaica Iranica. Ausgewählte Schriften. Vols. I–II. Ed. by C. Reck, D. Weber, C. Leurini, A. Panaino. Serie orientale Roma*. Roma: Istituto Universitario Orientale.

Sundermann, Werner. 2002. "The Book of the Head" and "the Book of the Limbs". A Sogdian Word List. In: P. Huyse (ed.). *Iran. Questions et Connaissances. Actes du IVe congrès européen des études iraniennes, organisé par la Societas Iranologica Europaea, Paris, 6-10 septembre 1999*. *Cahiers de Studia Iranica Vol. I*: 135–161. Paris: Peeters.

Sundermann, Werner. 2007. Eine Re-Edition zweier manichäisch-soghdischer Briefe. In: M. Macuch, M. Maggi, W. Sundermann (eds.). *Iranian Languages and Texts from Iran and Turan*. *Iranica Vol. 13*: 403–422. Wiesbaden: Harrassowitz.

Sundermann, Werner. 2016. On the Growing Turkicization of the Sogdian Language. In: É. A. Csató, L. Johanson, A. Róna-Tas, B. Utas (eds.). *Turks and Iranians. Interactions in Language and History*: 29–36. Wiesbaden: Harrassowitz.

Tedesco, Paul. 1923. a-Stämme und aya-Stämme im Iranischen. *Zeitschrift für Indologie und Iranistik* 2: 281–315.

Tedesco, Paul. 1926. Ostiranische Nominalflexion. *Zeitschrift für Indologie und Iranistik* IV: 94–166.

Thierry. 1999. Sur les monnaies des Türgesh. In: M. Alram, D. Klimburg-Salter (eds.). *Coins, Art and Chronology*: 321–349. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Tremblay, Xavier. 2006. Ist die Aktivendung 3Pl -āra in einigen ostiranischen Sprachen inneriranische Entwicklung oder indogermanisches Erbe. *Münchener Studien zur Sprachwissenschaft* 62(2002): 259–287.

Vaissière, Étienne de la. 2005. *Sogdian Traders. A History*. Leiden / Boston: Brill.

Vydrin, Arseniy P. 2025. Copulas in Yaghnobi. *Indo-Iranian Languages* 1(1): 193–227. doi: 10.30842/30346800.2025.1.1.9.

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

Wendtland, Antje. 2011. *Die Entwicklung von Demonstrativpronomen zu Artikeln im Soghdischen. Göttinger Orientforschungen. III. Reihe, Iranica*. Wiesbaden: Harrassowitz Verlag.

Yoshida, Yutaka. 1984. Sogdian Miscellany. *Studia Iranica* 13(1): 145–149.

Yoshida, Yutaka. 1996. The Sogdian Dhūta Text and Its Chinese Original. *Bulletin of the Asia Institute. New Series* 10: 167–173.

Yoshida, Yutaka. 2001. On the Sogdian Fragments of the St. Petersburg Collection. *Studies in the Inner Asian Languages* 3: 105–117.

Yoshida, Yutaka. 2007. Notes on the Khotanese Secular Documents of the 8th–9th Centuries. In: M. Macuch, M. Maggi, W. Sundermann (eds.). *Iranian Languages and Texts from Iran and Turan*. *Iranica Vol. 13*: 463–472. Wiesbaden: Harrassowitz.

Yoshida, Yutaka. 2009a. Sogdian. In: G. Windfuhr (ed.). *The Iranian Languages*: 279–335. London / New York: Routledge.

Yoshida, Yutaka. 2009b. Turco-Sogdian Features. In: W. Sundermann, A. Hintze, F. de Blois (eds.). *Exegisti monumenta: Festschrift in Honour of Nicholas Sims-Williams*. *Iranica Vol. 17*: 571-585. Wiesbaden: Harrassowitz.

Yoshida, Yutaka. 2016. Sogdian Language I. Description. In: *Encyclopaedia Iranica. Online Edition*. Available online at: www.iranicaonline.org/articles/sogdian-language-01.

Yoshida, Yutaka. 2019. *Three Manichaean Sogdian Letters Unearthed in Bäzäklik, Turfan*. Kyoto: Rinsen Books Co.

Yoshida, Yutaka. 2022a. *Lectures in Sogdian Grammar* (ソグド語文法：解説と練習問題). Kyoto: Rinsen Books Co.

Yoshida, Yutaka 2022b. Sogdian Christians in China, Turfan and Sogdiana. In: G. Mikkelsen, K. Perry (eds.). *Byzantium to China: Religion, History and Culture on the Silk Roads. Studies in Honour of Samuel N.C. Lieu, Texts and Studies in Eastern Christianity. Vol.25*: 596–618. Leiden / Boston: Brill.

П. Б. Лурье. Еще раз к вопросу о взаимосвязи согдийского и ягнобского языков

В статье анализируются вопросы взаимосвязи двух восточноиранских языков, особенно близких друг другу: согдийского и ягнобского. После краткого введения с общей характеристикой мертвого согдийского языка I тысячелетия н. э. и современного языка долины Ягноба в горах центрального Таджикистана, автор рассматривает историографию исследований об их сходствах и расхождениях. Особый параграф в конце этого раздела обосновывает целесообразность двустороннего сопоставления согдийского и ягнобского. Третий раздел посвящен ключевым различиям двух языков и вероятной дате их размежевания. В качестве таковых рассмотрены согдийский ‘ритмический закон’, положение аугмента в имперфекте, окончания первого и третьего лица множественного числа, дифтонги, образованные вследствие умлаута у и развитие кластеров *δr и *θr. Мы обосновываем период между III в. до н. э. — III в. н. э. как разделение основных изоглосс. Это время после образования согдийского письма и эллинского влияния

и до «Старых писем», самой ранней крупной группы согдийских текстов (до 313 года). В следующем разделе мы исследуем общие инновации, имевшие место после размежевания праязыков. Ряд таких черт являются следствием интенсивных контактов с носителями тюркских языков. Раздел 5 посвящен некоторым культурным словам и экстралингвистическим, в первую очередь, археологическим данным. Мы стремимся доказать, что протоягнобцы принимали самое скромное участие в развитии согдийской торговли и космополитичной высокой культуры. В разделе 6 мы рассматриваем литературный согдийский как койне, которое использовалось носителями разных близкородственных диалектов, и стремимся показать, что некоторые изоглоссы, которые можно увидеть в тех или иных согдийских текстах, были основаны на диалектных особенностях их авторов, происходивших, в том числе, из восточного Согда или Уструшаны, где речь была ближе протоягнобскому, чем самаркандский стандарт. Мы приходим к выводу, что протоягнобский занимал свое (быть может, весьма особенное) место среди диалектов согдийских городов-государств. К статье прилагается список из 150 согдийско-ягнобских лексических соответствий, выявленных позже словаря в книге «Ягнобские тексты» (1957). Часть их основывается на архивных материалах Б. А. Алборова и М. Н. Боголюбова, примерно в половине случаев эти соответствия были выявлены автором. Дополнительные указатели доступны в формате онлайн-приложения к статье.

Ключевые слова: близкое родство языков; восточноиранские языки; согдийский язык; ягнобский язык; языковые изоглоссы; абсолютные даты языковых признаков.

Issue 2 / Часть 2

Articles / Статьи

George Starostin
National Research University Higher School of Economics, Moscow / Santa Fe Institute, New Mexico; gstarst1@gmail.com

Lexicostatistical Studies in Khoisan III/III: Reconstructing a Swadesh wordlist for Proto-Khoe (items 51–100)[1]

This paper is the third part of an ongoing lexicostatistical analysis of the basic lexicon for languages belonging to the Khoe family of South Africa, revised and expanded in comparison to the author's previously published attempt. Here, I focus on the second, generally less stable half of the Swadesh wordlist with the aim of investigating lexical replacement from Proto-Khoe to intermediate proto-stages and modern languages and offering an onomasiological reconstruction for at least some parts of the wordlist. Results of this analysis, combined with similar earlier treatment of the first half of the wordlist, constitute the basis for a robust phylogenetic classification, based on objective statistical analysis and manually examined shared innovations (to appear in the future as the last part in this series).

Keywords: Khoisan languages; Khoe languages; lexicostatistical analysis; onomasiological reconstruction; historical semantics.

Introductory notes

In my previous two publications on the diachronic (etymological and lexicostatistical) analysis of languages belonging to the Khoe family (Starostin 2024, 2025), I have focused on the "ultra-stable" part of the Swadesh wordlist, i.e. 50 lexical items selected for their overall higher average rates of retention (following the stratification algorithm in S. Starostin 2007, further refined in Starostin 2013). Given the relatively large number of languages and subgroups involved (as compared to the other two "Khoisan" families, Ju and Tuu), it was impossible to publish a detailed account of all the lexical evidence at once. However, since Khoe is, by all accounts, not a very deep linguistic family, and the lexicostatistical distance between its constituents (especially all the subgroups within the large Kalahari Khoe branch) is fairly small, it was also clear that a meaningful phylogenetic analysis of the family is impossible without at least extending the scope of the data to the entirety of the wordlist, which is why I have so far refrained from producing a new phylogenetic tree or engaging into lengthy discussions on the controversial and ambiguous moments in the classification.

Consequently, in this paper I continue the analysis begun in Starostin 2024 and 2025, offering an attempt of onomasiological reconstruction for the second, "less stable" part of the 100-item wordlist, further supporting it with 10 additional entries (originally selected by Sergei Yakhontov from the full 200-item Swadesh wordlist and included by Sergei Starostin in his standard computerized version of the lexicostatistical wordlist) for the sake of extra precision. Due to the fact that this part contains etyma which are commonly known to be more liable to lexical replacement around the world (as we shall see below, this holds more or less true for the Khoe family as well), the reconstruction process may be a little more complicated, yet, on the other hand, this part of the investigation actually promises more insight into the process of dispersion and diversification of the family, and more reliable results as to the internal

[1] The present work was conducted as part of the research project "From Antiquity to Modernity" (National University Higher School of Economics, 2026).

classification of Khoe on the level of small language clusters. Unfortunately, progress is still hampered by the lack of data — lacunae in the wordlists, dialectal interference, and occasional inaccuracy in phonetic transcriptions and semantic glossing can be seriously problematic, though active data collection and publication over the last decade (especially concerning the Eastern areas of Kalahari Khoe) are starting to gradually alleviate these issues.

As before, for a general overview of the surveyed languages, specific details on the reconstruction of Proto-Khoe, and the chosen methodology, the first two parts of the study (Starostin 2024, 2025) should be consulted, as well as my earlier works on the subject (e.g. Starostin 2013). For the sake of readers' convenience, I shall once again repeat the list of principal data sources and notes on transcription.

(a) **Data**. The list of principal sources remains the same as in the previous papers; below I provide a brief recapitulation, distinguishing between primary and auxiliary sources.

I. **Khoekhoe (KK)**: Nama (N) — Haacke & Eiseb 2002 (primary), Rust 1969, Haacke et al. 1997 (auxiliary); !Ora (K) — Meinhof 1930 (primary), Wuras 1920, Du Plessis 2018 (auxiliary); Xri — Haacke & Snyman 2019 (primary), Meinhof 1930 (auxiliary).

II. **Non-Khoekhoe** (**NKK**) (= Kalahari Khoe).

II.1. **Khwe**: ǁAni, ǀGanda, Buga — Vossen 1997 / Vossen et al. 1988 (primary), Vossen 1986 (auxiliary); Khwe — Kilian-Hatz 2003 (primary), Vossen 1997 (auxiliary);

II.2. Naro-ǂHaba: Naro — Visser 2001 (primary), Barnard 1985, Vossen 1997 (auxiliary); ǂHaba —Vossen et al. 1988 / Vossen 1997 (primary);

II.3. **ǀGwi-ǁGana**: ǀGwi, ǁGana —Vossen et al. 1988, Vossen 1997 (primary), Tanaka 1978, Nakagawa 1996, Nakagawa 2006 (auxiliary).

II.4. Shua: Cara, ǀXaise, Deti, Danisi — Vossen et al. 1988, Vossen 1997 (primary). *New*: I also add the recently made available dictionary of Shua (Fehn et al. 2025) as an auxiliary source (it is difficult to use it as a primary one due to the conflation of several dialects of Shua, although the authors usually take care to note specific dialectal differences — however, they do not easily correlate with the ones originally traced by Vossen).

II.5. Ts'ixa: Vossen et al. 1988, Vossen 1997 (primary), Fehn et al. 2024 (primary), Fehn 2014 (auxiliary). Compared to the previous papers, I now treat Vossen's earlier data and the recently published dictionary by Fehn et al. on equal footing; differences between the two sources are very rarely lexical and mainly include phonetic discrepancies (sometimes very important, as Fehn et al. often note Ts'ixa variants with click loss instead of or in addition to variants with the palatal click, attested by Vossen).

II.6. Tsua: Kua, Tsua — Vossen et al. 1988, Vossen 1997 (primary); Tjwao (Tshwao) — Phiri 2021 (primary); Hie (Hietshware) — Dornan 1917 (primary).

II.7. **Cua**: Welstood 2024 (primary); Collins & Wellstood 2025 (auxiliary).

(b) **Transcription**. I follow the current standards of the Global Lexicostatistical Database (GLD), which, in their turn, are largely based on standard IPA conventions. The only major changes involve consistent transcription of the palatal glide as *y* (instead of *j*) and modifications to the affricate system (IPA *ts*, *dz* = GLD *c*, *ʒ*; IPA *tʃ*, *dʒ* = GLD *č*, *ǯ*; IPA *tɕ*, *dʑ* = GLD *ɕ*, *ʑ*). In transcribing clicks, I also adopt Rainer Vossen's convention of representing clicks with voiced effluxes with an underline tilde (ǀ̲, ǂ̲, etc.), while clicks with nasal effluxes get a superscript tilde (ǀ̃, ǁ̃, etc.).

Some of the newer sources on Khoe languages use more complicated notation to transcribe both click and non-click phonemes, concentrating more on the actual phonetic values of the sounds rather than their phonological status. In order to provide more comparative transparency and avoid confusion for readers who are not so well versed in the specific intricacies of Khoe phonetics, we have simplified those transcriptions: thus, e.g., the exact same click that

is transcribed as *ǂx* (velar fricative efflux) or *ǂχ* (uvular fricative efflux) in different sources — which, quite often, does not even reflect a true dialectal/idiolectal difference in articulation, but rather a transcriptional convention — will be consistently unified to *ǂx* so as to avoid creating the illusion that we may be dealing with different phonemes. Likewise, the velar/uvular ejective affricate, alternately transcribed as *kʼ*, *kx*, *kxʼ*, *qχ*, or *qχʼ* depending on the source, shall be consistently simplified and unified to *kx*.

Khoe basic lexicon: Items 51-110

51. ALL

- KK: ***hoa** (N *hòá*, K *hoá*, Xri *hoa ~ wa*). ◊ At least in N, the word is polysemous (may mean 'all', 'both', or 'entire' depending on the grammatical number of the head noun).
- Khwe: Not reconstructible. ◊ The concept is poorly documented in published Khwe sources. The most common equivalent in ǁXom is the complex form *yó-kà-xā*, derived from the noun *yó* 'size, quantity'.
- Naro-ǂHaba: ***we** (Naro *wě*, ǂHaba *wéː*).
- |Gwi-ǁGana: ***we** (|Gwi *wèē* [Na.], *wé-kae* [Ta.], ǁGana *wé-ha* [Ta.]).
- Shua: Not reconstructible due to lack of attestation, but cf. Shua *ʔíī-yé* 'all' (Fehn et al. 2025: 4), in agreement with Ts'ixa and Tsua forms below.
- Ts'ixa: **ʔíī-yè**.
- Tsua: ***ʔi(-ye)** (Hie *i-he*, Tjwao *ʔiː-ye ~ ʔiː-ra*). ◊ Attested only in Hie and Tjwao, but external data corroborates the archaic nature of the stem.
- Cua: **yéː**.
- NKK: This is a rare case where an isogloss actually distinguishes "Western" from "Eastern" Kalahari Khoe: Naro and |Gwi-ǁGana allow to reconstruct **we*, while Shua and Tsua agree on sharing the root **ʔi-*, usually in the extended morphological form **ʔi-ye* (in all likelihood, contracted to just *yéː* in Cua). External data (Khoekhoe) speak in favor of **we*; internal data would rather suggest **ʔi*, but it is somewhat more likely to represent a secondary formation from a verbal root (cf., perhaps, Cua *ʔīː* 'to be like'?).
- CK: ***hoɛ**. ◊ The root is reconstructible on the basis of PKK **hoa* and PNKK (Western) **we*, with two caveats: (a) the more accurate PNKK reconstruction should be **wɛ* (the form is not attested in languages which consistently differentiate between **e* and **ɛ*, such as ǁAni), which would account for **oɛ* → **oa* in PKK; (b) word-initial **h-* is irregularly lost in PNKK, although this may be a specific development in front of a labial diphthong (basically a development like **hw-* → **w-* in NKK; for a similar example, cf. PKK **hui* 'to help' vs. |Gwi *wùí* id., quoted in Vossen 1997: 148). It may be tempting to further speculate on possible derivation of this stem from the verbal root **ho* 'to find, get, obtain' (N *hòò*, K *ho*, Naro *hòː*), though outside of KK it is only found in Naro; such a derivation would still already have to be projected onto the PCK level.

52. BARK (of tree)

- KK: ***soro-** (N *sòrṓ-b* 'peel, skin (of fruit), rind; *rare* bark', K *sòró-b*). ◊ Not attested in Xri. A more common equivalent for the meaning 'bark of tree' in modern N is *ǂʔàm̋-mi*, although it seems to more generally denote 'bark stripped off tree', being derived from the verb *ǂʔàm̋* 'to (de)bark'. Another close synonym is *ǁʔúú-b* 'periderm (of bark); (dry, dead) bark', also

with somewhat more narrow semantics than is required. The primary meaning of **soro-*, as is seen both from synchronic and comparative data, is rather 'shell', 'cover' in general; however, examples from the older dictionary of Krönlein & Rust 1969 show that in the 19th century at least, the word could definitely apply to tree-bark (cf. *sac go ǂʔawi sorob ãsa ǁxuːs gumo* "there is the thorn-tree whose bark you are peeling" on p. 349), so formally we are justified in reconstructing **soro-* 'shell, peel, tree-bark' on the PKK level.

- Khwe: ***ǁkxũ̋** (ǁAni, |Ganda, Buga, ǁXom *ǁkxṹ*).
- Naro-ǂHaba: ***ǁkxũ̋** (Naro *ǁkxṍː*, ǂHaba *ǁʔṹˀ*). ◊ The ǂHaba form most certainly belongs here, though simplification of the click efflux and (compensatory?) appearance of pharyngealization are quite irregular (possibly erroneous).
- |Gwi-ǁGana: ***gure** (|Gwi, ǁGana *gure* [Ta.]). ◊ Attested only in Tanaka's dictionary, but well corroborated by external data.
- Shua: ***ǁʔũ̋** (Cara *ǁʔṹ*, |Xaise *ǁʔṹ*). ◊ Note also Shua *gōré* 'bark' (Fehn et al. 2025: 20).
- Ts'ixa: **ǁʔṹ** [Vossen], **ǁkxṹ**: [Fehn].
- Tsua: ***ʒore** (Hie *ǯoreː*, Tjwao *ǯore)*. ◊ Kua has *ǁʔṹ* 'bark' in Vossen's records, but no other Tsua lect seems to preserve traces of this root.
- Cua: **ʒŭrē**.

○ NKK: ***ǁkxũ.** ◊ Of the five main branches of NKK, three have **ǁkxũ* as the default equivalent for this meaning, while two (|Gwi-ǁGana and Tsua) prefer the bisyllabic **gure*. However, the situation is not so black-and-white overall, since **gure* is also found in some Shua lects, while **ǁkxũ* is attested at least in Kua. As of now, evidence is lacking as to whether both forms could be encountered in the same lect with any semantic differentiation. However, a serious argument can be put up in favor of a secondary origin for **gure*: it functions as the main equivalent for 'bark' in !Xóõ, where, furthermore, the form is morphologically segmentable (sg. *gú-le*, pl. *gú-n*, see Traill 1994: 173). This means that **gure* in the NKK area is most likely a Taa root, diffused wide enough to replace the original CK root in some of the languages.

○ CK: ***ǁkxũ.** ◊ PNKK **ǁkxũ* is a transparent match with N *ǁʔṹú-b* '(dry/dead) bark', allowing to easily project the root onto the PCK level. As to the precise semantics, there are two possibilities: (a) reconstructing the semantic opposition **ǁkxũ* 'bark of tree' : **coro* 'cover, shell, peel' for PCK, with several semantic shifts in PKK; (b) reconstructing **ǁkxũ* 'dead bark' : **coro* 'cover, shell, peel, bark /general/', with the inverse semantic shifts in PNKK or some of its daughter branches (note that on some rare occasions, reflexes of **coro* in NKK languages also include the semantics of 'bark', though this issue demands very fine-grained investigation). Given, however, that even in N *sòrṓ-b* is rarely encountered in the meaning 'bark', solution (a) seems altogether preferable at our current state of knowledge.

53. BELLY

○ KK: ***ǃ̃a-b** (N *ǃ̃àá-b*, K *ǃ̃àː-b*). ◊ Not attested in Xri. Always with polysemy: 'belly'/'stomach'/'innards'.

- Khwe: ***ʃ̃ʰa** (ǁAni *ʃ̃nă*, |Ganda */à*, Buga */á*, ǁXom */āá*). ◊ The notation **ʃ̃ʰ-* is here rather an arbitrary designation of the rare correspondence where a nasal efflux in ǁAni matches a zero efflux in the other Khwe lects (as in, e.g., **ǁ̃ʰabo* 'shoe' → ǁAni *ǁ̃nàbó*, but |Ganda *ǁábò*, ǁXom *ǁàvòo*).
- Naro-ǂHaba: ***ʃ̃a** (Naro *ʃ̃ãː*, ǂHaba *ʃ̃nà*).
- |Gwi-ǁGana: ***ʃ̃a** (|Gwi *ʃ̃nàː*, ǁGana *ʃ̃nâ*).
- Shua: ***|a** (Cara, Danisi */ă*, |Xaise */à*, Deti */â*).
- Ts'ixa: **|â** [Vossen], **|āà** [Fehn].
- Tsua: ***|a** (Kua */à*, Tsua */â*, Tjwao */aː*).

- Cua: *|āà* ‘stomach’. ◊ No separate word for ‘belly’ is given, so it makes sense to suspect polysemy.

○ NKK: ***ʃ̃ʰa**. ◊ All the forms are clearly related; whether the fluctuation between nasal and non-nasal efflux reflects an original complex click efflux in PNKK or is due to other (unknown) factors is really a subject for a separate study.

○ CK: Not properly reconstructible. ◊ Despite some phonetic similarity, PKK **l̃a-* and PNKK **ʃ̃ʰa-* (whatever its efflux might have actually been) cannot be expected to correspond to each other. The closest possible parallel to the KK root I can find is perhaps Naro *l̃áè* ‘middle; waist’ (with a different suffix?), but without any further NKK comparanda.

54. BIG

○ KK: ***kai** (N *kấí*, K *kái ~ gái*, Xri *kai*).

- Khwe: ***|áú** (‖Ani */áú*, |Ganda, ‖Xom */éú*).
- Naro-ǂHaba: Unclear. ◊ Naro *kāī-ā* ‘big’ is somewhat suspicious due to the rarity of this root in other NKK languages; a relatively recent borrowing from KK cannot be excluded. The ǂHaba equivalent is *‖ʔ̂û* (Vossen 2022: 120; transcription *‖ʔû* in Vossen 1988: 72 is likely erroneous in light of the later entry as well as external data), with further parallels in the Eastern branches, so it might actually be more archaic. (On the other hand, see also notes on MANY).
- |Gwi-‖Gana: Unclear. ◊ Vossen (1988: 72) contrasts |Gwi *ǃúrí* with ‖Gana *‖́ó*; however, Tanaka’s dictionary, despite the usual orthographic problems, shows that both roots can be found in both of these languages (|Gwi *ǃuri*, *ǂdó*; ‖Gana *guri*, *ǂdó*), without any clear semantic differentiation. In light of external data, **‖́o* seems to be a likely candidate for proto-status.
- Shua: Not reconstructible. ◊ Vossen’s data are highly contradictory here, with each lect reflecting a different root: Cara *gùrí*, |Xaise *‖́ô*, Danisi */áú*, Deti *kái* (Fehn’s data support many of these synonyms but do not allow to construct a clear historical picture).
- Ts’ixa: **|áú** [Vossen, Fehn].
- Tsua: Unclear. ◊ Kua *‖ʔ̂û*, Tsua *‖ʔ̂û* (Vossen’s data) are opposed to Hie *‖o: ~ ‖xo:*, Tjwao *‖́o:*; despite phonetic similarity, these two groups of forms are not really compatible (reflecting **‖ʔ̂ũ* and probably **‖́o* respectively).
- Cua: **‖qʔ̂ũ̀**.

○ NKK: Not properly reconstructible. ◊ The concept ‘big’ seems to be quite unstable in NKK languages. At least four different roots display an almost random distribution here: (a) **|au* (all of Khwe + Danisi and Ts’ixa); (b) **‖(q)ũ* (ǂHaba + Kua-Tsua and Cua); (c) **ǃ̃uri* (|Gwi-‖Gana + Cara); (d) **‖́o* (|Gwi-‖Gana + |Xaise + Hie-Tjwao), not to mention **kai* (in Naro and Deti, perhaps diffused from a KK source). Of these, (a) does look like a Khwe innovation, which could then have areally penetrated into Danisi and Ts’ixa; for a possible external correlate, cf. perhaps N *ǀ̃àú-* ‘manner, way, style, mode; amount’. The other roots are all encountered both in the Western and Eastern parts of the NKK continuum, and without precise semantic descriptions it is impossible to differentiate between the three.

○ CK: Not reconstructible. ◊ PKK **kai*, reasonably stable in this small branch, is sharply contrasted here with a variety of roots in NKK languages; selecting a single optimal candidate is impossible at our current level of knowledge.

55. BITE

○ KK: ***ba** (K *bá*, Xri *baa*). ◊ Although K and Xri agree in this case, and the antiquity of the root is further confirmed by external data, N has no traces of it; instead, the equivalent is *nấấ́ ~ nấá ~*

(Hai‖om) *ndáá*, without any known Khoe parallels and with phonetic features hinting at a non-Khoe origin (word-initial *n-*, including an even less regular reflex *nd-* in the Hai‖om dialect).

- Khwe: ***pá** (‖Ani *pá*, |Ganda, Buga, ‖Xom *pá*).
- Naro-ǂHaba: ***kǎˢ** (Naro *kǎˢ*, ǂHaba *kǎˢ*).
- |Gwi-‖Gana: ***pá** (|Gwi *pá:*, ‖Gana *pá*).
- Shua: ***pá** (Cara, |Xaise, Danisi, Deti *pá*).
- Ts'ixa: **pá** [Vossen], **pá**: [Fehn].
- Tsua: ***pá** (Kua, Tsua *pá*, Hie *pʰa*, Tjwao *pa:*).
- Cua: **pá**:.

○ NKK: ***pa**. ◊ This is a stable root, only replaced in Proto-Naro-ǂHaba with the innovation **kǎˢ*; given the presence of pharyngealized articulation, this can be compared with the frequently attested root **qa* 'to open /mouth/' (→ ‖Xom *qá:*, Ts'ixa *qáā*, Shua *káà*, Cua *qáà*, etc.) with a trivial semantic shift.

○ CK: ***pa**. ◊ The isogloss between PNKK and PKK is quite transparent and shows that the word belongs to the very small group of PCK items with word-initial **p-* (the only other item in this group is PCK **pi-* 'milk'). Although the statistical rarity is suspicious, there are no formal grounds on which the item could be refused PCK status.

56. BREAST (= CHEST)

○ KK: ***‖xai-b** (N *‖xáì-b*, Xri *‖xaip*). ◊ This term denotes '(male) chest' in N and probably in Xri as well (glossed as 'chest'); it is lexically opposed to N *sáḿ-s* '(female) breast' = K *samm-a* (only in Maingard's data), Xri *samm-i* id. For K, both Meinhof and Wuras quote the compound form *tʰam-kxari-b* (Mh.) = *tʰam-kxarri-b* (Wu.), glossed as 'Brust' and 'chest' respectively. Its first component can be identified as *tʰam* 'soft' (which would rather imply 'female breast' as the original semantics?), but the second component is unclear. Maingard's texts consistently indicate the use of *samma* to denote 'female breast', but contain no examples for 'male chest'. Still, at least attestation in Xri confirms that ***‖xai-** can be formally traced back to PKK in the required meaning (and the compound form in K is very likely to be a recent innovation).

- Khwe: ***‖̰u** (‖Ani, ‖Xom *‖̰ǔ*).
- Naro-ǂHaba: ***‖̰u** (Naro *‖̰ùū*, ǂHaba *‖̰úù*).
- |Gwi-‖Gana: ***‖̰u** (‖Gana *‖̰û*). ◊ For |Gwi, cf. the form *g̃!u-ša* in Tanaka 1978: 21.
- Shua: ***‖̰u** (Cara |Xaise, Danisi, Deti *‖̰ǔ*).
- Ts'ixa: **‖̰ūú** (Fehn), **‖̰ǔ** (Vossen).
- Tsua: ***‖̰u** (Kua *‖̰ǔ*, Tsua *‖ǔ*, Hie *|ʰu:*).
- Cua: **‖̰ǔū**.

○ NKK: ***‖̰u**. ◊ A very stable root for 'male chest' throughout NKK (more stable, actually, than the equivalents for 'female breast', which include reflexes for PNKK **pi*, more common in the Western area, and PNKK **sam*, more common for Eastern languages).

○ CK: Not properly reconstructible. ◊ The only known NKK parallel to the KK root is Naro *‖xáì* 'breastbone' (suspicious since it could actually be a KK borrowing). On the other hand, a curious parallel in KK to PNKK **‖̰u* may be N *‖̰ū-s* 'sternum of ostrich' (according to Haacke & Eiseb, only found in an old source, so also suspicious). No way to properly resolve the binary split.

57. BURN (tr.)

○ KK: ***dao** (N *tàò*, K *dao*, Xri *dãu*). ◊ In N, this is primarily a verb denoting transitive action (= 'to burn smth.'); for K and Xri, the situation is more blurry, but all three languages also

agree on sharing the common intransitive verb *ǂʰ*ubi* ‘to burn’ (N *ǂʰùwí ~ ǂʰùbí*, K *ǂʰūbī*, Xri ǂʰubi), so it can be safely assumed this basic opposition was inherited from PKK. The nasalized variant in Xri finds confirmation in K *dãu-sa* ‘burning /of sun/’ (Mh.), but, as in many similar cases, reasons for additional nasalization remain unclear (external comparanda only show oral vowels).

- Khwe: ***dao** (‖Ani, Buga *dáò*, |Ganda *dàò*, ‖Xom *dàō*).
- Naro-ǂHaba: ***dao** (Naro *dàò*, ǂHaba *díò*).
- |Gwi-‖Gana: ***dao** (|Gwi *díò*, ‖Gana *dáó*).
- Shua: ***dao** (Cara, Danisi, Deti *dàò*, |Xaise *dáó*, Deti *dàò*).
- Ts’ixa: **dáò ~ dāò** [Fehn], **dáó** [Vossen].
- Tsua: ***dao** (Kua *dáó*, Tsua *dáò*, Hie *dʰau*, Tjwao *dao*/*-xo*/).
- Cua: **dàō**.

○ NKK: ***dao**. ◊ A highly stable verbal root, usually employed both in the transitive and intransitive functions throughout NKK (additional partial synonyms are scattered across languages, but hardly ever projected even to the intermediate proto-levels).

○ CK: ***dao**. ◊ Except for a stricter distinction between transitive and intransitive equivalents for this meaning, KK and NKK generally agree on the transitive equivalent.

58. CLOUD

○ KK: (?) ***ʃ̃anu-s** (N *ʃ̃ànű-s*, Xri *ʃ̃anu-s*). ◊ A clear isogloss between N and Xri, yet its PKK status is somewhat dubious due to the overall instability of the concept in KK (and the Khoisan area in general). For N, the meaning is glossed as ‘cloud, raincloud, nimbus’, with the masc. counterpart *ʃ̃ànű-b* meaning ‘rain; thunderstorm’. Additionally, the word *!ʔàǎ̋-s ~ !ʔàǎ̋-b* is glossed as ‘cloud’ for the Damara dialect and as ‘white cloud’ for the Topnaar dialect (Haacke & Eiseb 2002: 298), while the word *‖ʰàő-s* is glossed as ‘cloud’ for the Hai‖om and ǂĀkhoe dialects (Haacke & Eiseb 2002: 264). In older sources, *ʃ̃anu-s* is glossed as ‘cloud’ (Krönlein & Rust 1969: 272), while the word *‖ʰao* is only attested in its verbal meaning ‘to be covered with clouds (of sky)’ (Krönlein & Rust 1969: 162) and the equivalent of *!ʔàǎ̋-s* is not attested at all.

In K, the only known equivalent is */ʰumm-a* ‘cloud, sky’ = N */ʰòmm-i* ‘heaven, (sky)’ = Xri */ʰommi* ‘cloud’, going back to PKK *//ʰom-*; however, since out of these three languages only the N data are consistently reliable in the semantic aspect, it is probably more cautious to reconstruct the original meaning of the word as ‘sky’ rather than ‘cloud’.

- Khwe: Unclear. ◊ Cf. Buga *qóm̀*, |Ganda *qóm̀* (Vossen 1992: 385). These forms would seem to be comparable to ‖Xom *‖òm̄* ‘cloud’, but click loss in this case would be completely irregular.
- Naro-ǂHaba: ***tu-|ʔũ̀** (Naro *tú:-|ʔõ̀:*, ǂHaba *tú:-m-|ʔű̀ǜ̀*). ◊ This compound literally means ‘hair of rain’ (similar idiomatization is also encountered in various Ju dialects). Additionally, cf. Naro *‖̃òm* ‘cloudy’, *‖̃ōm-ā* ‘cloud’ (but without any known parallels in ǂHaba).
- |Gwi-‖Gana: Not attested.
- Shua: Not attested in Vossen’s data, but cf. Shua *‖ūm̄ ~ ‖̃ūm̄* ‘sky; cloud’ in Fehn et al. 2025: 22.
- Ts’ixa: **‖ūm̄** (Fehn).
- Tsua: ***‖um** (Hie. *ǂom*, Tjwao *‖um*). ◊ Not attested in Kua or Tsua; the Hie-Tjwao comparison is acceptable, assuming that Dornan is mistranscribing the lateral click.
- Cua: **‖qāì**.

○ NKK: ***‖um** (?). ◊ Although this reconstruction can be seen as reliable based on joint evidence from ‖Xom, Shua, Ts’ixa, and Tjwao, what makes it problematic are the divergent forms in Naro (with an unexpected nasal efflux) and Buga/|Ganda, with a uvular stop (*q-*) in

place of the lateral click. This can mean two things: either all these forms are tied together with a rare series of correspondences reflecting an equally rare click efflux (e.g. prenasalized uvular or something of the sort), or at least some of them are not genetically related, but areally diffused (e.g. different scenarios of borrowing from a substrate source). Still, formally the connection between at least the above-mentioned four languages can be counted as genetically valid.

○ CK: Not reconstructible. ◊ The concept of 'cloud' is generally unstable in the entire Khoisan area, so, not surprisingly, coming up with a credible PCK reconstruction is hardly possible at the present levels of knowledge.

59. COLD

○ KK: ***!xai** (N *!xàǐ*, K *!xai*, Xri *!xai*).

• Khwe: ***|xonu** (‖Ani */xónó*, |Ganda */xònú*, ‖Xom */xùnú* 'be cold; be okay, feel good'). ◊ For ‖Xom, this word seems rather to convey the semantics of 'cool' (with positive connotations); the actual meaning 'cold' is rather correlated with */ǂxává*, an isolated item that nevertheless correlates well with N */ɔàwǎ-s* 'North'.

• Naro-ǂHaba: ***!q^{h}ai** (Naro *!xàī*, ǂHaba *k^{h}áì ~ káì^{ʕ}*). ◊ Click loss in ǂHaba is regular; strange discrepancies in transcription between Vossen 1988 and Vossen 2022 (aspirated stop in the former, unaspirated stop + pharyngealization in the latter) may actually tie in with |Gwi-‖Gana data as being indicative of a rare aspirated uvular click efflux.

• |Gwi-‖Gana: ***!q^{h}ai** (|Gwi *!q^{h}āī* [Na.], *!^{h}àì* [Vossen]; ‖Gana *q^{h}āī* [Na.], *qáì* [Vossen]). ◊ Nakagawa's data clearly show that we are dealing with a uvular aspirated efflux (which Vossen seems to have struggled to transcribe).

• Shua: ***xai** (Cara *xáí*). ◊ In Vossen's data, this root is only found in Cara; it is also, however, confirmed as Shua *xāí ~ xáí* 'cold; to freeze; to shiver' in Fehn et al. 2025. |Xaise and Deti have *kárá* 'cold' with no external parallels. Danisi has */xónó* 'cold' = Shua */xun* 'to shiver' (Fehn).

• Ts'ixa: **|xúnú** (Fehn), **|xúnù** (Vossen). ◊ Another Ts'ixa equivalent is */ǂxábá* 'be cold, freeze' (Fehn), which, judging by this root's distribution, is almost certainly a loanword from Khwe.

• Tsua: Unclear. ◊ According to Vossen, the Kua equivalent is */xónù*, while Tsua has *hùrú* = Tjwao */xuru ~ /^{h}uru* with sporadic click loss. On the other hand, Hie. *hai:* 'cold' = Shua **xai*, etc.

• Cua: **k^{h}àī**.

○ NKK: ***!q^{h}ai**. ◊ The most widespread, phonetically and semantically certain equivalent for the required concept throughout NKK is this stem, although its original phonetic shape has probably only been preserved in (Nakagawa's) |Gwi. A weaker second candidate would be PNKK **|xonu*, displaying the required semantics in Khwe, Kua, Ts'ixa, and Danisi (because of this distribution, ruling it out as just another Khwe loan in Ts'ixa would be premature); this may, however, be a secondary semantic development from 'to shiver' (as a possible etymological cognate, cf. PKK **|xon* 'to itch' → N */xǒń*, K */xòn-/xón*, which would suggest the physical movement semantics as primary), and its overall distribution is narrower than that of **!q^{h}ai*.

○ CK: ***!q^{h}ai**. ◊ Although the click efflux in KK is predictably simplified, the sole PKK and the preferable PNKK candidates for this term present a solid match.

60. COME

○ KK: ***ha** (N *hàà*, K *há:*, Xri *ha:*). ◊ N uses the verb */xìǐ* as a full synonym for *hàà*, and its existence in K is also confirmed as */xi:* 'to come, arrive' by Maingard. However, it is absent from all main sources on K and Xri, and its archaic nature is not corroborated by external data.

- Khwe: ***ha** (‖Ani *hă*, Buga *hâ*, ‖Xom *yāá*). ◊ |Ganda *ǂàḿ* 'to come' = ‖Xom *ǂāḿ* 'to run, run into smth.', i.e. a transparent innovation (provided the meaning is accurate in the first place).
- Naro-ǂHaba: ***ha** (Naro *hà:*, ǂHaba *hă*).
- |Gwi-‖Gana: ***hâ** (|Gwi, ‖Gana *hâ*).
- Shua: ***ha** (Cara, |Xaise, Danisi *hă*, Deti *hâ*).
- Ts'ixa: **hāà** (Fehn), **hâ** (Vossen). ◊ Aligns with Shua rather than ‖Xom (although the development **ha* → ‖Xom *ya* is recent and not very diagnostic).
- Tsua: ***ha** (Kua, Tsua *yâ*, Hie. *ya*, Tjwao *ha: ~ a:*). ◊ Tjwao preserves the original laryngeal articulation, as corroborated by external data.
- Cua: **āà**.

○ NKK: ***ha**. ◊ Superstable root, (possibly) lost only in |Ganda.

○ CK: ***ha**. ◊ A clear match between PKK and PNKK.

61. EARTH (= SOIL)

○ KK: ***!ʰu-** (N *!ʰùú-*, K *!ʰu:-b*, Xri *!ʰu:-p*). ◊ Usually attested with polysemy 'earth/land/soil/country' (at least in N; K also shows signs of polysemy).

- Khwe: ***xóḿ** (‖Ani, |Ganda, Buga, ‖Xom *xóḿ*).
- Naro-ǂHaba: ***xóḿ** (Naro, ǂHaba *xóḿ*).
- |Gwi-‖Gana: ***xóḿ** (|Gwi *xóḿ*, ‖Gana *xóám*).
- Shua: ***xóḿ** (Cara, |Xaise, Danisi *xóḿ*, Deti *xòḿ*).
- Ts'ixa: **xúḿ** (Fehn), **xóḿ** (Vossen). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***xóḿ** (Kua, Tsua **xóḿ**).
- Cua: **xòm̄**.

○ NKK: ***xom**. ◊ In most NKK languages, there is no lexical differentiation between 'earth' and 'sand' (see SAND for further details).

○ CK: Unclear. ◊ KK is clearly opposed to NKK in its lexical differentiation between **!ʰu-* 'earth, soil, land' and **‖xae* 'sand', whereas NKK merges both meanings in the single root **xom* — very likely related to the geological realities of Khoekhoe vs. Kalahari Khoe-speaking peoples. Unfortunately, linguistic evidence does not allow to decisively establish which situation is primary, because neither of the KK roots finds any etymological correlates in KK. There is, however, a hint that the nominal stem **xom* may have originally been present in KK as well, due to the preservation of the potentially related verb N *xòḿ* 'to wipe off/up (liquid)' (Haacke & Eiseb) = *xom* 'abstauben, abbürsten' (Rust), i.e. 'to dust'. But even if the two words are etymologically related, this does not disprove the possibility of an original opposition like **!ʰu-* 'soil' vs. **xom* 'dust/fine sand', so the issue remains unresolved.

62. FAT (n.)

○ KK: ***kau-** (N *kắú-b ~ kắú-s*, K *kou-b*). ◊ This is the noun denoting 'solid fat', as opposed to PKK **‖ui-b* 'liquid fat; ointment'; the root also functions as the verbal/adjectival stem **kau* '(to be or become) fat' (this is the only meaning in which it is attested in Xri, so we are not completely sure of the situation with denoting different types of 'fat' in that language).

- Khwe: ***‖̃úì** (‖Ani *‖̃núì*, |Ganda, Buga, ‖Xom *‖̃úì*).
- Naro-ǂHaba: ***‖̃úì** (Naro *‖̃úì*, ǂHaba *‖̃núì*).
- |Gwi-‖Gana: ***‖̃úì** (|Gwi, ‖Gana *‖̃úì*).
- Shua: ***‖̃úì** (Cara, Danisi, Deti *‖̃úì*, |Xaise *‖̃núì*).
- Ts'ixa: **‖̰núì** (Fehn), **‖̃úì** (Vossen). ◊ Aligns with both Shua and ‖Xom.

- Tsua: ***ǁ̃ui** (Kua *ǁúì*, Tsua *ǁùí*, Hie *ǂ̰wi*, Tjwao *ǁ̰nui*).
- Cua: **ǁ̰úì**.
- ○ NKK: ***ǁ̰núì.** ◊ No NKK language has been shown so far to distinguish (at least, on the level of different lexical roots) between the sub-meanings of 'liquid fat, oil' vs. 'hard/congealed fat'.
- ○ CK: ***ǁ̰núì.** ◊ KK **kau-* can be seen as a secondary nominal derivative of the verb **kau-* 'to be fat' which, in turn, is comparable to PNKK **cau* 'to be fat' (ǁAni, Naro, Cua etc. *cáú*) with a rare, but recurrent correspondence pattern (cf. PKK **k*^{h}*ao* 'to dig' : PNKK **c*^{h}*ao* id.); since none of the NKK languages derive the noun 'fat' from this root, this speaks in favor of an innovation in PKK.

63. FEATHER

- ○ KK: ***!ʔam-i** (N *!ʔàm̃-mi*, K*!ʔamm-a*). ◊ Both N and K systematically distinguish between **|ʔũ-* 'human hair (on head or body)' and this term for 'feather'; Snyman's data on Xri, however, yield the same equivalent (Xri *|ʔũ-b*) for both concepts. Since there is no easy way to explain the N-K isogloss as an innovation, while the semantic merger of 'hair' and 'feather' in one lexeme is typologically common, it is preferable to treat the N-K situation as more archaic (in the context of PKK proper).
- Khwe: Unclear. ◊ Attested only in ǁXom: *|ʔũ̋ũ̄*, the same word as HAIR q.v.
- Naro-ǂHaba: Unclear. ◊ Attested only in Naro: *|ʔō̄ò̄*, the same word as HAIR q.v.
- |Gwi-ǁGana: Not attested.
- Shua: Not attested (in Fehn et al. 2025, same word as HAIR q.v.).
- Ts'ixa: Unclear. Fehn lists *ǁābò* with polysemy 'feather/wing', but this word typically means 'wing' in NKK languages; additional checkups are necessary to understand if there really has been a subtle semantic shift in Ts'ixa compared to the rest of NKK or not.
- Tsua: Unclear. ◊ Not attested in Kua or Tsua; Hie *ǂ̰o* and Tjwao *|ʔũ:* are the same word as HAIR q.v.
- Cua: **|ʔṵ̃**: (same word as HAIR q.v.).
- ○ NKK: ***|ʔṵ̃**. ◊ Although the concept is poorly attested across the group, most attested examples in different subbranches agree that there was probably no lexical distinction between the concepts HAIR and FEATHER in PNKK.
- ○ CK: Unclear. ◊ While the basic term for HAIR remains the same across both of the major subdivisions of Khoe, a separate term for FEATHER (**!ʔam-*) is exclusive to Khoekhoe, and has no clear parallels in the Kalahari languages. Whether this situation is archaic or innovative cannot be easily determined.

64. FISH

- ○ KK: ***ǁʔau-b** (Nama *ǁʔàù-b*, !Ora *ǁʔau-b*, Xri *ǁʔau-p*).
- Khwe: ***ǁʔau** (ǁAni, |Ganda, Buga *ǁʔáù*, ǁXom *ǁʔéú*).
- Naro-ǂHaba: ***ǁʔau** (Naro *ǁʔāù*, ǂHaba *ǁʔáù*).
- |Gwi-ǁGana: ***ǁʔáù** (|Gwi, ǁGana *ǁʔáù*).
- Shua: ***ʔau** (Cara, |Xaise, Deti *ʔáú*, Danisi *ʔáù*).
- Ts'ixa: **ǁʔáū** (Fehn), **ǁʔáù** (Vossen). ◊ Aligns with both Shua and ǁXom.
- Tsua: (?) ***ʔau** (Tjwao *ʔau*). ◊ Not attested in Kua or Tsua. Hie. *t*^{h}*api* 'fish' is a transparent borrowing from Setswana.
- Cua: Not attested.

- NKK: ***‖ʔau**. ◊ A surprisingly stable term throughout Kalahari Khoe (despite the well-known difficulties of eliciting terms for 'fish' in the San area). Note the irregular (so far unexplained) lateral click loss throughout the Eastern area (but not in Ts'ixa, where preservation of the click may be tentatively attributed to Khwe influence).
- CK: ***‖ʔau**. ◊ The term is safely reconstructible based on evidence from both major branches.

65. FLY (V.)

- KK: Not properly reconstructible. ◊ Only N has a special word for the concept of 'flying' (*‖xànà*), although even that one is occasionally replaced in dialects with *kʰãĩ̀* 'to stand up, rise, fly up'. K and Xri nominally agree on *doe* 'to fly', but even in K the semantics fluctuates between 'fly' and 'fly away, flee' = N *dʊ̀è* 'to migrate, move away', and in NKK the respective cognates for this item almost always have the same meanings as in N; therefore, even if the semantics of 'flying' (no specific direction) is indeed present in K and Xri *doe* (and is not simply the result of inadequate glossing), it is most likely innovative. On the other hand, N *‖xànà* is not easily explainable as a recent innovation and may represent a PKK retention in this language.
- Khwe: Unclear. ◊ The only attested form is ǁXom *‖̀èvùū* 'to fly (of birds)', very likely related to the general CK term for 'wing' (**‖abo*), although the voiced articulation of the click is unexpected.
- Naro-ǂHaba: ***ʒãĩ̃ˢ** (Naro *cãĩ̀ˢ* [Visser], *cãĩ̂ˢ* [Vossen], ǂHaba *ʒãĩ̂ˢ*). ◊ Although fromally reconstructible on this low level, this word has no obvious parallels in the rest of NKK, but can be identified as a borrowing from Taa, cf. !Xóõ *ʒãĩ̃ˢ* 'to fly'.
- |Gwi-ǁGana: Not attested.
- Shua: Not attested. Fehn et al. (2025) offer several dialectal synonyms: *péē* (← PNKK **pe* 'to jump'), *tāñ* (← PNKK **tan* 'to stand up, rise'), *‖ābū* (again, connected to *‖ābò* 'wing').
- Ts'ixa: **‖ābū** (Fehn). ◊ Aligns with both Shua and ǁXom.
- Tsua: Not attested, except for several competing Tjwao synonyms (*‖abu* as in Shua; *tan* ← 'to rise'; *ʒaro*, perhaps = PNKK **ʒaro* 'to gather') and Hie. *fofa*, borrowed from Setswana.
- Cua: Not attested.
- NKK: (?) ***‖abu**. ◊ After excluding several roots with originally distinct semantics ('to jump', 'to rise', etc.) and borrowings, the only remaining candidate is **‖abu*, found in the Eastern area as well as ǁXom (albeit with phonetic deviations) — however, it seems to be morphologically related to the nominal stem 'wing' and is probably derived from it rather than vice versa.
- CK: Not reconstructible. ◊ The concept of 'flying' is generally unstable in the Khoisan area, as the meaning tends to colexify with a variety of other concepts, and CK is no exception. The only etymon represented by its own root is N *‖xànà*, but the complete lack of comparanda in other languages makes it ineligible for PCK status at our present level of knowledge.

66. FULL

- KK: ***|kxoa-sa** (N */ʔòà-sấ*, K */kxóá*, Xri */kxoa-sa ~ /ʔoa-sa*). ◊ Adjectival derivate from the verbal root **/kxoa* 'to be full / to fill'.
- Khwe: ***|kxóè** (‖Ani, |Ganda, Buga, ǁXom */kxóè*).
- Naro-ǂHaba: ***|kxoe** (Naro */kxōè*, ǂHaba */kxóè*).
- |Gwi-ǁGana: ***|kxoɛ** (|Gwi */kxóé*, ǁGana */kxòè*).
- Shua: ***|kxoe** (Cara, Deti */ʔóé*, |Xaise */ʔóè*, Danisi */kxóé*).
- Ts'ixa: **|kxóè** (Fehn), **|ʔóè** (Vossen). ◊ Aligns with both Shua and ǁXom.

- Tsua: ***|kxoe** (Kua */ʔóè*, Tsua */ʔóé*, Hie *ǃwea*, Tjwao */ʔoe-ha*).
- Cua: **|kxóè**.
- ○ NKK: ***|kxoɛ**. ◊ This root is extremely stable throughout the entire group.
- ○ CK: ***|kxoɛ**. ◊ The dipthong **-oɛ* is explicitly preserved in some of the NKK languages and undergoes regular development → **-oa* in PKK.

67. GIVE

- ○ KK: ***ma** (N *màà*, K *mà ~ mā́*). ◊ For Xri, Snyman's data only quote *au* in the meaning 'give' = N *áú* 'to share in; to let smbd. have something (food, etc.)', K *au* 'to give up (abgeben)', i.e. a word with a very close, but narrower meaning. Whether this is a genuine replacement or a deficiency of the wordlist can only be guessed.
- Khwe: ***ǂxa** (‖Ani *ǂxá*, |Ganda, Buga, ‖Xom *ǂxǎ*). ◊ For ‖Xom, cf. also *mā́:* 'to distribute, give (out) as gift', as well as *xàròó* 'to give, give away, hand in'.
- Naro-ǂHaba: ***mâ** (Naro, ǂHaba *mâ*).
- |Gwi-‖Gana: ***mã̂** (|Gwi *mã̂:* [Na.], *mâ* [Vossen], ‖Gana *mã̂*). ◊
- Shua: Unclear. ◊ Vossen does not record any Shua equivalents; Fehn et al. (2025) have *kʰáà* and *xáró* as synonyms.
- Ts'ixa: **kʰāà** (Fehn). ◊ Aligns with data from Shua, but not from ‖Xom.
- Tsua: (?) ***ma**. ◊ Attested as Tjwao *ma:* and Hie *ma:* 'to lend, to give to, to forgive'; the more basic Hie equivalent for 'to give' seems to be *tʰeː*.
- Cua: **tʰèē** and **mā́ǎ̀** are given as synonyms.
- ○ NKK: ***ma** ~ ***mã**. ◊ This root seems to still function as the default equivalent for the Swadesh meaning only in Naro-ǂHaba, |Gwi-‖Gana, and (importantly) in Tjwao; some other languages still preserve it in slightly shifted meanings.
- ○ CK: ***ma** ~ ***mã**. ◊ A solid isogloss between KK and NKK. However, there is a clear tendency in different branches to replace the original root with more specialized terms from the same semantic field. In KK, one such term is **au* 'to share, partake' (not found in NKK except for Naro *āū* 'give' /polite/, possibly a loan from KK); in the Shua cluster, it is **kʰa* = Naro *kʰáà* 'to give generously', N *kʰàà* 'to give' (especially as a gift; said to be "obsolete"). Khwe **ǂxa* and Tsua/Cua **tʰe* have no transparent etymologies for now, but may be speculated to represent recent replacements of a similar nature.

68. GOOD

- ○ KK: ***!ãĩ** (N *ǃẫĩ̀*, K *!ãĩ*, Xri *!ãĩ ~ ǃãĩ*). ◊ In KK, this root mainly conveys the semantic of moral or intrinsic goodness/quality, as opposed to KK **ʔĩ* 'good' = 'handsome/beautiful' (although K *ĩ-sa* is also occasionally rendered as 'good' in European translations).
- Khwe: ***t'on[i]** (‖Ani, Buga *t'ûĩ̀*, |Ganda *t'óǹ*, ‖Xom *t'óǹ ~ t'úǹ ~ t'úíǹ*). ◊ Phonetic variation here is probably best explained through two morphological variants (one with the verbal suffix **-i* and one reflecting the pure root). In ‖Xom, this root seems to merge both the sub-meaning referring to "moral goodness" and to "physical beauty".
- Naro-ǂHaba: ***!ãĩ** (Naro *!ã́ɛ̃̀*, ǂHaba *kʸî*). ◊ Click loss in ǂHaba is regular; monophthongization may somehow be tied to the palatalization of initial velar (**!ãĩ → *kãĩ → *kʸãĩ → kʸĩ*), although it is not possible to verify this on additional examples for now. In any case, both items can be traced back to a single proto-source with sufficient confidence.
- |Gwi-‖Gana: Unclear. ◊ Sources here are conflicting and not sufficiently accurate. For |Gwi, Vossen has *c'ɛ̂* 'good' and *qádè* 'good/pleasant/comfortable/sweet'; Nakagawa has *ɕ'ɔ̃́ɛ̃̀* 'beau-

tiful' = *ekyuen* 'beautiful/good' in Tanaka's dictionary. For ǁGana, Vossen has *kâî* 'good' and *qádè* with the same meanings as in |Gwi, as well as *t'ûî* 'good, beautiful' = *ɕ'ôè* 'beautiful' (Nakagawa) = *ekyuen* 'beautiful/good' (Tanaka).

- Shua: [a] ***t'ũĩ** (Cara, Deti *t'ű́í*, |Xaise, Danisi *t'ű̂ĩ*); [b] ***kãĩ** (Cara *kã́í*, |Xaise, Deti, Danisi *kâĩ*). ◊ The semantic difference between these two roots cannot be established based on extant data.
- Ts'ixa: [a] **ťű̀ì**; [b] **kã́ì** (Fehn, Vossen). ◊ Meanings of these two items overlap, but item [a] seems to primarily refer to physical beauty, while item [b] is also attested in the verbal meaning 'to enjoy (smth.)' and more commonly refers to abstract qualities.
- Tsua: [a] ***t'ũĩ** (Kua *t'ű̂ì*, Tsua *t'ű̂ì*, Tjwao *t'ũĩ ~ t'u(i)n*); [b] ***kãĩ** (Kua, Tsua *kâĩ*, Hie *kai-e* 'to be nice, pretty, agreeable'). ◊ Exact semantic difference between these roots is difficult to figure out based on available sources. It should be noted that a huge variety of partial synonyms is also attested for Hie (*čo* 'good, goodness'; *hwi* 'good, well', *!xo:* 'good', *tweŋe* 'fine, pretty'), complicating the situation even further.
- Cua: [a] **ťű̂ì** 'beautiful'**;** [b] **kã́ì** 'delicious, good, well'.

○ NKK: ***!ãĩ**. ◊ Despite significant technical difficulties such as an abundance of partial synonyms and inaccuracy of semantic glossing, many, if not most, of the NKK branches point to a fundamental lexical dichotomy between 'good = of suitable quality' and 'good = physically beautiful', just as in KK. The required Swadesh meaning is more typically expressed by reflexes of **!ãĩ*, while the meaning of 'physical beauty' is more commonly rendered by **t'on[i]* (→ **t'ũĩ* in most of the subbranches). In Khwe (at least in ǁXom) the latter root may have also taken on the functions of the former as well, but this seems to be an exceptional situation rather than the norm.

○ CK: ***!ãĩ**. ◊ Reliably reconstructible based on data from both branches; distribution of the root in NKK speaks against a scenario of diffusion from KK. However, the basic terms for 'good = physically beautiful' are completely different in KK and NKK and cannot be etymologized on any mutual grounds.

69. GREEN

○ KK: ***!kxam** (N *!ˀáḿ*, K *!kxam*, Xri *!kxam ~ !ˀam*).

- Khwe: Unclear. ◊ The only well-attested form is ǁXom */kxáó* 'green', a color term applicable to leaves and other green objects.
- Naro-ǂHaba: Unclear. ◊ Attested forms for this meaning are similar in both languages: Naro *cã́ˢ ~ zã́ˢ* (Visser), ǂHaba *ʒǎ̃* (Vossen). However, variation *c- ~ ʒ-* in Naro is unusual, and the forms are impossible to separate from Sasi (ǂHoan) *ʒaˢa* 'green'; note also, perhaps, !Xóõ *ǀāiˢ* 'green', which is phonetically quite similar despite the substitution of affricate by a dental click. It is difficult to construct a single reliable historical scenario for all these forms, but the overall distribution rather points out to a local *Wanderwort* than an inherited term.
- |Gwi-ǁGana: Unclear. ◊ The only well-attested form is |Gwi *ckxāē* [Na.] 'blue, green', with a very rare initial cluster and no external parallels.
- Shua: Unclear. ◊ Cf. in Fehn et al. 2025: Shua *c'áē* 'blue, green'; *t'ālá* 'blue, green'.
- Ts'ixa: **ťálà**.
- Tsua: Unclear. ◊ Not attested in Kua or Tsua. Hie.(in addition to the probable loan *barat^{h}i*) has *čabe* 'green = Tjwao *c'abe (*in Tjwao, cf. also *kˀoara ~ kura* 'green').
- Cua: **cáˢbé** 'green'; **ckxāē** 'green'.

○ NKK: Not reconstructible. ◊ Although some links may be drawn between certain branches (most notably |Gwi *ckxāē* = Shua *c'áē* = Cua *ckxāē*, pointing to a hypothetical prototype like **ckxae*), the concept is overall highly unstable, prone to borrowing, and areal diffusion.

○ CK: Not reconstructible. ◊ It should be noted that KK does have a stable, phonetically unexceptional and semantically well-defined term for ‘green’, sharply contrasting with NKK in this matter. Whether this situation may be projected onto the PCK level, however, remains unknown.

70. KNEE

○ KK: ***‖oa-** (N *‖ő́á-s*, K *‖óá-b*, Xri *‖oa-p ~ ‖õã-p*).
- Khwe: [a] ***‖óɛ́** (‖Ani, ‖Xom *‖óɛ́*); [b] ***kuru** (|Ganda *kúrù*, Buga *kúdù*, ‖Xom *kûrù*). ◊ Both words are attested in ‖Xom, where *kûrù* is glossed simply as ‘knee’, while *‖óɛ́* is given with polysemy (‘knee / joint / node of reed shaft’). The other lects, if Vossen’s data are to be trusted, tend to choose between one of the two synonyms.
- Naro-ǂHaba: ***!úrù** (Naro, ǂHaba *!úrù*).
- |Gwi-‖Gana: Unclear. ◊ Cf. ‖Gana *qúrù*, |Gwi *‖ʔǘ-kœ̀rì* (Vossen). The first term is phonetically similar to Khwe ***kuru**, but genetically incompatible if the first consonant is indeed uvular. The second is a compound formation, but the individual components are obscure.
- Shua: [a] ***‖oe** (Cara, Danisi *‖óé*, Deti *‖òé*); [b] ***kukuru** (Cara *kúkúrù*, |Xaise *kúkúrù*). ◊ Fehn et al. 2025 also list the same two synonyms (Shua *‖óé*, *kúkūrū*) without any notes on semantic or dialectal differentiation. Same situation as in Khwe (see above).
- Ts’ixa: **‖óɛ́** (Fehn), **‖úɛ́** (Vossen). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***‖oe** (Kua *‖ùì*, Tsua *‖óé*, Tjwao *‖oe*). ◊ Irregular vocalic reflexes in Kua. Hie has *kukuru* ‘knee’, just like in Shua, but all the other lects in the cluster go against it.
- Cua: **‖óé**.

○ NKK: ***‖oɛ**. ◊ This root is well attested and reconstructible for Khwe, Shua, and Tsua without any phonetic difficulties, and is further corroborated by external data in KK. Its main competitor is Naro-ǂHaba **!úrù*, which can be genetically correlated with Khwe **kuru* (with regular click loss); far more problematic, however, is its relation with Eastern forms such as Shua **kukuru* and Hie *kukuru* — forms that do not display a phonotactic structure typical of CK. It is tempting to speculate about a contraction from a trisyllabic structure (**kukuru*) to a bisyllabic one with secondary click formation (**kukuru* → **!ᵏuru*), but this would not be a normal development from PNKK to daughter subbranches; at best, such a scenario could only be entertained from an areal perspective, i.e. different scenarios of accommodation of a non-Khoe word (**kukuru*) in the Western and Eastern areas of Kalahari Khoe.

○ CK: ***‖oɛ**. ◊ A straightforward and fully regular isogloss between PKK and PNKK (reflecting the regular shift **-oɛ* → **-oa* in PKK). No support exists in the KK branch for the archaic nature of forms like **!uru* or **kukuru*, which, at this point, are better treated as areal intrusions across the spread zone of NKK.

71. KNOW

○ KK: ***ǂʔan** (N *ǂʔáń*, K *ǂʔán*, Xri *ǂʔan*).
- Khwe: ***ʔã̌** (‖Ani *ʔã̌*, |Ganda, Buga *ʔã́*, ‖Xom *ã́*).
- Naro-ǂHaba: ***!ʔã̄** (Naro *!ʔã̄ː*, ǂHaba *kʔâ̌*).
- |Gwi-‖Gana: ***!ʔã̌** (|Gwi *!ʔã́ː*, ‖Gana *ʔâ̌*).
- Shua: ***ʔã** (Cara, |Xaise *ʔã́*, Danisi, Deti *ʔã̌*).
- Ts’ixa: **ʔã̃ã́** (Fehn), **ʔã̌** (Vossen). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***ʔã** (Kua, Tsua *ʔã́*, Hie *an*, Tjwao *ʔana-ha ~ ʔãː*).
- Cua: **ʔã́ã̄**.

- NKK: ***!ʔã**. ◊ This root is well preserved in all daughter languages, although it is commonly subject to regular (or at least recurrent) click loss.
- CK: (?)***ǂʔan**. ◊ This is a semantically complicated situation. In KK, the semantic opposition between 'to know' and 'to think' is usually expressed as **ǂʔan* 'to know' vs. **ǂʔãĩ* 'to think' (→ N, K, Xri **ǂʔãĩ*), where the latter looks like a causative derivation of the former (with contraction: **ǂʔan-i* → **ǂʔãĩ*). In those NKK languages where this root has been preserved, it commonly means 'to think' by itself (cf. |Gwi, ǁGana *ǂʔáǹ*, Cara, Danisi, |Xaise *ʔyàǹ*, Ts'ixa *ʔyāǹ ~ ǂʔāǹ* /Fehn/), which, in light of the N data, could be thought of as a semantic shift from 'to know', perhaps under the influence of the same old (unpreserved) causative forms. Meanwhile, PNKK **!ʔã* 'to know' corresponds to K *!ʔã* 'to listen; to understand' (from which also comes the K word for 'head') and, most likely, to the second component of the N compound verb *hȍò-!ʔãã̀* 'to feel, perceive; to come to learn, become aware of' (*hȍò* = 'to find'). Again, a semantic shift from 'perceive, learn /new information/' to 'know' is rather trivial. The opposite scenario, according to which the NKK situation is original, requires a more complex chain of assumptions, though it cannot be ruled out by any fully waterproof arguments.

72. LIE

- KK: ***ǁoe** (N *ǁőé*, K *ǁóé*, Xri *ǁoe*).
- Khwe: ***ǁoe** (ǁAni, Buga *ǁóé*, ǁXom *ǁòē*).
- Naro-ǂHaba: ***ǁóé** (Naro, ǂHaba *ǁóé*).
- |Gwi-ǁGana: ***ǁóé** (ǁGana *ǁóé*). ◊ Not attested in |Gwi.
- Shua: ***ǁóé** (ǁGana, Cara, |Xaise, Danisi *ǁóé*, Deti *ǁòé*).
- Ts'ixa: **ǁóé** (Fehn, Vossen). ◊ Aligns with both Shua and ǁXom.
- Tsua: ***ǁoe** (Kua *ǁóé*, Tsua *ǁúé*, Hie *!^{h}oe*, Tjwao *ǁoe*). ◊ The Hie word is glossed as 'to be asleep', but since the meaning 'to sleep' already corresponds in Dornan's vocabulary to a different lexical root (*ǁom*), it may be cautiously assumed that 'to be asleep' here is actually = 'to be lying down'.
- Cua: **ǁóé**.
- NKK: ***ǁoe**. ◊ Highly stable root, preserved everywhere.
- CK: ***ǁoe**. ◊ Perfect phonetic and semantic match between PCK and PNKK.

73. LIVER

- KK: ***kxãĩ-b** (N *ãĩ́- ~ ãĩ̀-*, K *kxãĩ-b*, Xri *kxãĩ-p*).
- Khwe: ***kxãĩ** (ǁAni, Buga *kxãĩ̀*, |Ganda *kxãĩ̀*, ǁXom *kxãĩ́*).
- Naro-ǂHaba: ***kxãĩ** (Naro *kxãẽ*, ǂHaba *kʔãĩ́*).
- |Gwi-ǁGana: ***kxãĩ** (|Gwi *kʔãẽ* [Na.], *kʔãĩ́* [Vossen], ǁGana *kʔãĩ́*).
- Shua: ***kxãĩ** (Cara, Deti *kʔãĩ́*, |Xaise *kʔĩ́*, Danisi *kxãĩ́*).
- Ts'ixa: **kʔãĩ́** (Fehn, Vossen). ◊ Aligns with both Shua and ǁXom.
- Tsua: ***kʔĩ** (Kua, Tsua *ɕ'ĭ̃́*, Hie *če*, Tjwao *kxĩ̃ː ~ kʔĩ̃ː ~ čʔĩ̃ː*). ◊ Tjwao is the only lect here that still preserves (at least in some sub-lects) the original velar articulation of the first consonant.
- Cua: **ɕ'ĭ̃ĭ̃**.
- NKK: ***kxãĩ**. ◊ Highly stable root, preserved in all languages.
- CK: ***kxãĩ.** ◊ Perfect phonetic and semantic match between PCK and PNKK.

74. LONG

- KK: ***ga-xu** (N *kàì-xùű ~ kàà-xùű*, K *gá-xù*). ◊ Not attested in Xri. In N, largely restricted to the Damara dialect, but clearly archaic in light of the direct correspondence in K. Variation be-

tween *kai-* and *kaa-* is unclear (perhaps due to contamination with **kai* 'big'?), as is the nature and origin of the second component (perhaps to be equated with N *xùű* 'to let go' and the correlated terminative suffix **-xu*). In K, cf. also *gú-xū* 'length; thin, high', as well as *ku-xa* 'long' and *kui-hu-b* 'length' (Wuras); perhaps all these forms are actually from **ga(i)-xu-* with vocalic assimilation, although this is contradicted by the opposition between *gá-xù* and *gú-xū* in Meinhof's data.

- Khwe: ***!áò** (‖Ani, Buga *!áò*, ‖Xom *kʸáò*).
- Naro-ǂHaba: ***!áò** (Naro, ǂHaba *!áò*).
- |Gwi-‖Gana: ***!au** (|Gwi *!àù*, ‖Gana *!áù*).
- Shua: ***kao** (Cara *káó*, Danisi, Deti *káò*).
- Ts'ixa: **káò** (Vossen, Fehn). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***kao** (Kua *!áò*, Tsua *!áò*, Hie *ǂxau*, Tjwao *kao*).
- Cua: **!áò**.

○ NKK: ***!ao**. ◊ The reconstruction seems quite straightforward, but distribution of the reflexes causes some concern: the word shows an unusually high rate of conservation of the alveolar click, even in such languages as Kua and Tsua, descendants of Proto-Tsua where the alveolar click must have normally disappeared already at the proto-level (another such exception is the word for 'white', see below). Reasons for such abnormal persistence remain unclear; scenarios such as gradual eastward diffusion from, e.g., Naro (where the alveolar click is most consistently preserved) would seem rather far-fetched, while a scenario of independent recent borrowings from an unknown substrate finds no supporting evidence.

○ CK: Not reconstructible. ◊ As of now, PKK **ga-xu* and PNKK **!ao* find no clear etymological parallels in the other branch of the family, and both stems are somewhat problematic (the former due to its morphological complexity and the latter due to the suspicious status of the alveolar click in daughter languages).

75. MAN

○ KK: ***kxao-b** (N *áò-b*, K *kxao-b* ~ *kxao-kʰoe-b*). ◊ In KK, the meaning 'man = male human being' may either be expressed by the masculine form of the word for 'person = human being', **kʰoe-b*, or by the separate root **kxao-* whose original semantics may have been wider, i.e. = 'male' (of humans and animals). In modern N, however, the meaning 'male' more typically corresponds to the derived form *àò-ré*, while K frequently prefixes the root *ǂxara-* (← PKK **ǂxara* 'testicle') to terms for male animals. Here and also below (in the case of 'woman') we adopt a unified lexicostatistical strategy to both KK and NKK: (a) if the language in question has documented usage of a *separate* lexical root for the concept of 'man' or 'woman', we choose it as the default equivalent; (b) if the language does not document any such roots, we take the respective form of **kʰoe* 'person' as the default equivalent. Thus, for Xri the use of **kxao-b* as 'man' is not attested in Snyman's data, which requires us to accept Xri *kʰoe-p* as the default term and postulate a lexical divergence between Xri, on one hand, and N + K, on the other (allegedly, this could be merely a deficiency of the limited lexical corpus).

- Khwe: (?) ***kxao**. ◊ The only clear attestation here is ‖Xom *kxá-kʰòè* 'man/husband' (contracted from **kxao-kʰoe* 'male-person'). ‖Ani has *kʰóé-mà* 'man' and *kxáò* 'male, manly' (Vossen's data), but the semantic glossing is not very reliable here.
- Naro-ǂHaba: ***kʰóè-ba** (Naro, ǂHaba *kʰóè-ba*). ◊ Text examples from Naro show that this is indeed the most common word for 'man', although a rarer, more specialized *kxáò-kʰōè* 'man = male human being' also exists (in Visser's data). For ǂHaba, Vossen explicitly quotes an opposition between *kxáò* 'male' and *kʰóè-ba* 'man'.

- |Gwi-‖Gana: (?) ***ǂxao(-k^{h}oe)** (|Gwi, ‖Gana *ǂxáó* ‘male’). ◊ In Tanaka 1978: 60, the meaning ‘man’ is rendered by the form *ǂxao-kue* (= *ǂxao-k^{h}oe*) for both languages.
- Shua: (?) ***ǂxao(-k^{h}oe)** (Cara, |Xaise *ǂ'áò*, Danisi *ǂxáò*, Deti *ǂ'áó* ‘male’). ◊ Cf. Fehn et al. 2025: *ǂ'á: ~ ǂ'â: ~ ǂ'áò* ‘male’, *ǂ'á:-k^{h}òè ~ ǂ'â:-k^{h}òè ~ ǂ'â:-k^{h}ō ~ ǂ'áò-khòè* ‘man’. Judging by Fehn’s data, here, too, the primary carrier of the meaning is the root **ǂxao* rather than **k^{h}oe*.
- Ts’ixa: **ǂ'á:-k^{h}óè** (Fehn). ◊ Aligns with both Shua and ‖Xom.
- Tsua: (?) ***ǂxao(-k^{h}oe)** (Kua *ǂ'áò*, Tsua *ǂ'áò* ‘male’; Hie *kau-čo*, Tjwao *ǂxao-čo ~ ǂ'ao-čo*). ◊ Same situation as in Shua (Hie *kau-čo*, Tjwao *ǂxao-čo* ← **ǂxao-k^{h}oe*).
- Cua: **ǂxáò-k^{h}ò**.

○ NKK: ***ǂxao-(k^{h}oe)**. ◊ The most common equivalent for the concept ‘male human being’ in NKK is a compound formation, in which the first morpheme (**ǂxao*) carries the semantics of ‘male’ and the second (**k^{h}oe*) restricts the denoted objects to humans.

○ CK: ***ǂxao**. ◊ Situations in KK and NKK are comparable, but KK more rarely uses the (typical for NKK) compound formation **ǂxao-k^{h}oe*, preferring instead to substitute it with simple masculine forms of the second root (**k^{h}oe-b*). It is interesting to note that of all the subbranches of NKK, only Naro-ǂHaba seems to mirror this situation — a fact that is difficult to separate from Naro as the only direct point of areal contact with Khoekhoe in the Kalahari Khoe continuum.

76. MANY

○ KK: ***ǂui** (N *ǂ̲ùí*, Xri *ǂui*). ◊ This root is not found in extant sources on K; instead, Meinhof lists the equivalent */xóá-sá*, which may be the same word as ‘full’ (q.v.) with an erroneously transcribed click efflux. Reliability is questionable.

- Khwe: Unclear. ◊ The only data are on ‖Xom, where there are two equivalents: (a) *kó ~ á=kó*, the adjectival equivalent of the noun *kó* ‘quantity, multitude’, with multiple confirming examples; (b) *t^{h}íyà* ‘be much, be many / be four’ (no textual examples).
- Naro-ǂHaba: ***káí** (Naro *káí*, ǂHaba *kí:*). ◊ Obviously the same root as **kai* ‘big’ (q.v.), but ǂHaba apparently preserves it only with the semantics of ‘many’.
- |Gwi-‖Gana: Not attested. ◊ Tanaka (1978: 60) lists the form *!tui* ‘many’ for both languages, which would be tempting to compare with KK **ǂui* (since Tanaka’s click transcription may, in fact, easily represent a palatal click with zero efflux); however, ambiguity of the unprofessional transcription and lack of corroborating evidence from other NKK sources make the comparison questionable.
- Shua: Unclear. ◊ Several Shua synonyms are listed in Fehn et al. 2025: *gūrí* ‘many, much’, *t^{h}íyà* id. (western dialect; borrowed from Khwe?), *c^{h}íyà* id. (Deti; variant of the former?), *‖xárà* id. (cf. also *‖xárà-k^{h}òè* ‘crowd; family’, lit. ‘many-people’).
- Ts’ixa: **t^{h}íyà**. ◊ This seems to be a borrowing from Khwe.
- Tsua: Unclear. ◊ Cf. Hie *cao*; Tjwao *‖xara ~ ‖^{h}ara* (see Shua above).
- Cua: **c^{h}ǎà**.

○ NKK: Not reconstructible. ◊ The concept is both poorly attested (Vossen does not usually include it in his survey data) and glaringly unstable.

○ CK: Not reconstructible. ◊ It should, however, be noted that KK (at least Nama and Xri) have a separate and quite basic root for the concept, while most of the NKK equivalents are either polysemous, phonetically unusual, or borrowed.

77. MOUNTAIN

○ KK: (?) ***|ʔui-** (N */ʔűí-*, K */ʔuy-b* [Wuras]). ◊ Same word as ‘stone’. For K, Meinhof lists a different equivalent: *!àrè-b* = N *!̲àrè-s* ‘(granite) boulder’ (Haacke & Eiseb), cf. also *!̲are-b* ‘einzeln

stehender Fels' (Krönlein & Rust). Xri *ǂam-p* = N *ǂa̰a̰̋-b* 'steep, abrupt slope; cliff face'. Another N synonym is *!ʰõmm-i*, without external parallels in K or Xri. Clearly an unstable concept overall, and the link between N and one of the K dialects suggests that lexical differentiation between 'stone' and 'mountain' may have been recent, which explains so many different strategies in between lects.

- Khwe: Not attested. ◊ The only reliably attested equivalent is ǁXom *ndúndù*, a word clearly of non-Khoe origin, judging by its phonetic structure.
- Naro-ǂHaba: Unclear. ◊ Cf. Naro *ǁàbì* 'mountain'; not attested in ǂHaba.
- |Gwi-ǁGana: Unclear. ◊ Tanaka (1978: 63) lists *!kabi* 'mountain' for both lects, which theoretically could reflect **ǁabi* and be consequently comparable with Naro, but this needs confirmation.
- Shua: Not attested. ◊ No data in Vossen's records. Fehn et al. (2025) record three different synonyms for Shua, without differentiation: *kárō*, *|ʾúì*, *ǁai* (the first two words are polysemous with 'stone').
- Ts'ixa: **ǁnóá** (Fehn).
- Tsua: Unclear. ◊ Cf. Hie *ǀoa*; Tjwao *karo*, also *gae*.
- Cua: Not attested.

○ NKK: Not reconstructible. ◊ An unstable concept, sometimes merging with 'stone' and at other times represented by various areal lexemes.

○ CK: Not reconstructible. Perhaps there was no lexical difference between 'stone' and 'mountain' at the proto-level, but other scenarios are possible as well.

78. NECK

○ KK: ***!kxao-** (N *!ʾáó-s*, K *!kxau-b* [Wuras], Xri *!kxao-p* ~ *!ʾao-p*). ◊ Variation between glottalic and affricative efflux is odd (cf. also K *!ʾáó-b* 'nape of the neck' in Meinhof's data), but overall the data support a reconstruction with **!kx-*, further corroborated by external data.

- Khwe: ***!kxao** (ǁAni *!kxáó*). ◊ The ǁAni form is supported by external data; other lects, however, feature an innovation – |Ganda *xèrí*, ǁXom *xèrèɛ́*.
- Naro-ǂHaba: ***!kxáó** (Naro, ǂHaba *!kxáó*).
- |Gwi-ǁGana: ***!kxáó** (|Gwi, ǁGana *kxáó*).
- Shua: ***kʾao** (Cara *kʾáó*, |Xaise *kʾàò*, Deti *kʾàó*). ◊ Innovation in Danisi: *gʸànú* 'neck'.
- Ts'ixa: **ʒānú** (Fehn), **dʸànú** (Vossen). ◊ Aligns not just with Shua in this particular case, but even more specifically with the lexical innovation in Danisi.
- Tsua: ***kʾao** (Kua *kʾàó*, Tjwao *kʾao* ~ *kxao*). ◊ Innovation in Tsua (*kínù*), and possibly the same word in Hie (*ǂano*, but only if *ǂ-* somehow ended up transcribing the palatal stop *ɕ-*).
- Cua: **!kxàō**.

○ NKK: ***!kxao**. ◊ Relatively stable across subbranches, with a few innovations of unclear origin across branches. Note that Danisi *gʸànú* + Ts'ixa *ʒānú*, on one hand, and Tsua *kínù*, on the other hand, probably go back to different roots: the first two reflect **ǂanu*, the second may go back to **kVnu* or **!Vnu*.

○ CK: ***!kxao**. ◊ CK and NKK data are in good agreement with each other.

79. PERSON

○ KK: ***kʰoe** (N *kʰõè-*, K *kʰoe-*, Xri *kʰoe-*). ◊ The root itself carries the semantics of 'human being', which can be restricted to 'male' or 'female' categories by means of gender markers.

- Khwe: ***kʰóé** (ǁAni, |Ganda, Buga, ǁXom *kʰóé*).
- Naro-ǂHaba: ***kʰóè** (Naro, ǂHaba *kʰóè*).

- |Gwi-‖Gana: ***k^{h}oe** (|Gwi *k^{h}óé*, ‖Gana *k^{h}óè*).
- Shua: ***k^{h}oe** (Cara, |Xaise, Deti *k^{h}óé*, Danisi *k^{h}óè*).
- Ts'ixa: **k^{h}ōē ~ k^{h}óè** (Fehn), **k^{h}óè** (Vossen). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***k^{h}oe** (Kua *k^{h}òé*, Tsua *c^{h}òé*, Hie *čo ~ čwa*, Tjwao *čoa*). ◊ Regular palatalization in most lects, but Kua still preserves the original velar.
- Cua: **k^{h}ǒè**.

○ NKK: ***k^{h}oe**. ◊ Super-stable root, preserved in all languages.

○ CK: ***k^{h}oe**. ◊ NKK and CK are in perfect agreement with each other in this case (unlike the situation with the concepts of MAN and WOMAN).

80. RED

○ KK: **|kxaba** (N *|ʔàwā́ ~ |ʔàpā́*, K *|kxābā*, Xri *|kxaba*).

- Khwe: ***ʃ̃oa** (‖Ani *ʃ̃nòá*, Buga *ʃ̃óá*, ‖Xom *|ōá* 'be clean / be red / be holy'). ◊ Loss of nasality in ‖Xom is unclear. Additionally, this root does not seem to be the principal equivalent for the required Swadesh meaning in ‖Xom, with the more common term being *‖qéú-ʆ̃i* '(be) red, reddish', derived from the noun *‖qéú* with polysemy 'red colour, reddish colour/sunset' = |Ganda *‖xɛ́ú* 'red' (acc. to Vossen's data). No external parallels for this item have been found (could the original meaning be simply 'sunset'?), but external data clearly mark it as an innovation.
- Naro-ǂHaba: ***ʃ̃oa^{ʕ}** (Naro *ʃ̃òà^{ʕ}*, ǂHaba *ʃ̃nòá^{ʕ}*).
- |Gwi-‖Gana: ***ʃ̃oa** (|Gwi *ʃ̃òà*, ‖Gana *ʃ̃nóà*).
- Shua: ***|oa** (Cara, |Xaise *|òá*, Deti *|óá*). ◊ Alternate synonym: Danisi *qâ-sè* (see Ts'ixa below), of unclear origin but clearly an innovation, based on external data.
- Ts'ixa: **qāá-sī** (Fehn), **qâ-sì** (Vossen). ◊ Aligns not just with Shua in this particular case, but even more specifically with the lexical innovation in Danisi (see NECK for a similar situation).
- Tsua: (?) ***|oa** (Tjwao *|oa ~ |qoa*). ◊ Archaic nature of the Tjwao item is corroborated by external data. Kua *ʔúdú*, Tsua *wúdú* are a shared innovation of unknown origin; Hie *ʆ̃eye* is unclear due to ambiguity of Dornan's transcription (alternately, cf. also *ǯuruŋe* 'red' and *ta: kue* 'redness, red', both of which are also rather obscure).
- Cua: **‖qōlā**.

○ NKK: ***ʃ̃oa**. ◊ The original root tends to be replaced in modern languages (presumably by more specialized partial synonyms, mostly of unclear origin), but every major subgroup preserves it at least partially. Phonetic reconstruction is uncertain; note the presence of pharyngelization in Naro-ǂHaba, the occasional unpredictable loss of nasality (‖Xom, Shua), and even an occasional uvular efflux (Tjwao *|qoa*), all of which suggest the possibility of a more complex onset (e.g. **ʃ̃q-*?).

○ CK: Not reconstructible. ◊ PKK *|kxaba* and PNKK **ʃ̃(q)oa* 'red' seem to be in fully complementary distribution and have no mutual etymological parallels in the other branch.

81. ROAD

○ KK: ***dao-b** (N *tàò-b*, K *dāō-b ~ dau-b*). ◊ This term is common for N and K and is corroborated by external data. Presumable innovations include N *ʆ̃áǹn-i* 'road, street' (more commonly applicable to modern 'streets'), as well as Xri *ǂaro-p* 'road' = K *ǂārō-b* 'road, path' = N *ʆ̃áró-b* 'footpath'. This evidence points to a PKK opposition between **dao-b* 'road' (large) and **ǂaro-b* 'footpath'; it is unclear if Xri truly merged both concepts or if this glossing simply represents another deficiency of Snyman's lexical corpus.

- Khwe: ***dao** (‖Ani, Buga *dáò*, |Ganda *dáù*, ‖Xom *dáó ~ dàó*).
- Naro-ǂHaba: ***dao** (Naro *dàò*, ǂHaba *díó*).
- |Gwi-‖Gana: ***dao** (|Gwi *díó*, ‖Gana *dáò*).
- Shua: ***dao** (Cara, |Xaise, Deti *dáó*, Danisi *dáò*).
- Ts'ixa: **dáò** (Fehn, Vossen). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***dao** (Kua *dáó*, Tsua *dáó*, Hie *d^{h}au*, Tjwao *dao*).
- Cua: **ǂóró**. ◊ Meaning given as 'road'; the common NKK root *dǎò*, on the other hand, is glossed as 'path, spoor, track' (of animals).

○ NKK: ***dao**. ◊ A very stable root, preserved almost everywhere; detailed dictionaries usually stress the lack of lexical differentiation between 'road' (large) and 'path' (small), the only clear exception being Cua.

○ CK: ***dao**. ◊ A possible scenario would involve an original opposition between 'road' (as used, e.g., by cattle-herders) and '(foot)-path' in PCK, preserved in PKK but neutralized in PNKK due to the speakers' loss of the corresponding semantic distinction. It should be pointed out that of all semantically and phonetically adjacent words in NKK, Cua *ǂóró* is the one closest to PKK **ǂaro-b* 'footpath' (phonetically identical, in fact, if one assumes that the first *o* is the result of vocalic assimilation); however, accepting this relationship is complicated by the necessity to suggest a semantic reversal of meanings in Cua, as well as independent loss of **ǂaro-* in all the other subbranches of NKK, which is highly dubious. Perhaps the similarity is accidental, after all.

82. ROOT

○ KK: ***ǃ̃oma-b** (N *ǃ̃omǎ-b*, K *ǃ̃omã̄-b*, Xri *ǃ̃oma-p*).

- Khwe: ***toɛ** (Buga *tóɛ̄*, ‖Xom *tòɛ̄*). ◊ Not attested in ‖Ani and |Ganda.
- Naro-ǂHaba: ***to^{ʕ}be** (Naro *tò^{ʕ}bē*, ǂHaba *tò^{ʕ}bé*).
- |Gwi-‖Gana: (?) ***tabe** (|Gwi, ‖Gana *tabe* [Tanaka]). ◊ The forms are only attested in Tanaka's dictionary and may not be phonetically accurate; however, the word itself is quite likely the same as in Naro-ǂHaba.
- Shua: Unclear. ◊ The concept is not attested in Vossen's comparative data, but cf. Shua *tóbē* 'root' in Fehn et al. 2025 (clearly the same word as in the Western subgroups above). For the western dialect of Gweta, the same source, however, records *ŋōmá ~ ŋūmá* 'root'.
- Ts'ixa: **tóbē** (Fehn). ◊ Phonetically aligns with Shua rather than Khwe.
- Tsua: (?) ***kxari** (Tjwao *kxari ~ kˀari*). ◊ Hie. *papasi* 'root', judging by its unusual phonetics, is almost certainly a borrowing. Tjwao *kxari ~ kˀari* 'root', although not supported by data from other lects, directly corresponds to the Cua form below and is therefore somewhat more promising.
- Cua: **kxàdī**.

○ NKK: (?) ***to^{ʕ}be**. ◊ Most of the major clusters of NKK show evidence for a form like **tobe*, with the Naro-ǂHaba cluster adding support for original pharyngealization of the main root vowel. However, a form like PNKK **to^{ʕ}be* is formally irreconcilable with the phonetically similar Proto-Khwe **toɛ*, since there are no supporting cases of deletion of intervocalic **-b-* in Khwe. Either the Khwe forms do not belong here, or, perhaps, we are dealing with two old morphological variants, such as simple **toɛ^{ʕ}* (→ Proto-Khwe **toɛ*) and extended **toɛ^{ʕ}-be* (→ **to/^{ʕ}/be* in most of the other clusters).

○ CK: Not reconstructible. ◊ PKK **ǃ̃oma-b* and PNKK **toɛ^{ʕ}(-be)* are more or less in complementary distribution, but Gweta (western Shua) *ŋōmá ~ ŋūmá*, transparently going back to **ǃ̃oma-*, raises some concern: since the form can hardly be a KK borrowing, its attestation would imply the presence of a **ǃ̃oma-* in PNKK, lost (or having avoided detection) in every single lect of

the family except for one small Shua dialect (!). For now, it is perhaps advisable to regard the situation as aberrant (result of a stray contact with some Khoekhoe speakers? independent semantic shift from a common PCK root with a highly specialized meaning? accidental resemblance?).

83. ROUND

- KK: ***!ubu** (N *!ű̱wú ~ !ű̱pú*, Xri *!ubu*). ◊ N actually distinguishes between this root (as 'round = cylindrical') and *ǂù̱wù ~ ǂù̱bù* (as 'round = spherical'); however, only the former is reliably traceable back to PKK. (Could the phonetic variation between two such phonetically and semantically close roots be explained by the factor of sound symbolism?). K *!ʔūm* 'round' is also quite similar phonetically, but impossible to directly compare with N and Xri due to the difference in click efflux articulation.
- Khwe: Not attested. ◊ ǁXom *mbírì*, judging by word-initial *mb-*, is not of Khoe origin.
- Naro-ǂHaba: Unclear. ◊ Not attested in ǂHaba; Naro *gàrù-gàrù* (also *gau-garu*, *ga-garu* id.) looks like an expressive lexeme.
- |Gwi-ǁGana: Not attested.
- Shua: Not attested in Vossen's data. ◊ Cf. *mbárí ~ mbádí* 'round' in Fehn et al. 2025 (of non-Khoe origin; same word as ǁXom *mbírì* above).
- Ts'ixa: **kúlú**.
- Tsua: Not attested.
- Cua: Not attested.
- NKK: Not reconstructible. ◊ Highly unstable and poorly documented concept overall.
- CK: Not reconstructible. ◊ It should be noted, however, that the concept is better attested and somewhat more stable in KK than in NKK.

84. SAND

- KK: ***ǁxae-b** (N *ǁxàé̋-b* 'river-sand'; K *ǁxāːēː-b*, Xri *ǁxae-p*). ◊ N distinguishes between *ǁxàé̋-b* 'river-sand' and *!ò̱wā̋- ~ !ò̱pā̋-* 'fine sand; dune'. The latter is probably a better candidate for the required Swadesh meaning, but has no parallels in K or Xri; only **ǁxae-b* is reliably reconstructed onto the proto-level. Xri also has *|ʔoma-p* 'sand' without external parallels.
- Khwe: Unclear. ◊ Attested only in ǁXom: *qàn̄* 'sand', very likely the same root as in the verbal stem *qáń ~ qā́ǐ̋* 'to bury' (in the sand?). It must be noted that the word *xóḿ* 'earth' is also listed with polysemy 'earth, sand, desert' (but cf. compounds such as *qàn-xóm̀* 'sandy ground' which show that there is sufficient semantic differentiation between the two roots).
- Naro-ǂHaba: ***xóḿ** (Naro, ǂHaba *xóḿ*). ◊ Same root as 'earth' q.v.
- |Gwi-ǁGana: ***xóḿ** (|Gwi, ǁGana *xóḿ*). ◊ Same root as 'earth' q.v.
- Shua: Unclear. ◊ Poorly attested. Fehn et al. 2025 list multiple synonyms for 'sand', including *xóḿ ~ xúḿ* (same word as 'earth') and *ǁ̱árá* 'sand' with parallels in Ts'ixa and ǁXom (see below).
- Ts'ixa: **xúḿ** (Fehn). ◊ Same word as 'earth'. Ts'ixa also has a separate word, *ǁ̱nárá*, for '(white) sand of river' = Shua *ǁ̱árá* 'sand'.
- Tsua: (?) ***xom** (Hie *hom*, Tjwao *xum*). ◊ Not attested in Kua or Tsua specifically in the meaning 'sand' (only 'earth'). Tjwao also has *ǁ̱nara* as an additional synonym (see above).
- Cua: **xòm̄**. ◊ Same word as 'earth'.
- NKK: ***xom**. ◊ There is no convincing evidence that PNKK may have lexically distinguished between 'earth, soil' and 'sand'. ǁXom: *qàn̄* 'sand' most likely goes back to a verbal stem: cf. ǁAni *qáń* 'to cover with sand', |Gwi *qàǹ* 'to bury', and even N *kā́ń* 'to cover with hot ash'. The

Eastern term *ǁnara ‘river-sand’ may be of PNKK origin (cf. ǁXom ǁná:rá ‘white sand /appearing when digging a grave’), but with more restricted semantics than required.

○ CK: Unclear. ◊ See notes on ‘earth’ above which tentatively suggest that the NKK situation may be more archaic than the one observed in PCK.

85. SAY

○ KK: ***mĩ** (N *mĩ̋ĩ̋*, K *mĩ́*, Xri *mĩ*).

• Khwe: ***mĩ́** (ǁAni, ǁXom *mĩ́*). ◊ In the English-Khwe index section of her ǁXom dictionary, Kilian-Hatz also attaches the meaning ‘say’ to such verbs as *ǀʰî* and *ǁâ*, but these are really transitive predicates meaning ‘to say, tell (smth.), inform of (smth.)’, less eligible for inclusion.

• Naro-ǂHaba: ***mi** (Naro *mé:*, ǂHaba *mí*). ◊ Nasalized variants of this root also exist in Naro at least (*mĩ́* acc. to Vossen 1997, *mí: ~ mĩ́:* acc. to Barnard 1985).

• ǀGwi-ǁGana: (?) ***me** (ǀGwi, ǁGana *me* [Tanaka]). ◊ Attested only in Tanaka’s dictionary (so the transcription is not fully reliable), but well supported by external data.

• Shua: ***mi ~ *mĩ** (Cara *mí*). ◊ Not attested in most of Vossen’s data sources, but cf. Shua *mĩ́:* in Fehn et al. 2025.

• Ts’ixa: **mĩ́:** (Fehn). ◊ Aligns with both Shua and ǁXom.

• Tsua: ***mi** (Hie *me*, Tjwao *mi:*). ◊ For Hie, the meaning is glossed as ‘to say to’ (*me-he* ‘to say to oneself’), as opposed to *kwa* ‘to say’, but only the first item finds strong external support.

• Cua: **mĩ́:**.

○ NKK: ***mi ~ *mĩ**. ◊ Although there are some problems with attestation, this basic verb is still well represented in all the clusters. Fluctuation of nasal/non-nasal articulation of the vowel remains poorly understood.

○ CK: ***mĩ**. ◊ KK and NKK are in solid agreement with each other.

86. SEE

○ KK: ***mũ** (N *mũ̀ũ̀*, K *mũ̄*, Xri *mũ*).

• Khwe: ***mũ̂** (ǁAni, ǀGanda, Buga, ǁXom *mũ̂*).

• Naro-ǂHaba: ***mô** (Naro *mô ~ bô*, ǂHaba *mô*).

• ǀGwi-ǁGana: ***m̂** (ǀGwi, ǁGana *m̂*).

• Shua: ***mũ̂** (Cara, ǀXaise, Danisi, Deti *mũ̂*).

• Ts’ixa: **mṹũ̀** ~ **mũ̄ũ̀** (Fehn), **mũ̂** (Vossen).

• Tsua: ***mũ** (Kua *mũ̂*, Tsua *mũ̂*, Hie *mo:*, Tjwao *mũ:*).

• Cua: **mṹũ̀**.

○ NKK: ***mũ**. ◊ This root undergoes occasional irregular developments (mostly in the Naro-ǂHaba / ǀGwi-ǁGana areas, where it can lose nasal articulation of the vowel or even contract to the state of a syllabic nasal), but on the whole is extremely stable.

○ CK: ***mũ**. ◊ KK and NKK are in perfect agreement. Note, however, that only KK derives the nominal stem for ‘eye’ from this root, while NKK preserves the more archaic separate stem **ǂxai* (see Starostin 2024: 230). Of note, perhaps, is the similarity between this root and Proto-Bantu **-bon- ~ *-mon-* ‘to see’ (Guthrie 1970: 56).

87. SEED

○ KK: ***!xum-** (N *!xóm̀-s*, K *!xùmm-a*). ◊ In N this word (with polysemy: ‘seed/descendants, offspring’) contrasts with *ǂʔárá-s* ‘pip, seed (of e.g. date/pumpkin/peach)’, and the latter word, glossed as *ǂkxara-p ~ ǂʔara-p*, is also the only known equivalent for the meaning ‘seed’ in Xri;

there is no way to ascertain if Xri had really merged both sub-concepts into one or if the lack of attestation for **!xum-* was merely accidental. PKK **ǂkxara-* 'pip' could, perhaps, be cautiously compared with the PKK verb **ǂkxara* 'to spit' (→ N *ǂʔáárà*, K *ǂkxára*), itself derived from PKK **ǂkxa-* 'spit, saliva' (i.e. 'pip, kernel /of fruit/' as 'smth. spat out').

- Khwe: ***|xuri** (‖Ani */xúrí*, Buga, ‖Xom */xùí*). ◊ Not attested in |Ganda. Buga and ‖Xom show a completely irregular loss of intervocalic *-r-*; this can only be reasonably interpreted in terms of contamination with some other lexeme (cf., perhaps, ‖Xom */xűì* 'to vomit, spit out', esp. in light of the proposed internal etymology for N *ǂʔáárá-* above?). Additionally, cf. ‖Xom *ǂʔáró* 'flesh of fruits and tubers; fruit, seed, nut (*gen.*)', suspiciously similar to N *ǂʔáárá-* but probably unrelated, both because of the discrepancies in word-final vocalism and the fact that the N word likely goes back to **ǂkxara*, while in ‖Xom there are no signs of a velar affricate efflux.
- Naro-ǂHaba: ***|xuri** (Naro */xùrī*, ǂHaba */xúrí*).
- |Gwi-‖Gana: ***|xúrí** (|Gwi, ‖Gana */xúrí*).
- Shua: ***|xúrí** (Cara, Danisi, Deti */xúrí*). ◊ For |Xaise, Vossen lists the form *ɕódó* 'seed' ← PNKK **coro* 'shell, pod'. Curiously, Fehn et al. 2025 do not have any reflexes of **/xúrí* at all for the meaning 'seed' in Shua; instead, they give *ɕxáí ~ čxáí*, i.e. the same word as 'eye' (clearly a metaphoric shift from the latter meaning).
- Ts'ixa: **ǂʔáró ~ ʔʸáró** (Fehn). ◊ Meaning listed as 'pip/pod/seed', also 'grin' (?).
- Tsua: ***|xuri** (Kua */xúrí*, Tsua */xúrí*, Hie *čai=xai=!xuri*, Tjwao */xuri*). ◊ The Hie form is a compound where the last morpheme likely transcribes a misheard **/xuri* (see external data), the first morpheme = *čai* 'eye', and the second remains unidentified, but may be the same as in *ča-kai* 'eyes (pl.)' (where *kai* is, however, not a productive plural morpheme). The whole expression is thus, literally, 'eye-smth.-seed' or 'eyes-seed'. Actually, the simple root */xuri* is also listed by Dornan, under the meaning 'a grain of seed'.
- Cua: **|xódī**.

○ NKK: ***|xuri**. ◊ This stem is well preserved in all daughter branches. It seems to have disappeared in Ts'ixa; the replacement *ǂʔáró ~ ʔʸáró* is perhaps best interpreted as a loan from Khwe, since no parallels for it have so far been discovered in sources on the Eastern NKK languages. (Note, however, that the presence of Naro *ǂʔáró* 'body, fruit, self' shows that this word is more widespread in the Western area than just the Khwe cluster, and may be reconstructed for PNKK).

○ CK: Not reconstructible. ◊ The opposition between PKK **!xum-* 'seed' and **ǂkxara-* 'pip', on one hand, and PNKK **/xuri* 'seed', on the other, is absolute; none of these words find any etymological support in the other branch (and the similarity between PKK **ǂkxara-* and PNKK **ǂʔaro-* 'fruit, seed', as has been mentioned above, is probably accidental).

88. SIT

○ KK: ***ǂ̃u ~ *ǂ̃ũ** (N *ǂõ-ã̀*, K *ǂú*, Xri *ǂo-a ~ ǂõ-a*). ◊ The pure root is seen in K as well as in N *ǂùű* 'to sit down', and generally corresponds to the dynamic meaning; the static 'to sit' is formed with the suffix **-ã*, probably grammaticalized from the verb 'to be, to stay' (N *hã̀ã̀*).

- Khwe: ***ǂ̃u ~ *ǂ̃ũ** (‖Ani *ǂnű*, |Ganda, Buga *ǂnú*, ‖Xom *ǂű* 'to sit down', *ǂù-ī̃ ~ ǂù-ī ~ ǂòē* 'to sit'). ◊ Most of the attested data (Vossen) corresponds to the meaning 'sit down'; ‖Xom shows that the grammatical opposition is to be reconstructed as **ǂ̃u ~ *ǂ̃ũ* 'to sit down' : **ǂ̃ũ-i* 'to sit' (with the common passive/intransitive suffix **-i*).
- Naro-ǂHaba: ***ǂ̃ũ** (Naro *ǂõ:* 'to sit down', *ǂõ:-à* 'to sit', ǂHaba *ǂnű* 'to sit down'). ◊ Naro conveys the opposition between 'sit' and 'sit down' in exactly the same way as KK; for ǂHaba, the relevant information is missing.

- |Gwi-‖Gana: ***ǂ̃u** (|Gwi *ǂ̃nú*, ‖Gana *ǂ̃nù-è*). ◊ Only the forms for ‘to sit down’ are attested in Vossen’s data, but it is not likely that the static concept would be represented by a different root.
- Shua: ***ɲu ~ *ɲũ** (Cara, Danisi *ɲú*, |Xaise *ɲű*, Deti *ɲǔ* ‘to sit’). ◊ Only the forms for ‘to sit down’ are attested in Vossen’s data. For Shua, Fehn et al. 2025 list *ɲű:* ‘to sit; to sit down; live (dwell)’ and *ɕũĩ ~ čũĩ* ‘to sit’, which is interesting since Shua *ɕ-* typically goes back to PNKK **ǂ-* and thus points to a denasalized variant **ǂũĩ* (← **ǂ̃ũ-i*) — dissimilation before a nasal diphthong?
- Ts’ixa: **ɲú** ‘to sit down’ (Vossen); **ɲű:** ‘to sit down’, **ɲ̃ũ̃ĩ̃** ‘to be seated’ (Fehn). ◊ Phonetically aligns with Shua rather than ‖Xom.
- Tsua: ***ɲũ** (Kua *ɲű*, Tsua *ɲǜ*, Hie *ɲo*, Tjwao *ɲũ: ~ nũ:*). ◊ Only the simple root form is attested in all these languages, glossed either as ‘to sit down’ or ‘to sit’ depending on the source.
- Cua: **ǂ̃ű:** ‘to sit down (eventive); to stay, to live’, **ǂűì̀** ‘to be in a sitting position’. ◊ Of note is the same difference in click effluxes as in Shua, with probable denasalization in the second case.

○ NKK: ***ǂ̃u** ‘to sit down’. ◊ The pure root verb, barring the unpredictable fluctuations of vowel nasalization, is well preserved in all lects. The corresponding static form ‘to be sitting’ may be reconstructed as **ǂ̃u-i* based on data from Khwe, Shua, Ts’ixa, and Cua (the only explicitly contradicting form is Naro *ǂ̃õ:-à*, which may have been areally influenced by Khoekhoe).

○ CK: ***ǂ̃u** ‘to sit down’. ◊ Although the main grammatical means forming the static meaning ‘to sit’ are different between KK and NKK, the base root is always the same.

89. SKIN

○ KK: ***k^{h}o-b** (N *k^{h}òő-b*, K *k^{h}ō:-b*, Xri *kx^{h}o:-p*). ◊ Odd orthography in Xri (a unique case of *kxh-* in Snyman’s transcription), but the word is clearly the same.

- Khwe: ***k^{h}o** (‖Ani, Buga, ‖Xom *k^{h}ǒ*, |Ganda *k^{h}ò*).
- Naro-ǂHaba: ***k^{h}o** (Naro *k^{h}òō*, ǂHaba *k^{h}ô*).
- |Gwi-‖Gana: ***k^{h}ô** (|Gwi, ‖Gana *k^{h}ô*).
- Shua: ***k^{h}o** (Cara, |Xaise, Deti *k^{h}ò*, Danisi *k^{h}ǒ*).
- Ts’ixa: **k^{h}ōó** (Fehn), **k^{h}ǒ** (Vossen). ◊ Aligns with both Shua and ‖Xom.
- Tsua: ***k^{h}o** (Kua *k^{h}ǒ*, Tsua *c^{h}ǒ ~ k^{h}ǒ*, Hie *čo*, Tjwao *čo:*). ◊ Regular palatalization in most languages (Vossen’s data on Tsua are ambiguous).
- Cua: **k^{h}ǒō**.

○ NKK: ***k^{h}o**. ◊ The original root seems to be well preserved in all lects.

○ CK: ***k^{h}o**. ◊ KK and NKK data are in perfect agreement with each other.

90. SLEEP

○ KK: ***‖ʔum** (N *‖ʔóm̀*, K *‖ʔūm*, Xri *‖ʔom ~ ‖ʔõm*). ◊ See SEED for the same vocalic correspondence.

- Khwe: Unclear. ◊ ‖Ani *‖óé*, |Ganda *‖úí* ‘to sleep’ are the same as the verb ‘to lie’ q.v. ‖Xom *ǂʔóm* ‘to sleep’ (no relation to the forms with the lateral click listed above for KK and below for NKK) is clearly related to N *ǂʔòḿ-s* ‘sleep’ (noun), Xri *ǂʔum-i* ‘sleepy’, Naro *ǂʔōm̄* ‘sleepy, drowsy’; meanwhile, ‖Xom *‖ʔóm̀* (the probable reflex of the most widespread NKK root) means ‘to perch at dusk, to roost /of bird/’, a possible narrowing of the original meaning ‘to sleep’. Choosing an optimal candidate for Proto-Khwe with this distribution is hardly possible.
- Naro-ǂHaba: ***‖ʔóm̀** (Naro, ǂHaba *‖ʔóm̀*).
- |Gwi-‖Gana: ***‖ʔóm̀** (|Gwi *‖ʔóaḿ*, ‖Gana *‖ʔóm̀*).
- Shua: (?) ***‖oe** (Cara, |Xaise, Deti *‖óé*). ◊ For Danisi, Vossen lists *‖ʔùḿ* in the meaning ‘to sleep’, which, upon first sight, would seem to mean that Danisi alone preserves the old root for this

concept, while the other lects merge it with 'to lie'. However, the situation in Ts'ixa (see below) suggests that we may really be dealing with a situation in which only the old *static* meaning 'to sleep' (= 'to be asleep') has merged with the concept 'to lie', while the *dynamic* meaning 'to go to sleep', 'to fall asleep' has remained with the old root. (In this case, Vossen's listing for Danisi may simply be semantically inaccurate).

- Ts'ixa: **‖óé** (Fehn, Vossen). ◊ Fehn explicitly opposes Ts'ixa *‖óé* 'to sleep' and *‖ʔūḿ* 'to fall asleep', which may be a reflection of the entire Proto-Shua situation. Here Ts'ixa fully agrees with Shua rather than with ‖Xom.
- Tsua: ***‖oe** (Kua *‖óé*, Tsua *‖úé*, Hie *|oe ~ !we*, Tjwao *‖oe*). ◊ The same word as 'to lie' is used in most of the lects. For Hie, however, in addition to the (probably) mistranscribed words going back to the same root, Dornan also lists *‖om* 'to sleep', which shows that the old root must still have been present in the Tsua cluster as a synonym ('to fall asleep', as in Ts'ixa?)
- Cua: **‖ʔúm̀**.

○ NKK: ***‖ʔum.** ◊ Some of the NKK clusters merge the meanings 'to be asleep' and 'to fall asleep' under one lexical root (this seems to be the case with Naro-ǂHaba, Cua, and maybe |Gwi-‖Gana), while others prefer to merge the first one with the root 'to lie'. Based on mainly the external data of KK, it seems reasonable to view the first situation as primary and the second as the result of semantic innovation.

○ CK: ***‖ʔum.** ◊ PKK and PNKK are in general agreement. It makes sense to reconstruct the verbal roots **‖oe* and **‖ʔum* for PCK as 'to lie' and 'to sleep' respectively, with the former encroaching upon the latter's semantic territory over time.

91. SMALL

○ KK: ***|a** (K *|ā:*, Xri *|a:*). ◊ This original KK root for 'small' is still attested in N as *|a:*, at least in a limited number of contexts, in Krönlein & Rust's dictionary (1969: 74); it is also found in K as the main equivalent for the required meaning. All modern dialects of N, however, replace it with *ǂxàrí̋*, a word of unknown origin; Xri has both *|a:* and *ǂxari* listed as synonyms (as well as *ǂxami*, which could hint at *ǂxa-* being the original base root — or, conversely, could be the result of contamination with the unrelated PKK stem **ǂxam* 'young').

- Khwe: ***kű** (‖Ani, |Ganda, ‖Xom *kű*). ◊ Note that the attested forms may reflect an earlier **!ű*, with regular alveolar click loss in all lects.
- Naro-ǂHaba: **|kxáré** (Naro, ǂHaba *|kxáré*). ◊ Cf. Naro *|kxānà* 'useless, old, without value; small thing (something despisable)', *|kxāī-|kxáì* 'very small' (Visser 2001: 8), which may hint at an original monosyllabic *|kxa-* as the base root.
- |Gwi-‖Gana: (?) ***|kxare** (|Gwi *|kxáré*). ◊ The |Gwi equivalent is the same as in Naro-ǂHaba and is therefore likely to be more archaic. For ‖Gana, Vossen (1988: 97) lists *‖òm̀* 'small', which is somewhat dubious since (a) this is really the same word as SHORT q.v., and (b) Tanaka (1978: 88) gives *|are* (= *|kxare*) 'small' for both languages.
- Shua: ***ɕű̋ĭ̋** (Cara, |Xaise, Deti *ɕű̋ĭ̋*). ◊ This is the most widespread root in the required meaning (should go back to an earlier **ǂű̋ĭ̋*); cf. also Shua *ɕű̋ĭ̋ ~ čű̋ĭ̋* 'small; young; few' in Fehn et al. 2025. For Danisi, Vossen (1988: 97) originally quotes the form *ʔòré* 'small', later also adding *|ʔáré* 'small' (Vossen 1997: 457) — however, this word in all other Shua lects rather means 'thin', 'narrow' (in Fehn et al. 2025, the meanings for Shua *|ʔáré* are given as 'narrow; small; thin').
- Ts'ixa: **ʔòré** (Vossen), **ʔōré** 'be small / be narrow' (Fehn). ◊ Aligns not just with Shua, but quite specifically with the innovation in Danisi.
- Tsua: Unclear. ◊ Kua and Tsua (Vossen's data) have *|ʔàré* 'small' (just like in Danisi above). For Hie, Dornan lists the form *hai-ni* 'small' without any clear etymology. Finally, Tjwao has

cũĩ ‘small’ = Shua **ɕűĩ́*. Without a larger amount of semantically accurate data, the situation in Proto-Tsua remains beyond reliable reconstruction.

- Cua: **|kxáré**.

○ NKK: ***|kxare.** ◊ Although the concept is generally unstable throughout the family, reflexes of this root are found in both the Western and Eastern areas, whereas all the others are limited to one cluster or at least one geographical area (and do not easily find etymological parallels in the other clusters).

○ CK: Not reconstructible. ◊ PKK ***|a** and PNKK ***|kxare** remain without mutual etymological parallels for now, and it cannot be established which root is more archaic.

92. STAND

○ KK: ***mã** (N *mã́:*, K *mã́*). ◊ Not attested in Xri. Dynamic (‘to stand up’) and static (‘to stand’) sub-meanings are not distinguished (although static usage in N may historically result from contraction: *mã́:* + *hã̀ã̀* ‘to stay’, see notes on SIT).

- Khwe: ***tɛ́** (‖Ani *tɛ́*, |Ganda, Buga, ‖Xom *tɛ́*). ◊ Static verb, opposed to **tan* (→ ‖Ani, Buga *tàń*, ‖Xom *tàn̄*) ‘to stand up’.
- Naro-ǂHaba: ***te** (Naro, ǂHaba *té:*). ◊ Static verb, opposed to **ten* ~ **tin* ‘to stand up’ (→ ǂHaba *tìń*, Naro *tə̀ə̄* with vocalization of the coda). (In Vossen 2022: 128, ǂHaba *té:* is given with the meaning ‘stand up’; this is most likely inaccurate, since Vossen 1997: 418 lists *tìń* in the same meaning, which is rather to be expected).
- |Gwi-‖Gana: ***té** (|Gwi *ɕíé* [Na.], *té* [Vossen], ‖Gana *tê*). ◊ Static verb, opposed to **tàń* ‘to stand up’ (|Gwi *tàń* [Vossen] = *ɕěn* [Na.] with palatalization).
- Shua: ***te** (Cara *tè*, Danisi *té*, Deti *tě*). ◊ Static verb, opposed to **tàǹ* ‘to stand up’ (→ Cara, Danisi *tàǹ*). Cf. Shua *té:* ‘stand’, *tān̄* ‘stand up’ in Fehn et al. 2025.
- Ts’ixa: *tê* (Vossen), *tɛ́:* (Fehn). ◊ Static verb, opposed to *tān̄* (Fehn) ‘get up’. Aligns with both Khwe and Shua.
- Tsua: ***te** (Kua *té*, Tsua *tè*, Hie *tʰe*, Tjwao *te:*). ◊ Static verb, opposed to **tan* ‘to stand up’ (→ Kua *táń*, Tsua *tàǹ*, Tjwao *tan*).
- Cua: **té**: ~ **tɛ́**:. ◊ Static verb, opposed to *tàˀn̄* ‘to stand up’.

○ NKK: ***tɛ**. ◊ The opposition between **tɛ* ‘to stand’ and **tan* ‘to stand up’ is also safely traced back to PNKK (the latter root undergoes occasional irregularities in daughter languages, e.g. unclear pharyngealization in Cua and vowel fronting in Naro-ǂHaba, perhaps by analogy with **tɛ*).

○ CK: Not reconstructible. ◊ PKK **mã* and PNKK **tɛ* are in total complementary distribution, which is quite remarkable of verbal roots with such a basic meaning. Of note, perhaps, is the similarity between PKK **mã* and Proto-Bantu **-yim(ad)-* ‘to stand’ (reflexes often have the phonetic shape *-ma* ~ *-ima* in various daughter languages).

93. SWIM

○ KK: ***tʰã̀** (N *cã̀ã̀*, K *tʰã̄* ~ *tʰa:*, Xri *cʰã̀* ~ *tʰã̀*). ◊ Stable root; correspondences are regular (PKK **tʰ-* → N *c-* with fluctuations in Xri).

- Khwe: ***bara** (‖Ani, Buga *bàrá*, ‖Xom *bàrà-á*). ◊ Not attested in |Ganda.
- Naro-ǂHaba: Unclear. ◊ For Naro, Vossen lists *bàdá* ‘to swim’; this is likely the same word as *bàˀrā* ‘to go through’ in Visser 2001: 2, for which an association with ‘swimming’ only springs up for the derivative *baˀra-kxao* ‘one who goes through a difficult situation, one who suffers’ (lit. ‘go-through-male-person’), with one further example: *cʰa: baˀra-kxao-ba* ‘swimmer’ (lit. ‘water-go-through-male-person’). Visser himself gives the equivalent for ‘swim’ as

gǎ̃ã̄-tòè ← *gǎ̃ã̄* ‘duck /bird/’ + *tòè* ‘move’, which is, in turn, similar to ǂHaba *dá:-tùé* ‘swim’ (Vossen 2022: 128), but dá: is unclear (name for another bird? or a misprint for smth. like *gá:* ‘duck’, as in Naro?). Overall, it is clear that the concept is quite unstable in this branch.

- |Gwi-‖Gana: Not attested.
- Shua: ***bara** (Cara, Danisi, Deti *bàrá*, |Xaise *bàrà*). ◊ In addition to Shua *bārá*, cf. also *tuu* and *túbì* ‘to swim’ in Fehn et al. 2025 (no external parallels).
- Ts’ixa: **bàrá** (Vossen), **bārá** (Fehn). ◊ Aligns with both Khwe and Shua.
- Tsua: Not reconstructible. ◊ Not attested for Kua and Tsua. Hie *!xa* ‘to swim’ and Tjwao *‖ʔa:* id. are probably not related, but neither word has reliable external parallels.
- Cua: Unclear. ◊ Cua *dìbī* is sometimes glossed as ‘swim’, but a more detailed semantic notation puts it down as ‘to roll around in dirt, to play around in water’ (← PNKK **dabi* ‘to turn, roll’).

○ NKK: (?) ***bara**. ◊ This root is encountered in at least several Western and Eastern clusters and is formally reconstructible to the PNKK level in the meaning ‘swim’ — although its semantic “profile” in Naro, as described above, makes one wonder if ‘to swim’ was indeed the original semantics (rather than ‘go through smth.’).

○ CK: Not reconstructible. ◊ PKK *$t^h\tilde{a}$ and PNKK **bara* have no etymological connections to each other.

94. THAT

○ KK: ***‖̃a** (N *‖̃àá*, K *‖̃a-*, Xri *‖̃a:*). ◊ Adjectival pronominal stem which can also be nominalized both in N and K, e.g. N *‖̃àá-b* ‘that one (masc.)’, *‖̃àá-s* ‘that one (fem.)’.

- Khwe: Unclear. ◊ Detailed information (though somewhat contradictory, depending on the source) is available only on ‖Xom. According to Kilian-Hatz 2008: 190, the pure root *ǰ̃é* by itself denotes the (intermediate) distal degree, while the complex stem *ǰ̃é-e* denotes the far distal degree. The proximal degree (‘this’) is expressed by *ǰ̃ĩ̃ĩ́*, which looks like a contraction of the same base root (*ǰ̃e*) with yet another particle (*i* or *ĩ*). Cf. the situation in ‖Ani (Vossen 2000: 137), which is described as *ǰ̃né* ‘this’ vs. *ǰ̃né-tè* ‘that’. The two descriptions can be historically reconciled in a scenario which postulates but one base root for all pronouns denoting degrees of proximity (**ǰ̃é*), with the actual degrees further differentiated by means of spatial particles that vary from one language to another. Further complications are introduced by the presence of a whole set of referential pronouns (for Khoe, see the full list in Kilian-Hatz 2008: 190), which will not be discussed here. More information on lects other than ‖Xom and ‖Ani is necessary to fully clarify the situation.
- Naro-ǂHaba: ***ʔa ~ *ʔã̃** (Naro *ā́:*, ǂHaba *ʔà-sá-hà*). ◊ Visser gives the meanings for this Naro form as ‘that; there (far away)’; he also contrasts it with Naro *ē̃:* ‘that (thing) there (near addressee)’ and the extended stem *ē̃:-sí-hā̃:* ‘that there (far away)’ (= Vossen’s *ʔì-sí-hà*). It is possible that both **a* ~ **ã* and **e* ~ **ẽ* represented two different degrees of distal deixis in Proto-Naro-ǂHaba, but ǂHaba only presents comparative evidence for the first one.
- |Gwi-‖Gana: ***ʔa** (|Gwi *ʔáā* [Na.], ‖Gana *ʔá-séá*). ◊ For |Gwi, Vossen only lists the form *ǰ̃nè-si-ha*, which is basically *ǰ̃ne* ‘this’ (q.v.) with the same additional suffixal chain as in ‖Gana (*ʔá-séá* ← **ʔa-si-ha*) and in Naro-ǂHaba (see above). It is possible that both Vossen’s and Nakagawa’s data are correct, but the only root that is straightforwardly reconstructible in the meaning ‘that’ based on data from both languages is **ʔa* — and it is also the only root corroborated by external evidence from Naro-ǂHaba.
- Shua: ***(h)ŭ̃** (Cara, Danisi *ŭ*, Deti *hŭ*). ◊ Cf. Shua *hū̃ū́* ~ *ʔū̃ū́* ‘that’ in Fehn et al. 2025. Interestingly, Vossen’s data contain no indications of nasal vocalic articulation, as opposed to Fehn’s data. The status of initial *h-* is also uncertain (recent prothesis or archaic retention?).

- Ts'ixa: **mé** (Vossen), **m̃ĩ́** (Fehn). ◊ The only form of the distal deictic pronoun in Ts'ixa; does not align with either Khwe or Shua (in fact, has no clear cognates in any Khoe language). Note the same fluctuation between oral (Vossen) and nasal (Fehn) vocalism as in the (unrelated) Shua distal deictic pronouns.
- Tsua: ***(h)ŭ** (Kua, Tsua *ŭ*, Hie *ho*, Tjwao *u:*). ◊ Initial *h-* appears only in Hie and is therefore quite unreliable, but the occasional appearance of *h-* in the clearly related Shua forms somewhat raises the chances of its archaic nature. Note also that Dornan distinguishes between Hie *ho* 'that yonder' and *a* 'that'; the latter has no attested parallels in the rest of Tsua proper, but cf. the parallel in Cua below.
- Cua: **ʔáā**.

○ NKK: A precise reconstruction of the PNKK system of deictic pronouns is hardly possible at the moment, but certain scenarios can be proposed. First, it can be seen that there is a strong tendency in the Western area to form all or most of the deictic pronouns with the aid of a single stem (usually a reflex of **ǁe*) and various extensions; this contrasts with a clear preference for various monovocalic stems (**ʔa, *(h)i, *(h)u*) in the Eastern area. Monovocalic stems are, however, also found in the West (esp. **ʔa*), while reflexes of **ǁe* in the East are largely confined to Ts'ixa — the one language which, as we suspect, has been most seriously influenced by West Khoe. An elegant solution would be to reconstruct three degrees of proximity for PNKK (**ʔi* 'this /near/', **ʔa* 'that /intermediate/', **ʔu* 'that /far away/') plus an optional stem of general deixis (**ǁe*) which could be prefixed to any of the three. In this case, for instance, ǁXom *ǁé-e* 'that /far away/' ← **ǁe-a* (with assimilation), *ǁĩ́ĩ́* 'this' ← **ǁe-i* 'this'. The original system would not be preserved in this way in any single cluster, largely due to the instability of the three-way opposition, but very little, if any, of the remaining evidence flat-out contradicts such a scenario.

○ CK: Unclear. ◊ The distal deictic pronoun **ǁǁa* in PKK finds a very good etymological parallel in PNKK — but the respective pronoun **ǁǁa*, attested in multiple languages both in the Western and Eastern areas, is *referential* rather than deictic ('that mentioned previously'). This could be taken as a sign that PKK is more archaic in this respect (we would rather expect "deictic" → "referential" as a preferred direction of change rather than vice versa), but this is not certain. Note, however, that PKK has no traces of PNKK monovocalic deictic morphemes such as **ʔa*, **ʔu*, and it is probably easier to suggest their elimination in PKK rather than having them appear out of thin air in PNKK.

95. THIS

○ KK: Unclear. ◊ The basic pronominal stem of near deixis has two distinct phonetic shapes in KK languages: (a) with a nasal initial – N *nɛ̀è*; (b) with a laryngeal initial – K *he-*, Xri *he:*. These stems seem to behave in exactly the same way across both N and K, right down to the derived stems (cf. N *nɛ̀è-bá́* 'here' = K *hē-ba* id., N *nɛ̀è-tìí́* 'in this way' = K *hē-ti* id., N *nɛ̀è-cɛ̀ɛ̀* 'today' = K *hē-cē* id.), but tracing them back to the same ancestral form is difficult since the consonantal correspondence is completely unprecedented. See, however, the discussion below (CK section).

- Khwe: Unclear. ◊ See discussion on THAT above.
- Naro-ǂHaba: ***ǁe ~ *ǁẽ** (Naro *ǁêː*, ǂHaba *ǁné*). ◊ Should be distinguished from Naro *ǁǁà*, also listed in the meaning 'this' in Vossen 1997: 254, but better explained as a referential pronoun ('something or someone already referred to') in Visser 2001: 57.
- |Gwi-ǁGana: ***ǁe** (|Gwi *ǁné* [Vossen], *ǁnẽ* [Na.], ǁGana *ǁné*). ◊ Should be distinguished from *ǁǁná́* 'this (anaphoric)' (Nakagawa 1996: 114).

- Shua: ***(h)ǐ** (Cara, Danisi *ǐ*, Deti *hǐ́*). ◊ Cf. Shua *hĩ̃ĭ́* ~ *ʔĩ̃ĭ́* 'this' in Fehn et al. 2025. Same phonetic issues here as in the correlated forms for THAT q.v.
- Ts'ixa: **ǁné** (Vossen), **nēé** ~ **nĩ̃ĭ́** ~ **ŋ́** ~ **ŋ́ĭ** ~ **ǁĩ̃ĭ́** (Fehn). ◊ Forms with the dental click align with Khwe; it is, however, unclear if all the clickless variants listed in Fehn's vocabulary should be regarded as results of dialectal click loss (through frequency of usage?) or if they have a separate origin (for instance, represent a more archaic situation for Ts'ixa, whereas forms with clicks are, instead, recent borrowings from Khwe).
- Tsua: ***ǐ** (Kua, Tsua *ǐ*, Hie *e*, Tjwao *iː*). ◊ For Kua and Tsua, Vossen lists the form *ǁní* 'this' as an additional synonym, but it is not supported by data from Hie or Tjwao.
- Cua: **xá** ~ **xã̄-ì̀**. ◊ The form *xá* is clearly cognate with the anaphoric demonstrative pronoun *xa* in Tjwao (Phiri 2021: 89), whereas *xã̄-ì̀* almost certainly is the result of contraction of *xá* with **ʔĩ̀* 'this' in its nasalized variant (as, e.g., in Shua).

○ NKK: See discussion on THAT above.

○ CK: Unclear. ◊ Considering that even in NKK, the pronominal stem **ǁe* occasionally undergoes irregular click loss (although the only credible example is for Ts'ixa), it may be speculated that N *nèè* historically comes from **ǁe*; in fact, this scenario makes it easier to come up with a common origin for N *nèè* and K *he-*, provided that the original click efflux was more complex (e.g. **ǁʰe* with aspiration?) and click loss had its own separate variations in N and K. If this is true, we end up with an interesting set of semantic correspondences: (a) KK distal deictic pronoun = NKK referential pronoun (**ǁa*); (b) KK proximal deictic pronoun (*ne- ~ he-* ← **ǁʰe* ?) = NKK general (neutral) deictic pronominal stem (**ǁe*). Beyond that, determining which of the two systems is original and which one is the result of historical transformation is quite a challenge; but in any case, the deep discrepancy between KK and NKK in their "deictic strategies" is every bit as thought-provoking as are the etymological connections between separate elements of the two branches' deictic pronominal sub-systems.

96. WALK (= GO)

○ KK: ***!ũ** (N *!ű̀ù̀*, K *!ũ*, Xri *!ũː*).

- Khwe: ***kû̂** (ǁAni, |Ganda, Buga, ǁXom *kû̂*).
- Naro-ǂHaba: ***!û̂** (Naro *!ô̂*, ǂHaba *kû̂*).
- |Gwi-ǁGana: ***!û̂** (|Gwi *!û̂*, ǁGana *kû̂*).
- Shua: ***kû̂** (Cara, |Xaise, Danisi, Deti *kû̂*).
- Ts'ixa: **kû̂** (Vossen, Fehn). ◊ Aligns with both Khwe and Shua.
- Tsua: ***kû̂** (Kua, Tsua *kû̂*, Hie *xo*, Tjwao *kũːˀ*).
- Cua: **kű̀ù̀**.

○ NKK: ***!û̂**. ◊ Preserved everywhere; a highly stable root even in its tonal characteristics.

○ CK: ***!û̂**. ◊ Perfect lexical match between PKK and PNKK.

97. WARM (HOT)

○ KK: ***|am-** (N *|ǎḿ-sá*, K */am/-sa/*, Xri */am*). ◊ All known lects share the verbal root **|am* 'to be / to get hot' (cognate with the word SUN in Kalahari Khoe, see Starostin 2025: 107), but only for N it can be confirmed as the basis for the adjectival derivate 'hot'. In K, */am-sa* 'hot' is attested in data collected by Maingard and Engelbrecht, but not Meinhof (our main source), who instead lists the adjective *ǂkxumm-i* (no external parallels). Xri */am* 'hot' (without any adjectival suffixes) is present in Snyman's data, but is marked as questionable (Haacke & Snyman 2019: 304). With some reservations, the root can be projected onto the PKK level with the meaning

'hot /of objects that have been heated up, e.g. water etc./'; for the more narrow meaning '(naturally) hot /of fire ~ of sun/', one can also reconstruct PKK *ǁ*xõã* (→ N ǁ*xőã̀*, K, Xri ǁ*xõã*).

- Khwe: Unclear. ◊ In the meaning 'hot', the only attested form is ǁXom *čə̂*. The item 'warm' is, however, well reconstructible as Proto-Khwe *k^h*óbò* (→ ǁAni, |Ganda *k^hóbò*, ǁXom *k^hóvò*).
- Naro-ǂHaba: Unclear. ◊ Naro *kùrū-sá* 'hot/feverish' ← *kùrū* 'to get warm, be warm'; however, the adjectival stem 'warm, lukewarm' is formed from a different root = Naro ǁ*ʔōò*. For ǂHaba, however, Vossen (2022: 124) lists ǁ*ʔóò* 'hot', not 'warm' (!).
- |Gwi-ǁGana: (?) ***kuru** (|Gwi, ǁGana *kurú* [Ta.]). ◊ Attested (in the meaning 'hot') only in Tanaka's dictionary, but likelihood of the words representing the proto-item is significantly increased by the Naro cognate *kùrū-sá* (see above).
- Shua: Not attested in Vossen's data. ◊ Cf. Shua *cə̂ə̂* 'heat; hot, warm' in Fehn et al. 2025.
- Ts'ixa: **cū-cúú** ~ **čū-čúú** (Fehn). ◊ Reduplicated stem of unclear origin (does not align with either Khwe or Shua).
- Tsua: (?) ***cə̃** (Hie *ca:*, Tjwao *cə̃:*). ◊ The Hie form, glossed as 'hot', and the Tjwao cognate, glossed as either 'warm' or 'hot' in the default source, are corroborated by a close external parallel in Shua. The situation in Kua and Tsua is, however, not attested. For Hie, cf. also *ǂe:* 'hot' (no clear etymology), */^{h}we-hwe* 'to become hot' and *čobe-* 'warm' in Dornan's data.
- Cua: **hŏrò**.

○ NKK: (?) ***cə̃**. ◊ The isogloss between ǁXom *čə̂*, Shua *cə̂ə̂*, and the Hie/Tjwao forms is not likely to be solved in areal terms and stands as the best candidate as the inherited term for 'hot'; in this case, Naro-ǂHaba and |Gwi-ǁGana may be seen as showing a shared innovation for the verbal stem **kuru* in the meaning 'hot'.

○ CK: Unclear. ◊ Since PKK **/am-* 'hot' acquires the meaning 'sun' in PNKK, and the direction of the semantic shift is typically from the former to the latter, NKK **cə̃* 'hot' is more likely to be an innovation; furthermore, it is fully homonymous with PCK **cə̃* 'to boil, to cook' (Vossen 1997: 458), and the semantic shift 'boiling' → 'hot' would be trivial. Nevertheless, the concept as such is seemingly unstable in the area, and both roots could have co-existed in PCK with narrow sub-meanings such as 'to be hot /of sun/' and 'to be hot /of objects/'.

98. WHITE

○ KK: ***!ʔuri** (N *!ʔùrí*, Xri *!ʔuri*). ◊ K *xātí:* 'white' is also attested in Xri (*xati*) as an additional synonym, but its origin is unknown; the word's phonotactic structure (with an intervocalic *-t-*) is atypical of KK (and Khoisan in general) and suggests at a borrowed origin, but the source of borrowing remains unclear.

- Khwe: (?) ***!ʔé** (|Ganda *!ʔé*, ǁXom *!ʔé-ʃi*). ◊ This is a somewhat dubious root, given the presence of the alveolar click in both |Ganda and ǁXom (most Khwe words of NKK origin regularly undergo click loss with this phoneme). The only alternative, however, is ǁAni *xóɛ́* 'white' = ǁXom *xùɛ́-ʃi* 'be bright white, be very white' = Naro (Visser) *xõẽ-xõẽ-se* 'very white', cf. also ǁXom *xóɛ́* 'foam, froth', from which this adjectival stem is quite likely derived (further KK cognates most likely include N *gòá* 'foam, froth', although consonant correspondences are not fully regular).
- Naro-ǂHaba: ***!ʔu** (Naro *!ʔú:*, ǂHaba *!ʔú*).
- |Gwi-ǁGana: ***!ʔú** (|Gwi *!ʔú*, ǁGana *!ʔú*).
- Shua: ***xóé** (Cara, |Xaise, Deti *xóé*). ◊ Cf. Shua *xúɛ́* ~ *xóɛ́* 'white' in Fehn et al. 2025. Differently in Danisi: *!ô-sé* 'white', with a very rare case of an alveolar click in this lect.
- Ts'ixa: **!ô-sí** (Vossen), **!óò-sī** (Fehn). ◊ Aligns not with Shua in general, but specifically with Danisi (as in Danisi, cases of alveolar clicks in Ts'ixa are quite rare). Fehn also adds syno-

nyms such as *ɸ'ūrú-sì* and *cōū* 'white', with no semantic differentiation and no clear etymological origin.

- Tsua: (?) ***xoe** (Hie *hwe-he*, Tjwao *xue*). ◊ These forms contrast with Kua *!ʔǔ*, Tsua *!ʔú* 'white' in Vossen's data — but the latter are highly atypical for the Tsua cluster, where the alveolar click is normally supposed to be dropped already at the proto-level, and are perhaps better seen as areal re-borrowings, diffused from the western Khoe-speaking area.
- Cua: **!ʔú**:.

○ NKK: Unclear. ◊ A striking detail is that the various NKK branches give evidence for three different, but phonetically similar roots, all of which contain alveolar click phonemes (Khwe **!ʔé*, Naro-ǂHaba / |Gwi-‖Gana / Kua-Tsua **!ʔu*, Danisi-Ts'ixa **!ô*) — despite the fact that only Naro-ǂHaba and |Gwi-‖Gana languages preserve such phonemes on a consistent basis in the inherited lexicon. This raises doubts about their antiquity, and only **!ʔu* has external parallels in PKK (see below). The only other candidate that would be plausible from a phonetic and distributional angle is PNKK **xoɛ*; however, as shown above, its original semantics probably goes back to the concept of 'foam, froth', and it is unclear if the secondary meaning 'white' should really be projected onto PNKK.

○ CK: (?) ***!ʔu-**. ◊ PKK **!ʔuri* 'white' is hardly possible to separate from **!ʔu*, the main equivalent for this meaning in Naro-ǂHaba and |Gwi-‖Gana which also unexpectedly crops up in the Kua-Tsua cluster (with a unique case of initial *!ʔ-*); however, a genetic solution would require the analysis of the Nama form as a later suffixal extension. Further complicating things is the existence of Proto-Ju **!àʔú* 'white' (Starostin 2018: 51), which, if connected, should rather be seen as a possible donor to CK rather than recipient. Another possibility, however, is that all of these forms were independently gained from a "pre-Khoe" (maybe even "pre-Ju") substrate, which would (purely speculatively) explain all the attested variation. Nevertheless, on a strictly formal basis it is possible to reconstruct PCK **!ʔu* 'white', regularly preserved in two out of three Western clusters of NKK and (in an extended suffixal form) in PCK. For Kua and Tsua, however, one would still need to postulate a scenario of diffusion — in fact, it is more likely that the common ancestral stage for all of the Eastern clusters was **xoɛ*, and that forms with *!-* were introduced (or re-introduced) to some of the languages already after the split.

99. WOMAN

○ KK: ***tara-s** (N *tàrâ-s*, K *tara-s*). ◊ With polysemy: 'woman/wife'. This root is not attested in Snyman's Xri corpus, although Meinhof does mention *tara-s* as one of the equivalents for the meaning 'woman'. The other, also common for all attested lects, is PKK **kʰoe-s*, the feminine equivalent of **kʰoe-b* 'man'; see discussion on MAN above for reasons why it is convenient to lexically differentiate between MAN/WOMAN and PERSON on the root level for PKK.

- Khwe: (?) ***‖ɛ**. ◊ ‖Xom *‖ɛ̀ɛ̄-kʰòè* 'woman' consists of *‖ɛ̀ɛ̄* 'female' + *kʰòè* 'person'. This contrasts with the situation in ‖Ani, where the default equivalent seems to be *kʰóé-hɛ̀* ('person' + feminine gender morpheme *-hɛ* = N *kʰõè-s*, etc.). Symmetrically to the case with 'man' (q.v.), we prefer to view the situation in ‖Xom as more archaic, based on external data, but this is not certain.
- Naro-ǂHaba: Unclear. ◊ For Naro, Visser differentiates between *kʰóè-sa* 'woman' (= Khwe *kʰóé-hɛ̀*, N *kʰõè-s*, etc.) and *‖àē-kʰòè* 'young woman, wife' (lit. 'female-person'). For ǂHaba, the precise semantic situation is not attested. Given the parallel situation with MAN q.v., it is tempting to reconstruct **kʰoe-sa* 'woman' as the basic equivalent for Proto-Naro-ǂHaba, but formal evidence for such a reconstruction is lacking.

- |Gwi-‖Gana: ***‖̃ɛ** (|Gwi *‖̃ae-*, ‖Gana *‖̃ɛ-*). ◊ This root by itself means ‘female’; compound forms for ‘women’ are attested in Tanaka’s dictionary as |Gwi, ‖Gana *‖̃aí-kue ~ ‖̃aí-ko-ši* (the second component is **k^{h}oe* ‘person’ + optional feminine marker).
- Shua: ***‖̃ae-** (Cara *‖̃a-*, |Xaise, Danisi, Deti *‖̃ae-*). ◊ Vossen does not record the full stems for the lexeme ‘woman’, but cf. Shua *‖̃āà-k^{h}ō(ê) ~ ‖̃ɛ̄ɛ̀-k^{h}ōê* ‘woman’ in Fehn et al. 2025, showing that in Shua, just as in most of the other clusters, it is typically formed as a compound of **‖̃ae-* ‘female’ + **k^{h}oe* ‘person’.
- Ts’ixa: **‖̃ā:-k^{h}óè** (Fehn). ◊ Aligns with both Khwe and Shua.
- Tsua: ***‖̃ae-** (Kua *‖̃ae-*, Tsua *‖̃ae-*, Hie *‖̃ai-čo ~ ‖̃aie-čwe*, Tjwao *‖̃e:-xo ~ ‖̃e:-čo*). ◊ Vossen does not record the full Kua/Tsua stems for the lexeme ‘woman’, but Hie and Tjwao forms demonstrate that the standard Tsua equivalent is the same compound form as in Shua.
- Cua: **‖̃àē-ɕíà**. ◊ Compound form: *‖̃àē* ‘wife’ (← ‘female’) + unclear second part.
- ○ NKK: ***‖̃ɛ-k^{h}oe.** ◊ The compound formation for ‘woman’ seems well supported by most of the NKK clusters, with the possible exception of Naro-ǂHaba.
- ○ CK: Not reconstructible. ◊ Unlike ‘man’, the concept of ‘woman’ has no single lexical root that would be common for both PKK and PNKK: PKK **tara-* has no parallels in PNKK, whereas PNKK **‖̃ɛ-* has no known cognates in PKK. Both roots are equally likely to be archaic.

100. YELLOW

- ○ KK: Not reconstructible. ◊ N *!^{h}ùnì* ‘yellow’ has no external parallels. K */^{h}āī*, Xri */ʔ^{h}ai* ‘yellow’ = N */^{h}ǎí* ‘grey, ash-coloured’, making the original semantics of this color term difficult to establish. (Xri also has */ʔ^{h}ao* ‘yellow’ as a synonym, but this form has no external parallels).
- Khwe: Not reconstructible. ◊ The concept is attested only for ‖Xom, as *čèrè-tòè-ǂʔú*, lit. ‘silver terminalia (*Terminalia sericea*)’ (*čèrèé*) + ‘root’ (*tòɛ̄*) + ‘color’ (*ǂʔú*), with slight morphophonological changes of the individual stems — clearly a recent formation.
- Naro-ǂHaba: Not reconstructible. ◊ Naro *dàḿ ǂʔūbī* ‘yellow’ = ‘tortoise-egg’ (Visser 2001: 15). For ǂHaba, Vossen (2022: 130) suggests that ‘yellow’ is the same word as ‘white’, which is somewhat dubious (but agrees with the overall flimsiness of the concept).
- |Gwi-‖Gana: Unclear. ◊ For ‖Gana, Tanaka lists the form *dam-!ubi* ‘yellow’ = Naro *dàḿ ǂʔūbī* (see above); despite the obvious recent origin, it is quite possible that Naro and ‖Gana inherit this from a common intermediate ancestor.
- Shua: Not attested in Vossen’s data. ◊ Out of several synonyms listed in Fehn et al. 2025, Shua */xóā* (usually in the idiomatic expression */xóā-k^{h}òà-ʔíī*, like all the other color terms) looks like the only one with some chances of being relatively archaic.
- Ts’ixa: **|kxóà-sī** (Fehn). ◊ Aligns with Shua (but note the irregular discrepancy between click effluxes — which could, allegedly, be due to transcriptional error in one of the sources).
- Tsua: Not attested. ◊ The only known form is Hie *set^{h}a:*, a transparent Bantu borrowing (cf. Tswana *se:ƛ^{h}a* ‘yellow’).
- Cua: **gòē-‖̃úì**. ◊ Compound form: literally = ‘bullfrog-oil’ (see FAT for the second part).
- ○ NKK: Not reconstructible. ◊ This is arguably the least stable concept of all on the 110-item wordlist; new formations are constantly produced based on specific yellow-colored objects (‘silver terminalia’, ‘tortoise egg’, ‘bullfrog’, etc.), and few links exist between individual branches (Naro-ǂHaba and |Gwi-‖Gana being the only exception).
- ○ CK: Not reconstructible. ◊ In this case, KK and NKK seem to display comparable rates of (in)stability: just as in the rest of Khoisan, ‘yellow’ is simply not a part of the “default” color palette.

101. FAR

○ KK: ***ǃ̃u** (N *ǃ̃ùù ~ ǃ̃ùű*, K *ǃ̃u:*, Xri *ǃ̃u:*).
• Khwe: ***ǃ̃û*** (‖Ani *ǃ̃nû*, |Ganda, Buga, ‖Xom *ŋgû*).
• Naro-ǂHaba: ***ǃ̃û*** (Naro *ǃ̃û*, ǂHaba *ǃ̃nû*).
• |Gwi-‖Gana: Unclear. ◊ Vossen (1997: 507) lists *ŋú* ‘far’ for |Gwi (same root as in Naro-ǂHaba), but Nakagawa (1996: 116) mentions *ǂkʻáí* ‘far’ instead, and the latter is further supported by |Gwi *and* ‖Gana *ǂkʻai ~ ǂkʻai-ka* ‘far’ in Tanaka 1978: 34. This root is perhaps to be connected with the PNKK term **ǂkʻáí* ‘face, surface’, cf. especially such Naro derivates as *ǂkʻáí-a* ‘first’, *ǂkʻáí-xa* ‘regularly, usually, for a long time already, all the time’ (Visser 2001: 37) — which, in turn, raises issues of semantic accuracy in Nakagawa’s and Tanaka’s sources. Overall, it is difficult to say whether Proto-|Gwi-‖Gana indeed replaced the original PNKK root in the basic meaning ‘far’ or if Vossen’s attestation of |Gwi *ŋú* ‘far’ should be deemed more important.
• Shua: ***ŋu*** (|Xaise, Danisi *ŋgû*, Deti *ŋú*).
• Ts’ixa: **ŋgû** (Vossen), **ŋgúù** (Fehn). ◊ Aligns with both Shua and ‖Xom.
• Tsua: ***ŋu*** (Kua *ŋgû*, Tsua *gú*, Hie *gu-no*, Tjwao *ŋgu:*).
• Cua: **ǂkʻàī**.
○ NKK: ***ǃ̃u.*** ◊ Well reconstructible based on cohesive evidence from most branches. An interesting additional parallel can be made between |Gwi-‖Gana **ǂkʻai* and Cua *ǂkʻàī*, but if any of these languages truly share this root for the precisely required meaning ‘far’ = ‘not near’, this should probably be considered a homoplastic (semantically independent) development.
○ CK: ***ǃ̃u.*** ◊ Perfect agreement between KK and NKK.

102. HEAVY

○ KK: ***ǃum*** (N *ǃ̰ṓḿ*, K *ǃúm*, Xri *ǃom*). ◊ See SEED, SLEEP for the same vocalic correspondence.
• Khwe: ***ǃúḿ*** (‖Ani *ǃóḿ*, |Ganda *kúḿ*, Buga *kóḿ*, ‖Xom *kóm̄*).
• Naro-ǂHaba: ***ǃúḿ*** (Naro *ǃóḿ*, ǂHaba *ǃúḿ*).
• |Gwi-‖Gana: ***ǃóḿ*** (|Gwi *ǃóám*, ‖Gana *ǃóḿ*).
• Shua: ***kúḿ*** (Cara, |Xaise, Deti *kóḿ*, Danisi *kúḿ*). ◊ Cf. Shua *kúḿ ~ kóḿ* in Fehn et al. 2025.
• Ts’ixa: **kúḿ** (Vossen, Fehn). ◊ Aligns with both Shua and ‖Xom.
• Tsua: ***kúḿ*** (Kua *kóḿ*, Tsua *kúḿ*). ◊ Not attested in Hie or Tjwao.
• Cua: **kúḿ**.
○ NKK: ***ǃúḿ***. ◊ Stable and well-preserved in all branches.
○ CK: ***ǃum***. ◊ Another perfect match between KK and NKK.

103. NEAR

○ KK: ***|u** (N *ǀ̰úù*, K *|u-*, Xri *|ǂkʻu:se*). ◊ Adjectival use in K is only attested in Wuras’ dictionary (|u-se: ‘near’); Meinhof has */u-/u* ‘to bring near’ and */u-e:* ‘to approach, get near’. The appearance of a velar affricate in Xri is inexplicable (it is certainly not supported by any external data and looks like the result of “hypercorrection”), but cf. also */u:-/u:* ‘to bring near’ in the same corpus.
• Khwe: ***|û*** (‖Ani, |Ganda, Buga, ‖Xom */û*).
• Naro-ǂHaba: ***|û*** (Naro, ǂHaba */û*).
• |Gwi-‖Gana: ***|û*** (|Gwi, ‖Gana */û*).
• Shua: ***|û*** (Cara, |Xaise, Danisi, Deti */û*).
• Ts’ixa: **|û**. ◊ Aligns with both Shua and ‖Xom.

- Tsua: ***|û** (Kua, Tsua */û*, Hie */u*, Tjwao */u:*).
- Cua: **|ū**:.
- ○ NKK: ***|û.** ◊ Stable and well-preserved in all branches.
- ○ CK: ***|u.** ◊ Another perfect match between KK and NKK.

104. SALT

- ○ KK: ***ǂkxo-** (N *ǂʔòò-*, K *ǂkxo:-b*, Xri *ǂkxo:-p ~ ǂʔo:-p*).
- Khwe: ***dobe** (‖Ani *dóbè*, Buga *dòbé*, ‖Xom *dòvèē*).
- Naro-ǂHaba: ***daˢbe** (Naro *tàˢbē*, ǂHaba *dàˢbé*). ◊ Pharyngealization in Naro is confirmed by all the sources (cf. *dàˢbéˢ* in Vossen 1997: 481, *dăˢbé* in Barnard 1985: 98); for ǂHaba, Vossen lists the pharyngealized variant in Vossen 2022: 126 (although the same form is listed simply as *dàbé* in Vossen 1997: 481, perhaps by accident).
- |Gwi-‖Gana: ***daˢbe** (|Gwi *dàˢbē* [Na.], *dàˢbɛ́* [Vossen], ‖Gana *dàˢbè*).
- Shua: ***dobe** (Cara, Deti *dóbè*, |Xaise *dèbè*, Danisi *dóbé*). ◊ Vocalic assimilation in |Xaise.
- Ts'ixa: **dōbē** (Fehn). ◊ Aligns with both Shua and Khwe.
- Tsua: ***debe** (Kua, Tsua *dèbé*, Hie *debe*, Tjwao *debe:*). ◊ Note the same vocalic assimilation through the entire cluster as in |Xaise. Regarding the final vowel length in Tjwao, it is curious to note that for Hie, Dornan differentiates between *debe* 'salt' and *debe:* 'salt pan' (result of contraction with an unknown suffix?).
- Cua: **dòbē**.
- ○ NKK: (?) ***dobe**. ◊ The main difficulty here is that Khwe, Shua, and Tsua **dobe* (→ Tsua **debe*), on one hand, and Naro-ǂHaba / |Gwi-‖Gana **daˢbe*, on the other, are not easy to reconcile with each other. Pharyngealization in the latter two clusters could be archaic and regularly disappear in the other clusters, but the vowel correspondence (*o* : *a*) is utterly irregular. Furthermore, the variant **daˢbe* is hardly separable from N *tàbé ~ tàwé* 'to lick /brackish substance/', *tàbé-s* 'brackish spot, salt-lick', both of which also have dialectal variants with *n-* (*nàwé*). The latter observation makes it reasonable to seek a further connection with the word for 'tongue' (N *nàḿ-mi ~ tàḿ-mi*; note Naro *tàˢm̄*, also with pharyngealization!); an old verbal derivative such as **da/ˢ/m-e* would have likely yielded **da/ˢ/be* in accordance with the general laws of Khoe phonotactics. In this case, either PNKK **dobe* and Nama / Naro-ǂHaba / |Gwi-‖Gana **daˢbe* are accidental look-alikes (the former representing the original PNKK root for 'salt' and the latter being a new areal formation from the word for 'tongue'), or all the **dobe*-like forms are the result of areal diffusion of **daˢbe* with irregular vocalic correspondences formed through areal transmission.
- ○ CK: Not reconstructible. ◊ In light of the discussion above, the isogloss between N *tàbé* and NKK **dobe ~ *daˢbe* is not particularly telling from a genetic point of view, and even if all these forms were genetically related, we would still be in the dark as to the root's original meaning. In pure theory, an archaic opposition between **ǂkxo-* 'salt' (preserved only in KK) and **daˢbe* 'salt-lick' (transitioning to simply 'salt' in NKK) cannot be excluded.

105. SHORT

- ○ KK: ***ǃ̃ubu** (N *ǃ̃ùwú ~ ǃ̃ùbú*, K *ǃ̃ùbù*, Xri *ǃ̃ubu ~ ǃ̃obu*).
- Khwe: ***‖òḿ** (‖Ani, |Ganda, Buga *‖òḿ*, ‖Xom *‖ōḿ*).
- Naro-ǂHaba: ***‖om** (Naro *‖òm̀*, ǂHaba *‖óm̀*).
- |Gwi-‖Gana: ***‖òm̀** (|Gwi *‖òam̀*, ‖Gana *‖òm̀*).
- Shua: ***‖om** (Cara, |Xaise, Danisi *‖òḿ*, Deti *‖óm̀*). ◊ Cf. Shua *‖ōm̄ ~ ‖ūm̄* in Fehn et al. 2025.

- Ts'ixa: **‖óm̀** (Vossen), **‖ūm̀** (Fehn).
- Tsua: ***‖óm̀** (Kua *‖óm̀*, Tsua *‖úm̀*, Hie *‖om-ɲe*). ◊ Not attested in Tjwao.
- Cua: **‖ōm̀**.

○ NKK: ***‖om.** ◊ Stable and well-preserved in all branches.

○ CK: Not reconstructible. ◊ PKK **ǂ̃ubu* and PNKK **‖om* have no mutual etymological connections.

106. SNAKE

○ KK: ***|kxao-b** (N */ʔàó-b*, K */kxāō-b*, Xri */kxao-p*). ◊ An additional Xri synonym, *ǂ̃au-p*, is not corroborated by any external data (it could signify some specific kind of snake).

- Khwe: Unclear. ◊ Equivalents are only attested for ǁXom: *ǂʔíyō* 'snake' and *ǂqòyō* id., without any semantic differentiation. The *-yō* coda in both cases is very rare and atypical of Khoe root structure; given that from a distributional point of view, both terms are historical innovations, a borrowed origin for both (albeit purely speculative) is not out of the question.
- Naro-ǂHaba: ***|kxao** (Naro */kxāō*, ǂHaba */kxáò*).
- |Gwi-ǁGana: ***|kxáó** (|Gwi */kxáó* [Na.], */kxáú* [Vossen], ǁGana */kxáó*).
- Shua: ***|ʔáó** (Cara, |Xaise, Danisi, Deti */ʔáó*). ◊ For Deti, Vossen (1988: 97) also lists the synonym */ǎ-kà-kû̀*, which can be literally translated as 'walking-on-the-belly' (see BELLY, WALK /GO/ for the individual components) — probably a recent euphemism.
- Ts'ixa: **|ʔíī** (Fehn). ◊ A unique Ts'ixa lexeme, although phonetic similarity with ǁXom *ǂʔíyō* should not be ignored (the Ts'ixa form cannot be borrowed directly from ǁXom, but perhaps both can be retentions from some earlier substrate?). Judging by text examples, this is the primary word for the concept in the modern language; however, Fehn also adds the common Shua/NKK term */kxāó* 'snake' as an extra synonym, but in a very telling context — *Xuukhoedam a |kxao e* "in the language of the San, (snake) is */kxao*" (possibly hinting at the word's archaic status). Finally, Vossen 1988: 97 lists the Ts'ixa equivalent as */ǎ-kà-kû̀*, i.e. the same euphemistic expression as in Deti (see above).
- Tsua: ***|kxáó** (Kua */ʔáó*, Tsua */ʔáó*, Hie *ǀ̲auo ~ ǁ̲ao*, Tjwao */kxao*). ◊ Tjwao still preserves the original velar affricate efflux (which merges with the simple glottalic efflux in most of Tsua).
- Cua: **|kxáō**.

○ NKK: ***|kxáó.** ◊ Reliably reconstructible, based on data from at least four major clusters.

○ CK: ***|kxao**. ◊ KK and NKK data are in perfect agreement here.

107. THIN

○ KK: ***|ʔui** (N */ʔùì*, K */ʔu:ì*, Xri */ʔui*). ◊ In N, this word has the specific meaning 'slim, slender; narrow (of road); thin (of stick, rope)', e.g. 'thin' as applied to two-dimensional objects. A separate term, *ǁ̲àá*, has the meaning 'thin, flimsy (of smth. flat), sheer (of textiles), thin (of fluid), weak, thin (of hair)', e.g. 'thin' of one-dimensional objects; its only external parallel is K *ǁ̲ã* 'thin' (unspecified) in Engelbrecht's data, but it is quite possible that the opposition goes all the way back to PKK.

- Khwe: Unclear. ◊ The concept is attested only for ǁXom: */íní* 'be thin' (of objects; should be distinguished from *ǁʰầĩ́* '(be) meagre, thin; weak, feeble', presumably mainly of people).
- Naro-ǂHaba: Unclear. ◊ Visser lists Naro *ʃ̃ôẽ̄ˁ* 'thin' (= Vossen's *ʃ̃ùíˁ* in Barnard 1985: 120), distinct from *càīˁ* 'thin (of persons, animals)'. Not attested in ǂHaba.
- |Gwi-ǁGana: Not attested.
- Shua: Not attested in Vossen's data. ◊ Cf. Shua */ání ~ */íní* 'thin' in Fehn et al. 2025.
- Ts'ixa: **|íní** (Fehn). ◊ Aligns with both Shua and ǁXom.

- Tsua: Not attested. ◊ Hie *‖ai-ču* ‘lean, thin’ is likely the same root as ǁXom *ǁʰâĭ́* ‘meagre, thin’, but this is not the required semantics for this concept.
- Cua: **|xádī**.

○ NKK: Not reconstructible. ◊ ǁXom: */íní* = Shua */áni ~ /íní* = Ts’ixa */íní* is clearly the best candidate, but distribution of attested reflexes can be interpreted as areal rather than genetic. Overall, more data are necessary to make a properly informed decision.

○ CK: Not reconstructible. ◊ The concept is well represented in KK, but seems to be far less stable in NKK. Note an obvious parallel to KK in Ju|ʔhoan */ùʔí* ‘to be thin /of paper, etc.’ — probably an areal connection, but the direction of borrowing is hard to establish (although the phonetic structure of the Ju word would suggest borrowing into KK, with transposition of glottalic articulation to click efflux position).

108. WIND

○ KK: ***ǂʔoa-b** (N *ǂʔóà-b*, K *ǂʔōā:-b ~ ǂʔoa-b-i*). ◊ For Xri, Snyman lists the related form *ǂʔoa-p* in the meaning ‘strong wind’, while ‘wind’ as such is rendered by *‖ʔao-p*, a form of unknown origin.

- Khwe: ***ǂʔã́** (ǁAni, |Ganda, Buga, ǁXom *ǂʔã́*).
- Naro-ǂHaba: ***ǂʔã́** (Naro *ǂʔã́:*, ǂHaba *ǂʔã́*).
- |Gwi-ǁGana: ***ǂʔã́** (|Gwi, ǁGana *ǂʔã́*).
- Shua: ***ʔyã** (Cara, |Xaise, Danisi *ʔyã́*, Deti *ʔyã̌*).
- Ts’ixa: **ǂʔã́:** ~ **ʔʸã́:** (Fehn). ◊ Vossen’s shape of this item aligns strictly with ǁXom, but Fehn lists the click-containing (Khwe-like) and clickless (Shua-like) variants as doublets.
- Tsua: ***ʔyã** (Kua *ʔyã̂*, Tsua *ʔyã̂*, Hie *ɲa:*). ◊ Not recorded in Tjwao, where the basic equivalent for this concept seems to be *xai* (no known etymology).
- Cua: **ǂʔĭ́ã́**.

○ NKK: *ǂʔã́. ◊ Stable in all clusters and perfectly reconstructible for PNKK.

○ CK: Unclear. ◊ It is highly tempting to trace PKK *ǂʔoa- and PNKK *ǂʔã back to the same source, but the vocalic correspondence here is unique, since PKK *oa normally corresponds to either PNKK *oa or *oɛ. Although a sporadic irregularity (due to contamination with some other root?) cannot be excluded, we probably have to treat KK and NKK entries here as separate etymological entities, at least until it is demonstrated that the *oa : *a correspondence is actually recurrent. Note that a widespread equivalent for ‘wind’ in North Khoisan is the root *ǂʔa (Snyman 1997: 55), which cannot be a borrowing from KK for phonetic reasons.

109. WORM

○ KK: ***ǂkxoni-b** (N *ǂʔùnì-b*, K *ǂkxònì-b*, Xri *ǂʔuni-p*). ◊ Loss of velar affricate efflux in Xri is strange but not unprecedented (for many words with *-kx-*, doublets are attested in Snyman’s corpus).

- Khwe: Unclear. ◊ Attested only in ǁXom: *‖xèí* ‘worm/maggot/caterpillar’.
- Naro-ǂHaba: Not reconstructible. ◊ For Naro, Visser lists *l̃õ:ˢ* ‘worm in manure, beetle larva, caterpillar’ and */kxã̄õ̀* ‘ringworm (in the body)’. For ǂHaba, Vossen gives *‖̃nàíˢ* ‘worm’ = Naro *‖̃àēˢ* ‘(small) puff adder’ (no idea about the degree of semantic accuracy).
- |Gwi-ǁGana: Not attested.
- Shua: Not attested in Vossen’s data. ◊ Cf. Shua *‖āí* ‘worm’ in Fehn et al. 2025.
- Ts’ixa: **‖nāí** (Fehn). ◊ Aligns better with Shua than with ǁXom.
- Tsua: Not reconstructible. ◊ Only attested in Hie *‖xaie* ‘worm’, which may, however, be cognate with the forms attested in Shua, Ts’ixa, and Cua.

- Cua: **‖ǎĩˢ**. ◊ Polysemy: ‘worm/maggot’.

○ NKK: (?) ***‖naiˢ.** ◊ Although the word is not very well attested, most of the clusters seem to feature at least one or two reflexes of it. Naro-ǂHaba and Cua give evidence for pharyngealized articulation of the coda; Eastern languages show that the click efflux must have been complex (something like the pre-nasalized voiced click in Ts’ixa?).

○ CK: Not reconstructible. ◊ PKK **ǂˀxoni-b* and PNKK **‖naiˢ* are two different roots without any mutual etymological parallels.

110. YEAR

○ KK: ***kuri-b** (N *kùrí-b*, K *kuri-b ~ gūrīː-b*, Xri *kuri-p*).

- Khwe: ***kúrí** (‖Ani, |Ganda, Buga ‖Xom *kúrí*).
- Naro-ǂHaba: ***kúrí** (Naro *kūrī*, ǂHaba *kúrí*).
- |Gwi-‖Gana: ***kúrí** (|Gwi, ‖Gana *kúrí*).
- Shua: ***kúrí** (Cara, |Xaise, Danisi, Deti *kúrí*).
- Ts’ixa: **kúrí**. ◊ Aligns with both Shua and Khwe.
- Tsua: ***kúrí** (Kua *kúrí*, Tsua *cúrí*, Hie, Tjwao *čuri*).
- Cua: **ɕúdī**.

○ NKK: ***kúrí**. ◊ Stable and well-preserved in all the languages.

○ CK: ***kuri**. ◊ KK and NKK are in perfect agreement with each other.

Preliminary observations (by way of conclusion)

Following the same scheme that was earlier introduced in Starostin 2024, 2025, I summarize below (Table 1) the results of intermediate reconstructions for Proto-Khoekhoe (PKK) and Proto-Non-Khoekhoe (PNKK), as well as for the top level of Proto-Central Khoisan = Proto-Khoe (PCK). Reconstructions in square brackets represent probable lexical/semantic innovations in the respective intermediate protolanguage. Question marks in the PCK column note the impossibility of making an “optimal” choice between PKK and PNKK given our current state of knowledge on Khoe etymology (note that this can mean one of *three* possible options, including a scenario in which the original equivalent was replaced independently in both primary nodes of the family).

Additionally, I also list all the semantic connections (both polysemies and historically more or less certain semantic shifts) between the analyzed etyma and other meanings; for details on particular connections (names and number of languages, direction of shift, etc.) the actual data lists in the paper should be consulted.

Based on this table, we can make the following conclusions:

1. For almost half of the entries (29 out of 60), it is formally impossible to come up with a reliable and unambiguous *single* onomasiological candidate for Proto-Khoe — not just because of the transparent and strongly divergent binary branching of the family, but also because in most of these cases, the items reliably reconstructible for Proto-Khoekhoe and Proto-Kalahari Khoe find *no etymological traces* in the opposite branches, at least, not at the current level of our knowledge and investigation. Given the lack of detailed sources on many of the languages (particularly in the Kalahari area), it is quite likely that in the future, some mutual etymological support for these forms will be discovered, but overall, this is a rather surprising situation — and it begs the question of just how many of the forms listed in this table have, in fact, been inherited in the intermediate ancestral stages from Proto-Khoe itself.

Word	PCK	PKK	PNKK	Semantic connections
ALL	***hoɛ**	**hoa*	**ʔwɛ*	
BARK	***‖kxũ**	[**soro-*]	**‖kxũ*	‘shell, cover’; ‘dry/dead bark’
BELLY	**?**	**ǀ̃a-b*	**ǂ̃ʰa*	
BIG	**?**	**kai*	*?*	
BITE	***pa**	**ba*	**pa*	‘to open the mouth’
BREAST	**?**	**‖xai-b*	**‖̲u*	‘breastbone’; (?) ‘sternum of ostrich’
BURN (TR.)	***dao**	**dao*	**dao*	
CLOUD	**?**	**ǁ̃anu-s* (?)	**‖um* (?)	‘rain’; ‘sky’
COLD	***!qʰai**	**!xai*	**!qʰai*	‘cool’; (?) ‘to shiver’
COME	***ha**	**ha*	**ha*	
EARTH	**?**	**!ʰu-*	**xom*	‘sand’
FAT (N.)	***‖̲núì**	[**kau-*]	**‖̲núì*	‘fat (adj.)’
FEATHER	**?**	**!ʔam-i*	**ǀʔũ*	‘hair’
FISH	***‖ʔau**	**‖ʔau-b*	**‖ʔau*	
FLY (V.)	**?**	*?*	**‖abu* (?)	‘to rise’, ‘to jump’, ‘to migrate’, ‘wing’
FULL	***|kxoɛ**	**|kxoa-sa*	**|kxoɛ*	
GIVE	***ma**	**ma*	**ma ~ *mã*	‘to share’, ‘to distribute’
GOOD	***!ãĩ**	**!ãĩ*	**!ãĩ*	‘handsome, pretty’
GREEN	**?**	**!kxam*	*?*	
KNEE	***‖oɛ**	**‖oa-*	**‖oɛ*	
KNOW	***ǂʔan** (?)	**ǂʔan*	[**!ʔã*]	‘to think’, ‘to perceive/understand’
LIE (V.)	***‖oe**	**‖oe*	**‖oe*	‘to sleep’
LIVER	***kxãĩ**	**kxãĩ-b*	**kxãĩ*	
LONG	**?**	**ga-xu*	**!ao*	
MAN	***kxao**	**kxao-b*	**kxao-(kʰoe)*	‘male’, ‘person’
MANY	**?**	**ǂui*	*?*	‘big’, ‘quantity’
MOUNTAIN	**?**	**ǀʔui-* (?)	*?*	‘stone’
NECK	***!kxao**	**!kxao-*	**!kxao*	
PERSON	***kʰoe**	**kʰoe*	**kʰoe*	
RED	**?**	**|kxaba*	**ǁ̃(q)oa*	‘sunset’
ROAD	***dao**	**dao-b*	**dao*	‘path’
ROOT	**?**	**!ǒma-b*	**toʕbe* (?)	
ROUND	**?**	**!ubu*	*?*	
SAND	**?**	**‖xae-b*	**xom*	‘earth’, ‘to cover /w. sand or ash/’
SAY	***mĩ**	**mĩ*	**mi ~ *mĩ*	
SEE	***mũ**	**mũ*	**mũ*	
SEED	**?**	**!xum-*	**|xuri*	‘pip’, ‘fruit’
SIT	***ǂ̃u**	**ǂ̃u ~ *ǂ̃ũ*	**ǂ̃u*	
SKIN	***kʰo**	**kʰo-b*	**kʰo*	
SLEEP	***‖ʔum**	**‖ʔum*	**‖ʔum*	‘to lie’, ‘to roost’
SMALL	**?**	**|a*	**|kxare*	‘young’, ‘thin’, ‘narrow’
STAND	**?**	**mã*	**tɛ*	

Table 1. Khoekhoe, Non-Khoekhoe, and Proto-Central Khoisan reconstructions for Swadesh items 51–110.

Word	PCK	PKK	PNKK	Semantic connections
SWIM	**?**	**tʰã́*	**bara*	'to go through (smth.)'
THAT	**?**	**‖a*	**ʔa/*ʔu* (?)	
THIS	**?**	**ne/*he* (?)	**ʔi* (?)	
WALK (GO)	***!ũ**	**!ũ*	**!ṹ*	
WARM (HOT)	**?**	**|am-*	**cã́* (?)	
WHITE	***!ʔu-** (?)	**!ʔuri*	**!ʔu* (?)	'foam, froth'
WOMAN	**?**	**tara-s*	**‖ɛ-kʰoe*	'female', 'wife'
YELLOW	**?**	*?*	*?*	'grey', 'tortoise egg', 'bullfrog', 'silver terminalia'
FAR	***ǃ̃u**	**ǃ̃u*	**ǃ̃u*	
HEAVY	***!um**	**!um*	**!úḿ*	
NEAR	***|u**	**|u*	**|û*	
SALT	**?**	**ǂkxo-*	**dobe* (?)	'salt-lick', 'to lick salt'
SHORT	**?**	**ǃ̃ubu*	**‖om*	
SNAKE	***|kxao**	**|kxao-b*	**|kxáó*	'belly-walker'
THIN	**?**	**|ʔui*	*?*	
WIND	**?**	**ǂʔoa-b*	**ǂʔã́*	
WORM	**?**	**ǂkxoni-b*	**‖naiˢ* (?)	'maggot', 'caterpillar'
YEAR	***kuri**	**kuri-b*	**kúrí*	

Table 1. Khoekhoe, Non-Khoekhoe, and Proto-Central Khoisan reconstructions for Swadesh items 51–110 (continued).

2. On the positive side, these results are in solid agreement with the overall splitting of the Swadesh wordlist into two halves based on their average stability (as originally proposed in Starostin 2012): the first half only yielded around 10 items for which it was completely impossible to suggest the primary (more archaic) nature of the intermediate proto-item in either Khoekhoe or Kalahari Khoe. While the individual stability ratings for Khoe would naturally differ from the average ranking of the items (cf., for instance, such super-stable Khoe items as BURN, FULL, LIVER, PERSON, etc.), generally the family behaves as it would be expected to on this behalf.

3. It may be seen, even more clearly than before, that Khoekhoe and Kalahari Khoe sometimes display significant differences in their lexicalization of concepts. Among these are: (a) Khoekhoe's consistent lexical distinction between transitive and intransitive BURN, as opposed to Kalahari Khoe's tendency to use the same root for both; (b) Khoekhoe's colexification of SOLID FAT with FAT (adj.) rather than with LIQUID FAT = OIL in Kalahari Khoe; (c) Khoekhoe's possession of a special term for GREEN, unlike Kalahari Khoe here that meaning is quite "volatile"; (d) for WOMAN, a preference for using either a specialized lexical root (**tara-*) or a feminine form of PERSON (**kʰoe-s*) in Khoekhoe over the more common compound 'female-person' in Kalahari Khoekhoe, etc. These differences are likely to be explainable in areal terms and warrant comparison with the correlating semantic divisions in other African families, both those of the "Khoisan area" and of "non-Khoisan" lineages to the north of Khoe territory.

Although the assembled evidence, with complete lists published here and available in StarLing database format, permits to produce a new phylogenetic tree for all the included Khoe lects, publishing a "raw" version without first submitting the evidence to another round of historical analysis would be somewhat premature. As we have seen, in many cases there are reasons to suspect areal interference, sometimes mutual (e.g. Naro-Nama contacts), sometimes on the part of other "Khoisan" lineages (such as Ju or Taa), and sometimes on the part of uni-

dentified substrates or adstrates; failing to take those into consideration while producing a lexicostatistical tree may (and indeed will) result in glaring errors of both phylogenetic classification and glottochronological dating. Because of this (and also due to volume considerations), we shall postpone the relevant discussion for a final (fourth) part of this series, with a focus on the following questions:

1) does the major lexical rift between Khoekhoe and Kalahari Khoe truly reflect a deep chronological split, or should it rather be accepted as a result of substrate/adstrate interference (possibly related to the necessarily implied language shift on the part of some of the family's speakers)?;

2) does the data speak in favor of Rainer Vossen's earlier hypothesis of a binary split of Kalahari Khoe into the Western and Eastern branches, or is this split more of an illusion triggered by areal waves of phonetic change in the region? consecutively, is there sufficient lexical evidence to postulate *any* large clusters of a genetic nature within Kalahari Khoe?;

3) what is the status of Ts'ixa? should it be considered part of the Shua cluster, part of the Khwe cluster, or a separate entity on the Kalahari Khoe tree in between the two?;

4) what is the status of Cua? is it part of the Tsua cluster despite being so notoriously more archaic in its phonology, or is the lexical evidence against such a classification after all?;

5) an additional question is that of Khoe's relation to the extinct Kwadi (Fehn & Rocha 2024), with which it seems to form a generally accepted single family — in particular, how can the status of Kwadi as an outlier contribute to the onomasiological reconstruction of the Proto-Khoe wordlist itself?

Only after a serious look at all these issues on an etymon-by-etymon basis, and a subsequent correction of the assembled wordlists and their cognation indexes, would it finally make sense to present a lexicostatistical tree for the Khoe family. The final part of the present study will hopefully be able to address all of these problems individually and present a definitive classification (at least inasmuch as the currently available data permit to establish).

References

Barnard, Alan. 1985. *A Nharo Wordlist with Notes on Grammar*. Durban: University of Natal.

Collins, Chris, Zachary Wellstood. 2025. *A Grammatical Sketch of Cua*. Berlin / Bruxelles: Peter Lang.

Dornan, Samuel S. 1917. The Tati Bushmen (Masarwas) and their language. *Journal of the Royal Anthropological Society of Great Britain and Ireland* 47: 37–112.

Du Plessis, Ménan. 2018. *Kora: A lost Khoisan language of the Early Cape and the Gariep*. Cape Town: University of South Africa Press.

Fehn, Anne-Maria. 2014. *A Grammar of Ts'ixa (Kalahari Khoe). Ph.D. thesis*. Köln: Institut für Afrikanistik.

Fehn, Anne-Maria, Arnold Ketapilwe, Tshiamo Kebuelemang. 2024. *A dictionary of Ts'ixa: a Khoe language of northern Botswana. Manuscript Version*. Vairão: Network of CIBIOInBIO TwinLabs in Africa.

Fehn, Anne-Maria, Jorge Rocha. 2024. Lost in translation: A historical-comparative reconstruction of Proto-Khoe-Kwadi based on archival data. *Diachronica* 40(5): 609–665.

Fehn, Anne-Maria, Tshisimogo Leepang, Jorge Rocha. 2025. *A Dictionary of Shua. Manuscript Version*. Vairão: Network of CIBIO-InBIO TwinLabs in Africa.

Guthrie, Malcolm. 1970. *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages. A catalogue of common Bantu with commentary Part 2 (Volume 3)*. London: Gregg Press Ltd.

Haacke, Wilfrid, Eliphas Eiseb. 2002. *A Khoekhoegowab Dictionary with an English-Khoekhoegowab Index*. Namibia: Gamsberg Macmillan.

Haacke, Wilfrid, Eliphas Eiseb, Levi Namaseb. 1997. Internal and external relations of Khoekhoe dialects: a preliminary survey. In: Wilfrid Haacke, E. D. Elderkin (eds.). *Namibian languages: reports and papers*: 125–203. Köln: Rüdiger Köppe Verlag.

Haacke, Wilfrid, Jan W. Snyman. 2019. Lexical proximity of a Xri corpus to Khoekhoegowab. *Studies in African Linguistics* 48(2): 267–328.

Kilian-Hatz, Christa. 2003. *Khwe-English Dictionary*. Köln: Rüdiger Köppe Verlag.

Meinhof, Carl. 1930. *Der Koranadialekt des Hottentotischen*. Berlin: Dietrich Reimer Verlag.

Nakagawa, Hirosi. 1996. A first report on the click accompaniments of |Gui. *Journal of the International Phonetic Association* 26(1): 41–54.

Nakagawa, Hirosi. 2006. *Aspects of the phonetic and phonological structure of the G|ui language. Ph.D. Thesis*. Johannesburg: University of the Witwatersrand.

Phiri, Admire. 2021. *The Verbal and Nominal Morpho-syntax of Tjwao: A Grammaticalisation Approach. Ph.D. dissertation*. Cape Town: Stellenbosch University.

Rust, Friedrich Jr. 1969. *Nama Wörterbuch (Krönlein Redivivus)*. Pietermaritzburg: University of Natal Press.

Starostin, George. 2018. Lexicostatistical Studies in Khoisan I: The Ju-ǂHoan Relationship. *Journal of Language Relationship* 16(2): 19–61. doi: 10.31826/jlr-2018-161-205

Starostin, George. 2024. Lexicostatistical Studies In Khoisan III/I: Reconstructing a Swadesh wordlist for Proto-Khoe (items 1–25). *Journal of Language Relationship* 22(3–4): 216–240. doi: 10.31826/jlr-2025-223-404

Starostin, George. 2025. Lexicostatistical Studies In Khoisan III/II: Reconstructing a Swadesh wordlist for Proto-Khoe (items 26–50). *Journal of Language Relationship* 23(1–2): 97–119. doi: 10.31826/jlr-2026-231-208

Starostin, Georgiy. 2013. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoj klassifikacii. Tom I: Metodologija. Kojsanskije jazyki*. Moskva: Jazyki slav'anskoj kul'tury.

Starostin, Sergei. 2007. Opredelenije ustojčivosti bazisnoj leksiki. In: Sergei Starostin. *Trudy po jazykoznaniju*: 825–839. Moskva: Jazyki slav'anskix kul'tur.

Tanaka, Jiro. 1978. *A San Vocabulary of the Central Kalahari: G‖ana and G|wi Dialects*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Traill, Anthony. 1994. *A !Xóõ dictionary*. *Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 9*. Köln: Rüdiger Köppe Verlag.

Visser, Hessel. 2001. *Naro Dictionary*. Botswana: Printing & Publishing Co.

Vossen, Rainer. 1986. Zur Phonologie der ‖Ani-Sprache. In: Rainer Vossen, Klaus Keuthmann (eds.). *Contemporary studies on Khoisan (Festschrift Oswin R. A. Köhler), v. 2*: 321–345. Hamburg: Helmut Buske Verlag.

Vossen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe Verlag.

Vossen, Rainer, Sabine Neumann, Christina Patriarchi, Margit Rottland, Rainer Spörl, Beate Vagt. 1988. Khoe Linguistic Relationships Reconsidered: The Data. In: Rainer Vossen (ed.). *New perspectives on the study of Khoisan. Quellen zur Khoisan-Forschung / Research in Khoisan studies, Bd 7*: 67–108. Köln: Rüdiger Köppe Verlag.

Wellstood, Zachary. 2024. *Cua Dictionary*. Ms. (unpublished).

Wuras, Carl Friedrich. 1920. *Vokabular der Korana-Sprache. Herausgegeben und mit kritischen Anmerkungen versehen von Walther Bourquin*. Berlin: Dietrich Reimer Verlag.

Г. С. Старостин. Лексикостатистические исследования по койсанским языкам III/III: реконструкция списка Сводеша для працентрально-койсанского (пракхойского) языка (элементы 51–110)

Настоящая работа представляет собой третью часть лексикостатистического анализа базисной лексики языков семьи кхой (Южная Африка); впервые проводится попытка ономасиологической реконструкции второй, «нестабильной» (в среднестатистическом плане) половины списка Сводеша для пракхойского языка с целью выявления лексических замен от пракхойского состояния к промежуточным праязыковым состояниям и в конечном итоге к современным языкам. В совокупности с проведенной ранее реконструкцией первой половины списка результаты исследования формируют основу для построения устойчивой филогенетической классификации кхойской семьи на основе глоттохронологического метода и ручного анализа общих инноваций (соответствующую классификацию предполагается опубликовать в последней, завершающей статье автора на тему кхойской лексикостатистики).

Ключевые слова: койсанские языки; кхой языки; лексикостатистический анализ; ономасиологическая реконструкция; историческая семантика.

*Andrei F. Stoica**, *Xue C. R. Stoica*†
* Independent researcher (Chifeng); andreifstoica@outlook.com
† Independent researcher (Chifeng); xuecrstoica@outlook.com

On some Romanian substrate words with Albanian comparanda: etymologies and implications

The paper offers a new angle on the etymologies of some Romanian substrate words with close Albanian comparanda, focusing on distinguishing each word's language of origin. Two new Albanian loanwords are identified, *berc* 'with a short tail; tailless' and *drizǎ* 'hairy crabgrass', while *rangă* 'crowbar' and *urdă* 'a kind of cheese similar to ricotta' are identified as Thracian borrowings. Further, *groapă* 'pit' (< **gʰrobʰ-eh₂*) is attributed to a minor substrate that preserved PIE **o* — Paeonian or Brygian. Finally, *mărar* 'dill' is derived ultimately from Ancient Greek **maráthrion*, a diminutive of *márathron* 'fennel', through the mediation of both Thracian and Albanian.

Keywords: Romanian language; Romanian etymology; Paleo-Balkan languages; Albanian language; Thracian language.

1. Introduction

Romanian preserves a considerable number of words from Paleo-Balkan languages — the poorly attested languages native to pre-Roman Southeastern Europe. These words were acquired largely through the shift of Paleo-Balkan speakers to Latin and account collectively for the vast majority of the Romanian lexical substrate.[1] Currently, there is no consensus on the linguistic stratification[2] or periodization of this vocabulary,[3] and the problem of the Romanian substrate is intimately connected to the yet-unsolved problem of the Romanian Urheimat. Based on historical considerations, substrate words may have entered Balkan Latin as early as the 2nd century BC, when Latin began to spread in the interior of the Balkan Peninsula following the Third Macedonian War (171–168 BC), and as late as the final quarter of 1st millennium AD, prior to the cessation of Romanian-Albanian[4] contacts and the extinction of Daco-Moesian and Thracian.[5] This lengthy period of often intensive contacts with Paleo-Balkan languages[6] accounts for the significant influence that these languages exerted upon Romanian.

[1] Greek, usually not classed as a Paleo-Balkan language due to its extensive attestation (see Нерознак 1978: 7), also contributed to the Romanian substrate: Romanian exhibits a layer of early Greek loans of substrate typology — particularly names of small animals, such as *cărăbuș* 'cockchafer', *drepnea* 'blackswift', and *mormoloc* 'tadpole' (Diculescu 1927: 486–489). Additionally, any influences acquired from languages of non-Balkan origin that predated Latin in Southeastern Europe (e.g., Sarmatian) may be regarded as substratal.

[2] Research aimed at identifying layers within the Romanian lexical substrate has been limited. An early attempt, now outdated, was Philippide (1927: 695–743). Later, Ivănescu (1980: 251–257) distinguished an Albanian and a non-Albanian layer in the Romanian substrate, but without providing an extensive analysis.

[3] According to Ivănescu (1980: 171), the bulk of Romanian substrate words entered the language between 450 and 700 AD. Other authors propose earlier dates: Paliga (2024: 43–44) asserts that Thracian influenced Balkan Latin until the 6th century AD.

[4] Albanian, regarded by Matzinger (2018: 1790) as the continuer of an undocumented Balkan idiom distinct from both Illyrian and Thracian, is the sole modern descendant of a Paleo-Balkan language.

[5] Daco-Moesian and Thracian appear to have been still in widespread use at the time when the Slavs arrived in Southeastern Europe. See Yanakieva (2018: 59–61) on the situation of East Balkan languages in Late Antiquity.

Among the Romanian substrate words, particular attention has been given to those with close parallels in Albanian.[7] However, despite more than 160 years of research efforts, the problem of identifying and etymologizing Romanian substrate words with Albanian comparanda remains unsolved.[8] The identification of these words is complicated by the very late attestations of both languages[9] and the heterogeneity of the shared Romanian-Albanian vocabulary.[10] Among Romanian substrate words common with Albanian, reliable etymologies are largely limited to those analyzable as native in Albanian, of which some evidence Albanian-specific developments,[11] whereas others do not.[12] Several important Romanian substrate words shared with Albanian remain with no plausible Albanian etymology.[13]

In recent decades, it has often been assumed that this group of Romanian substrate words largely derives from a single language (Nagy 2003: 194). This presumptive source language has been regarded as ancestral or otherwise very closely related to Albanian (Brâncuș 2022: 46; Matzinger 2018: 1790; Paliga 2024: 46; Poghirc 1969: 314; Slușanschi 1981: 340). However, the difficulty in providing Albanian etymologies for a significant share of these words suggests that they did not originate in a single language or an "Albanoid" group[14] of closely related languages.

The linguistic diversity of the ancient Balkans offers a compelling alternative to the Albanoid hypothesis as an explanation for the Paleo-Balkan vocabulary shared by Romanian and Albanian. Multiple languages were spoken in pre-Roman times across the territory between present-day Romania and Albania, including Daco-Moesian, Thracian, Paeonian and Brygian.[15] While Albanian has itself been often suggested to continue either Daco-Moesian or Thracian (Brâncuș 2013: 419; Georgiev 1981: 132–140; Paliga 2024: 46; Witczak 1995: 310–311), it markedly differs from these languages in its treatment of several PIE consonants and clusters, particularly **sC* groups (Kaluzhskaya 1993: 70–71; Ködderitzsch 2001: 82), rendering such affiliations implausible. Albanian likewise diverges from Paeonian and Brygian, having shifted PIE **o* to **a* in pre-Roman times (Matzinger 2017: 39–40), whereas the latter languages both preserved it (Duridanov 1985: 134–135; 1993: 64–65).

Given these considerations, the essential principle guiding our investigation is taking into account the diversity of possible sources of the shared Paleo-Balkan vocabulary of Romanian and Albanian. This paper follows this framework to analyze six such words, offering arguments for their attribution to a particular language rather than to a generic substrate.

---

[6] See Rusakov (2017: 556–557) for information on the intensive Romanian-Albanian contacts, discernible at multiple linguistic levels.

[7] An early discussion of this subject is available in Miklosich (1861: 6–10).

[8] Reflecting this uncertainty, Brâncuș (2005: 48–50) estimates that Romanian has between 90 and 150 substrate words shared with Albanian. New candidates continue to be proposed, such as Rom. *înghesui* 'to cram, jam', linked by Schütz (2002: 19) to Alb. *gjesh* 'to knead'.

[9] Albanian and Romanian are virtually undocumented for most of their history. Substantive attestations first emerge in the 15th century AD for Albanian (Rusakov 2018: 1716) and around 1500 for Romanian (Maiden et al. 2020: 3–4).

[10] Romanian and Albanian also share extensive Latin and Slavic vocabulary layers besides the Paleo-Balkan one.

[11] One such example is Rom. *vatră* 'hearth' < Alb. *vatër* id., whose /v/ arose via the Alb. **ā- > *wɔ-* sound change in the earlier **ātrā* 'fire' (Schumacher 2009: 46).

[12] An example is Rom. *gard* 'fence' ~ Alb. *gardh* id. < **g^hord^h-o-* < PIE **g^herd^h-* 'to enclose, gird' (Schumacher 2013: 211). The word formation is straightforward and the phonetic developments are not diagnostic: PIE **o > a* and **g^h, *d^h > g, d* have also been attributed to Dacian and Thracian by Duridanov (1995: 45–46).

[13] An example is Rom. *copil* 'child' ~ Alb. *kopil* 'bastard'. The main attempt at an Albanian etymology starts from Alb. *pjell* 'to give birth (of animals)' and assumes a *ko-* prefix and an **'of bad birth' original meaning (Hubschmid 1983: 507–510). The speculative word formation and phonetic difficulties render this etymology unpersuasive.

[14] See a brief presentation of the Albanoid hypothesis in Friedman and Joseph (2025: 33–34).

[15] See Duridanov (1995: 45; 2000: 23–25) and Hill (2011: 411).

2. Etymologies

2.1. Rom. *berc* 'with a short tail; tailless'

This word has often been considered of unknown etymology (DLR 1913–1949, 1, pt. 1: 545; Șăineanu 1929: 62; Tiktin 1903–1925, 1: 177).

The explanation from a surmised Lat. **brevicus* < *brevis* 'short' (Bogrea 1921: 251) requires an arbitrary metathesis. Moreover, other Romance languages do not validate **brevicus*, and the adjective-forming Lat. *-icus* suffix has not been historically productive in Romanian.

The derivation from OHG *brecha* 'breaking' (Scriban 1939: 164) is semantically vague.[16] Formally and distributionally, the hypothesis presents further difficulties, hinging upon the same unexplained metathesis and requiring an unlikely OHG *x* > Rom. *k* outcome[17] as well as Old High German-Romanian contacts, which are not evidenced. Hypothetical mediation through an unattested Slavic form (Ciorănescu 2001: 88) fails to resolve these difficulties.

Substratal etymologies have also been proposed. Vinereanu (2023, 1: 240) derives it from PIE **mreǵʰ-* 'short' through Dacian, with an unexpected metathesis and velar stop.[18] Paliga (2024: 187) regards *berc* as a probable substrate word from a surmised *bar-/ber-* root extracted from toponyms and presumed to mean 'hill'. However, such a root is not validated, and the semantic link to 'hill' is not compelling.

Saramandu (1987: 126–130) connects *berc* with Rom. *bercă* 'sheep with short and curly wool'[19] and Alb. *berr* 'sheep; goat', but regards it as a diminutive of a lost **ber* through an earlier **berce*, with an unclear suffix. Saramandu (1987: 130) further suggests a 'sheep with short and curly wool' > 'with a short tail; tailless' semantic evolution, which is unconvincing.

We propose instead that *berc* was borrowed from Alb. *berk* 'goat'. The Albanian word is an internal diminutive of Alb. *berr* id. (Orel 1998: 21).[20] Early Romanians may have encountered *berk* as a term used to describe goats that resembled sheep in some way. A conspicuous difference between goats and sheep is that goats have much shorter tails than sheep; the historical importance of pastoralism among Romanians favored the generalization of an initial meaning *'sheep with a short tail' to 'animal with a short tail'.[21] Another conspicuous difference between goats and sheep is that goats have much shorter wool than sheep; this explains the meaning of *bercă*.[22] For the semantic shift from the name of an animal to a perceived attribute of it, cf. Hun. *ravasz* 'sly' < 'fox' (Kálmán 1972: 56).

Finally, it is worth noting that *berk* is a word of low distribution in Albanian — Orel (1998: 21) mentions it is used only in *era berk* 'stench of a goat'. However, given the large temporal gap between the age of intensive Romanian-Albanian contacts, dated to the mid-1st millen-

[16] One could, at most, surmise a connection with tail docking, a practice performed in farm animals such as pigs, sheep and cattle (Sutherland and Tucker 2011: 180).

[17] Such an outcome appears sporadically in auslaut in Old Church Slavonic loans, but the more frequent outcomes of OCS *-x* are Rom. *-f*, *-v*, *-u̯* (with dialectal variation), and *-h* in learned borrowings (Petrovici 1948: 190–192).

[18] PIE **ǵʰ* > Dacian *z* (Ködderitzsch 2001: 82).

[19] Rom. *bercă* lent Hun. *birka* 'sheep; lamb, mutton', Slovak *birka* 'short-wooled sheep', dialectal (Moravian) Czech *burka* 'sheep' (Pușcariu et al. 1926: 279), dialectal Russian *birka* 'lamb, sheepskin'. Zaicz (2006 81) instead considers Hun. *birka* of either native or Slavic origin, but without explaining its word formation.

[20] Related to PSl. **bôrvъ* 'boar, ram, wether' (Orel 1998: 21) < PIE **bʰor-wo-* (Derksen 2008: 58).

[21] Another semantic shift attributed to the historical predominance of pastoralism among Romanians is Rom. *făt* 'boy' < 'young animal' (Candrea and Densusianu 1907: 91–92).

[22] The feminine gender of Rom. *bercă*, in contrast with the masculine gender of Alb. *berk*, may reflect the influence of the feminine gender of Rom. *oaie* 'sheep'.

nium AD (Rusakov 2017: 557), and the first consistent attestations of Albanian, this circumstance does not strongly challenge our etymology. It may be that *berk* enjoyed wider usage in the mid-1st millennium AD, whether in Albanian as a whole or in a variety thereof that experienced greater contact with the Common Romanian dialect ancestral to Daco-Romanian. The absence of *berc* in Aromanian and Megleno-Romanian supports the latter scenario, suggesting that the word was borrowed from a northern Albanian variety which may have since disappeared through Romanization or Slavicization.

2.2. Rom. *drizǎ* 'hairy crabgrass'

MDA (2001–2003, 2: 226) provides no etymology for this rare word. DELR suggests an untenable derivation from German *Waldhirse* 'foxtail millet' through **valtirzǎ* and **valdirzǎ*.

The word is a transparent loan from Alb. *drizë* 'thornbush'. The Albanian word derives from **dri-* 'bush', a root shared with AGr. *dríos* id., with the Albanian diminutive suffix *-zë*.[23] See the discussion of *drizë* in Demiraj (1997: 146).

2.3. Rom. *groapǎ* 'pit'

There is little doubt that this word continues an o-grade result noun from PIE **gʰrebʰ-* 'to dig' (LIV 2001: 201–202). The comparison to Alb. *gropë* id., German *Graben* 'ditch', OCS *grobŭ* 'pit, hole' was already made by Miklosich (1861: 9). What remains is to identify the language of origin.

A Slavic (Cihac 1879: 129–130; Georgiev 1965: 76; Ungureanu 2016: 210), Thracian (Paliga 2024: 307)[24] or Dacian (Vinereanu 2023, 1: 704) origin cannot explain the /p/, while a borrowing from OHG *crôpa* 'pit, cavity' (Orel 1998: 124)[25] is distributionally impossible.

DLR (1913–1949, 2, pt. 1: 315) derives *groapǎ* from Albanian, which does not suffice to clarify the history of the word. Regarding the /p/, Hyllested (2016: 73–77) brings compelling arguments for a syllable-final **-b-* > **-p* development in early Albanian, occurring after the apocope of the unstressed vowel. Since the **-p* in the new word-final position was retained after vocalic suffixation, such as with the feminine suffix *-ë* (Hyllested 2016: 74), Albanian can account for the **bʰ* > *p* outcome. However, *gropë* remains difficult to explain as a native Albanian word. This is because Alb. /o/ does not continue PIE **o*, but PIE **eh*$_{1/2}$ (Demiraj 1997: 41–45), which **gʰrebʰ-* cannot provide. If starting instead from **gʰrēbʰ-*, /o/ would be expected, but this requires an arbitrary long grade.

A Slavic > Albanian > Romanian borrowing scenario has been suggested by Meyer (1891: 131) and Tiktin (1903–1925, 2: 703). Matzinger (2006: 213) posits that the early date of the borrowing explains the surmised *b* > *p* outcome. Yet this hypothesis is seriously marred by

[23] On this suffix, see Arapi (2012: 134). Other Romanian words borrowed from Albanian suffixed with *-zǎ* include *coacǎzǎ* 'currant' (Çabej 1964: 73) and *pupǎzǎ* 'hoopoe' (Orel 1998: 350).

[24] Paliga claims that PIE **bʰ* yielded Thr. *p*, giving Rom. *jupân* 'master' and *stǎpân* 'master, owner' as examples. However, *jupân* is not of Thracian but of eastern origin, related to Bulgar *zoapan* (Blažek 2006: 515) and Old Uyghur *čopan* 'chancellor' (Ramstedt 1951: 72), borrowed into Romanian via Proto-Slavic (Leschber 2021: 213). Cf. further related official titles: Chinese *chù bàn* (处半), Bactrian *sōpano*, Khotanese *chaupaṃ* (Zhan 2014: 77). Regarding *stǎpân*, Philippide (1927: 14–15) compares it to ARo. *stǎpuescu* 'I stop, I master myself' and further to Sanskrit *sthāpayati* 'to make to stand', implying a derivation from PIE **steh*$_2$- 'to stand' and an original meaning of *'one who makes (things, people) stay put', The **-p-* suffix is unclear, but a hypothetical earlier **-bʰ-* form is unevidenced. See a discussion of the origin of the possibly related Indic *-páya-* causative suffix in Khoshsirat and Byrd (2023).

[25] A variant of OHG *gruoba* 'pit, cavity' (Schade 1872–1882: 355) < PGm. **grōbō* 'hole, pit', an ō-stem derived from PGm. **graban* 'to dig' < PIE **gʰrebʰ-* id. (Kroonen 2013: 191; 185–186).

chronological difficulties. No loanwords in Albanian showing the *-*b*- > *-*p* outcome have been identified so far: the /b/ is preserved in Slavic loans *kob* 'treachery', *rob* 'serf', *cub*[1] 'short-tailed; with the tail cut off', *udob* 'easy, convenient' and even Gothic loan *cub*[2] 'robber' (Hyllested 2016: 67). If *gropë* is indeed of Slavic origin, the borrowing must have taken place very early. One may surmise that early Albanians borrowed *gropë* from Proto-Slavic just before the Alb. *-*b*- > *-*p* shift completed, and shortly thereafter borrowed *cub* from remnant Germanic populations of the Balkans. However, the earlier form of PSl. **gròbъ* 'grave' was **grabu* (Klotz 2017: 112), and the PSl. **a*, **u* > **o*, *ъ shifts only occurred after 600 AD (Leschber 2021: 208). Unless the borrowing of *cub*[2] from Gothic was as late as the 7th century AD, relative chronology demands that we start from **grabu*. This, in turn, would require an unevidenced early PSl. **a* > Alb. *o* outcome.

These difficulties suggest the need to move beyond the traditional hypotheses. The final-syllable *-*b*- > *-*p* shift has a notable Balkan parallel. The name of *Astibus*,[26] a town of Paeonia, an ancient region centered on the middle reaches of the Vardar (Axios) River (Duridanov 1985: 133), has been preserved in Macedonian as *Štip*. Displaying the loss of anlaut unstressed *a* characteristic of Albanian, the modern name has been regarded as transmitted by Albanian (Barić 1955: 49). While its *b* > *p* shift may have occurred in Albanian, an alternative possibility is that it occurred already in Paeonian[27] and the toponym was passed to Albanian as **Astip*-. This circumstance warrants consideration of the possibility that *groapă* originated in Paeonian.

Among the distinguishing features of Paeonian proposed by Duridanov (1985: 134–135) we remark the one needed for explaining *groapă*: the preservation of PIE **o*, based on the Paeonian hydronym *Pontos* (< PIE **pent-* 'find a path' through the o-grade, related to Gr. *póntos* 'sea' (Янакиева 2009: 112)) and ethnonym *Stoboi*. The preservation of PIE **o* is not exclusive to Paeonian among the eastern Paleo-Balkan languages, however: it also characterizes Brygian, another variety historically spoken near the Axios River (Duridanov 1993: 64–65).

Therefore, Rom. *groapă* and Alb. *gropë* continue a Paeonian or Brygian reflex of $*g^{h}rob^{h}$-eh_2. Most likely, Albanian mediated the Paeonian or Brygian word to Eastern Romance, as suggested by the Albanian transmission of *Astibus*.

2.4 Rom. *mărar* 'dill'

Long linked with Gr. *márathron* 'fennel' and Alb. *maraj*/*mërajë* id. (Capidan 1922: 458), this word is usually regarded as substratal, with no etymological solution provided (Boerescu 2018: 52; Brâncuș 1983: 96–97; Poghirc 1969: 345). The Greek word was attested in Mycenaean as *ma-ra-tu-wo*, and may have been borrowed from Pre-Greek (Beekes 2010, 2: 903–904).

Diculescu (1927: 460) proposes a borrowing from Gr. **maráthrion*, a reconstructed diminutive of *márathron*. However, the Gr. *th* > Rom. ∅ outcome would be unexplained.

Orel (1998: 244–245) regards *mărar* as borrowed from Alb. **mararja*, an earlier form of *maraj*. This is a plausible viewpoint; for Alb. *rj* > *j*, see Schumacher (2013: 247). The obsolete Rom. *mărariu* 'dill' (Boerescu 2018: 52) and ARo. *mărăl'u* 'dill' (Capidan 1922: 458) buttress this hypothesis: the Alb. *rj* > *j* sound change proceeded with a *ʎ* intermediary that survives in some Tosk varieties (Schumacher 2013: 247). Thus, *mărar* and *mărăl'u* were borrowed from Albanian, with the former an earlier loan than the latter.

Next, we will elucidate the link between *maraj* and *márathron*. First, the reconstructed diminutive **maráthrion* is necessary to account for the **rj* in the history of the Albanian word.

[26] Derived by Georgiev (1981: 172) from $*ad\text{-}tib^{h}os$ 'at the marsh'.

[27] Paeonia was classed as a distinct ethnic entity from Thrace, Illyria and Macedonia by some ancient writers, though conflicting reports describe Paeonians variously as Thracians, Phrygians or Illyrians (Katičić 1976: 117).

The loss of **th* can be explained through a **drj* > **rj* > *j* development, well-known in Albanian,[28] entailing an earlier **maradrjV-*, and an Alb. **d* ~ Gr. *th* correspondence. To account for it, Huld (1986: 247–249) proposes that *maraj* was borrowed from a Macedonian cognate of **maráthrion*,[29] putting forward two other loanwords in Albanian with a putative Macedonian treatment of PIE voiced stops:

— Alb. *mëllagë* 'mallow' ~ AGr. *malákhē* id.

— Alb. *bredh* 'fir tree' ~ AGr. *bráthu* 'savin'.

But these examples fail to evidence an original voiced aspirate. AGr. *malákhē* can be compared with Semitic words like Hebrew *malúakh* 'saltbush',[30] and the Albanian and Ancient Greek words may be independent non-Indo-European loans. Regarding AGr. *bráthu*, comparanda such as Aramaic *bərāṯ* 'cypress' and Assyrian *burāšu* id.[31] do not indicate an original **d*^{h}. Furthermore, Alb. *bredh* is rather a native word from PIE **b*^{h}*erHǵ-o-* 'birch' (Demiraj 1997: 107–108), and therefore it must be separated from *bráthu*. Thus, another explanation for the Gr. *th* ~ Alb. **d* correspondence is needed.

It may be more appropriate to start from **maráthrion* rather than from a postulated cognate of it. As fennel and dill are aromatic herbs — a category particularly susceptible to lexical borrowing through trade relations — the transmission of Gr. **maráthrion* to Alb. **maradrjV-* may have occurred through an intermediary language.

A plausible intermediary language can be identified. In Thracian, voiceless stops were voiced dialectally when adjacent to **r* or **n* (Duridanov 1985: 117–118; Georgiev 1983: 1171–1172). Remarkably, evidence from Romanian confirms this development. In Rom. *Medru* 'Demetrius' and *Sumedru* 'Saint Demetrius', the *tr* > *dr* shift, thought unexplained (Ivănescu 1980: 242) can be attributed to Thracian mediation. Regarding the specifics of these dialectal Thracian developments, Duridanov (1985: 118) proposes Thr. **t*^{h}*n* > **dn* and **nk*^{h} > **ng* shifts. However, it seems more typologically plausible that voiceless aspirates conditionally evolved to voiced aspirates, subsequently represented as plain voiced stops in the Greek and Latin alphabets as these lacked dedicated symbols to render these sounds. While this issue bears no direct relevance to Romanian etymology as Romanian does not distinguish voiced aspirates, we will opt to represent the Thracian sounds in question here as voiced aspirates.

We propose, thus, that a Thr. **marad*^{h}*rjV-* was borrowed from Gr. **maráthrion* and subsequently reached Albanian, where it yielded **maradrjV-*. A later version of **maradrjV-* subsequently lent Eastern Romance *mărar* and *mărar'u* before evolving to *maraj/mërajë*.

2.5 Rom. *rangă* 'crowbar'

This word has generally been considered unexplained (see MDA (2001–2003, 4: 124)). Vinereanu (2023, 2: 365) tentatively suggests a formally untenable derivation from PIE **reik-* ('to stretch up, to wobble; pole, bar' (Pokorny 1959: 862)).

Semantically, a derivation from PIE **wronk-eh*$_2$ (also reflected in PSl. **rǫkà* 'hand, arm' (Greenberg 2017: 529) and Late Lat. *branca* 'paw', a loanword (Pokorny 1959: 1155)) is quite satisfactory. The PIE noun derives from a verbal root **wrenk-* preserved in Lithuanian *riñkti* 'to gather' (Derksen 2008: 440).[32]

[28] See Hamp (1976: 34).

[29] PIE **b*^{h}, **d*^{h}, **g*^{h} > AGr. *p*^{h}, *t*^{h}, *k*^{h}, AMac. *b*, *d*, *g* (Georgiev 1981: 167–169).

[30] Beekes (2010, 2: 896) reconstructs a **mal*^{w}*ak-* pre-form.

[31] See Beekes (2010, 1: 234–235).

[32] Galician *ranco* 'a tool used to clean ashes and embers' appears to be another substrate word in a Romance language derived from**wrenk-*, paralleling *rangă* in its semantic evolution.

The Romanian word has a lookalike in Alb. *rangë* 'housework', plausibly derived from the same **wronk-eh₂* by Orel (1998: 365). Due to the wide semantic gap, Albanian cannot be the source of Rom. *rangă*, however.

Instead, the Thracian voicing of stops adjacent to **n* noted in the discussion of *mărar*, together with PIE **o* > Thr. **a*,[33] suggests Thracian provenance. While the attested Thracian material provides no reliable examples of the treatment of PIE **wr*,[34] Romanian evidence supports **r* as the Thracian reflex of PIE **wr*. Vinereanu (2023, 2: 380–381) derives the otherwise unexplained Rom. *rătie* 'twig; brushwood' and *rătină* 'small branch' from PIE **wreh₂d-* ('root', but perhaps also 'branch', as suggested by the Greek and Latin data (de Vaan 2008: 512–513)). Based on the implied PIE **d* > *t* outcome, attributable to Thracian (Duridanov 1985: 137; Georgiev 1981: 128),[35] *rătie* and *rătină* can be plausibly identified as Thracian loans.

Finally, we add that the preserved /a/ in *rangă* shows that the borrowing from Thracian occurred no earlier than the 6th century AD. This is because the Rom. *an* > *ɨn* shift, characteristic of native words and very early loanwords, also surfaces in Rom. *jupân* 'master', an early Slavic loan. The fact that Thracian still influenced Romanian at this relatively late date can be explained through Thracian's resistance to Latinization and Hellenization. As inferred from toponymic evidence, only a limited part of the Thracian population had adopted Latin or Greek at the time of the Slavic migration to the Balkans (Yanakieva 2018: 60–61).

2.6. Rom. *urdă* 'a kind of cheese similar to ricotta'

Exploiting the fact that this culinary product is prepared by boiling, derivations from PIE **werH-* 'to be hot' (LIV 2001: 689) have been proposed for *urdă* and Alb. *urdhë* id. (Orel 1998: 487–488; Russu 1981: 412). Semantically, this is plausible, but the formal aspect needs elucidation.

If assuming an Albanian origin, the suffix has to be identified with Alb. *-dhë*, but this is not comfortable, because the suffix only appears in a few words (see Ajeti (2017: 531–532)) and does not generally attach to verbal roots. A stronger objection to this scenario is that the *ur-* reflex cannot be explained through Albanian if starting from **werH-*. Orel (1998: 487) assumes either a zero-grade or an o-grade formation, but neither option can explain *urdhë*. PIE **C̥rHT* yielded Albanian C*raT*, while the normal result of PIE **o* is Alb. *a*, with *u* appearing as an outcome after *k*, *g* and *h* (Schumacher 2013: 211, 213, 230).

These difficulties prompt us to explore derivations from **werH-* through a pre-Roman language other than Albanian. The *ur-* appears to reflect a zero-grade derivative. In this regard, Thracian offers an option, as its reflex of PIE **r̥* was **ur/or*, distinguishing it from Dacian, whose reflex was **ri* (Georgiev 1983: 1189). Moreover, laryngeals following a syllabic resonant

[33] See Georgiev (1983: 1165).

[34] Mihaylova and Mircheva (2022: 210) plausibly argue that Thr. *briza* 'rye', comparable to Sanskrit *vrīhí* 'rice', is a Wanderwort bearing little significance to questions of Thracian historical phonetics. It has been claimed, though, that Thr. *bria* 'city' is related to Tocharian A *ri* id. and Tocharian B *riye*, entailing a **wriyā* proto-form (Duridanov 1985: 107) and a PIE **wr* > Thr. *br/vr* shift. But the apparent Thracian-Tocharian isogloss is isolated, and the Tocharian words admit another etymology: Lubotsky and Starostin (2003: 264) propose a Chinese origin ultimately from Proto-Sino-Tibetan **riəH*, whence also Jingpo *mǝre* 'town'.

[35] While the Thracian devoicing of the PIE mediae remains disputed (Yanakieva 2018: 47–48), some of the proposed examples have no alternative explanation. A well-known case is Thracian *sinupyla* 'white briony', contrasting with Dacian *kinouboila*[35] id. and Lithuanian *šùnobelė* 'crabapple', from a pre-form meaning 'dog's apple' (Duridanov 1995: 48). Further examples have been advanced recently: Mihaylova and Mircheva (2022: 215–216) compare Thr. *torelle* 'funeral song' with AGr. *odŭromai* 'to wail, lament'.

elided before a consonant in Thracian, as shown from the onomastic element *-turma*, continuing **dr̥h₂-mo-* < PIE **derh₂-* 'to run' (Mircheva and Mihaylova 2020: 57), while the PIE **wr* > Thr. **r* shift discussed for *rangă* suggests an analogous PIE **w* > Thr. ∅ / _ **r̥* development.

The above evidence establishes **ur-* as the probable Thracian reflex of PIE **wr̥H-*, making the language the likely source of *urdă* and *urdhë*. An additional factor speaking in favor of a Thracian origin is the suffix. Because of the Thracian voicing of stops near **r* also noted earlier, introducing a **-dʰ-* suffix is unnecessary. Instead, the pre-form of the word can be reconstructed as a participle formation in **-teh₂*: **wr̥H-teh₂*. Semantically, the formation is very close to Italian *ricotta*, etymologically 'recooked'. Subsequently, the initial Thracian result of **wr̥H-teh₂*, **urtʰā*, yielded **urdʰā*, which accounts for the discussed Romanian and Albanian words.

Regarding the pathway of borrowing, since Thracian is known to have significantly influenced Eastern Romance[36] but not Albanian, the Thracian word was most probably mediated to Albanian by Balkan Latin.

3. Summary and conclusions

We have provided novel etymologies for six Romanian substrate words with Albanian comparanda: *berc*, *driză*, *groapă*, *mărar*, *rangă* and *urdă*. The first two words are identifiable as Albanian loans. In *groapă*, Romanian preserves a relic from a minor Paleo-Balkan language — Paeonian or Brygian — likely acquired via Albanian. A chain borrowing scenario explains *mărar*, a word of Ancient Greek origin mediated to Romanian by Thracian and Albanian. The preserved /an/ in *rangă*, a Thracian loanword, imposes a 6th-century AD terminus post quem for the end of Thracian-Balkan Latin contacts. Finally, *urdă* is also analyzable as a Thracian loan, with the Romanian word being the most probable source of Alb. *urdhë*.

In addition to providing new solutions to problems of Romanian etymology, our work provides insights on extinct Paleo-Balkan languages — most notably, evidencing a PIE **w* > Thr. ∅ / _{**r*, **r̥*} shift. This result illustrates how data from surviving languages of Southeastern Europe can reveal phonological features of the sparsely attested extinct languages once spoken in the same area. Romanian, in particular, holds great promise in this regard. While we have chosen to limit the scope of this paper to substrate words with Albanian comparanda, Romanian may offer a considerably richer substrate corpus, given that the language reportedly contains over 1,200 words of unknown etymology according to recent counts (Paliga 2022: 235). In sum, we maintain that treating the Romanian substrate as a composite of distinct layers not only offers the most suitable framework for advancing its understanding, but also helps uncover features of Paleo-Balkan languages that would be inaccessible through the ancient attested data alone.

Abbreviations and symbols

AGr. — Ancient Greek; Alb. — Albanian; AMac. — Ancient Macedonian; ARo. — Aromanian; C — consonant; Gr. — Greek; H — Proto-Indo-European laryngeal; Hun. — Hungarian; Lat. — Latin; OCS — Old Church Slavonic; OHG — Old High German; PGm. — Proto-Germanic; PIE — Proto-Indo-European; PSl. — Proto-Slavic; Rom. — Romanian; T — plosive; Thr. — Thracian; V — vowel.

[36] An example of this influence is the Rom. *-esc-* suffix, one of the most important Romanian suffixes, which continues Thr. *-isk-* (Graur 1927: 549–551).

References

Ajeti, Idriz. 2017. Formimi i fjalëve dhe pasurimi i gjuhës shqipe. In: Rexhep Ismajli (ed.). *Studime për gjuhën shqipe*: 511–539. Prishtinë: Akademia e Shkencave dhe e Arteve e Kosovës.

Arapi, Ina. 2012. Albanisch-rumänische Parallelen im Bereich der Wortbildung: Die Suffixe. In: Wolfgang Dahmen, Günter Holtus, Johannes Kramer, Michael Metzeltin, Wolfgang Schweickard, Otto Winkelmann (eds.). *Südosteuropäische Romania. Siedlungs-/Migrationsgeschichte und Sprachtypologie. Romanistisches Kolloquium XXV*: 123–152. Tübingen: Narr Francke Attempto Verlag.

Barić, Henrik. 1955. *Hymje në historin e gjuhës shqipe*. Prishtinë: Shtëpija botuese "Mustafa Bakija".

Beekes, Robert S. P. 2010. *Etymological dictionary of Greek*. 2 vols. Leiden / Boston: Brill.

Blažek, Václav. 2006. Bohumil Vykypěl, Studie k šlechtickým titulům v germánských, slovanských a baltských jazycích. *Baltistica* 41(3): 511–518. DOI: https://doi.org/10.15388/baltistica.41.3.1363

Boerescu, Pârvu. 2018. *Elemente de substrat (autohtone) ale limbii române: — Compendiu –*. București: Editura Academiei Române.

Bogrea, Vasile. 1921. Elemente latine. *Dacoromania. Buletinul „Muzeului Limbei Române"* 1: 257–260.

Brâncuș, Grigore. 1983. *Vocabularul autohton al limbii române*. București: Editura Științifică și Enciclopedică.

Brâncuș, Grigore. 2005. *Introducere în istoria limbii române*. București: Editura Fundaţiei "România de Mâine".

Brâncuș, Grigore. 2013. Problemele substratului limbii române. *Limba Română* 4: 417–423.

Brâncuș, Grigore. 2022. Definirea substratului limbii române. *Limba Română* 32(3–4): 46–57.

Çabej, Eqrem. 1964. Studime rreth etimologjisë së gjuhës Shqipe XII: K. *Studime Filologjike* 18(1): 61–98.

Candrea, Ion Aurel, Ovid Densusianu. 1907. *Dicționarul etimologic al limbii romîne. Elementele latine. 2. Cere — fin*. București: Librăria Socec & Comp.

Capidan, Theodor. 1922. Raporturile albano-române. *Dacoromania. Buletinul „Muzeului Limbei Române"* 2: 444–553.

Cihac, Alexandru. 1879. *Dictionnaire d'étymologie daco-romane. Éléments slaves, magyars, turcs, grecs-moderne et albanais*. Francfort: Ludolphe St-Goar.

Ciorănescu, Alexandru. 2001. *Dicţionarul etimologic al limbii române*. București: Saeculum I. O.

de Vaan, Michiel. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden / Boston: Brill.

DELR = Academia Română. 2025. Dicționarul etimologic al limbii române. Available online at: https://delr.lingv.ro/

Demiraj, Bardhyl. 1997. *Albanische Etymologien (Untersuchungen zum albanischen Erbwortschatz)*. Amsterdam / Atlanta: Rodopi.

Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden / Boston: Brill.

Diculescu, Constantin C. 1927. Elemente vechi grecești din limba română. *Dacoromania. Buletinul „Muzeului Limbei Române"* 4(2): 393–516.

DLR = Academia Română. 1913–1949. *Dicționarul limbii române. 2 vols*. București: Librăriile Socec & Comp. și C. Sfetea / Imprimeria Națională / Tipografia Ziarului „Universul".

Duridanov, Ivan. 1985. *Die Sprache der Thraker: Vom Verfasser übersetzt und wesentlich erweitert*. Sofia: Hieronymus Verlag Neuried.

Duridanov, Ivan. 1993. Brygische Spuren im thrakischen Sprachgebiet. *ORPHEUS. Journal of Indo-European and Thracian Studies* 3: 63–68.

Duridanov, Ivan. 1995. Thrakisch — Dakisch — Pelasgisch. In: Michaela Ofitsch, Christian Zinko (eds.). *Studia onomastica et indogermanica: Festschrift für Fritz Lochner von Hüttenbach zum 65*: 45–57. Graz: Leykam Buchverlagsgesellschaft.

Duridanov, Ivan. 2000. Die Rolle der Paläobalkanistik für die Südosteuropa-Linguistik. In: Uwe Hinrichs, Uwe Büttner (eds.). *Die Südosteuropa-Wissenschaften im neuen Jahrhundert. Akten der Tagung vom 16.–19. Oktober 1999 an der Universität Leipzig*: 23–32. Wiesbaden: Harrassowitz Verlag.

Friedman, Victor A., Brian D. Joseph. 2025. *The Balkan languages*. Cambridge University Press.

Georgiev, Vladimir I. 1965. Le dace comme substrat de la langue roumaine. *Revue roumaine de linguistique* 10(1–3): 75–80.

Georgiev, Vladimir I. 1981. *Introduction to the history of the Indo-European languages*. Sofia: Publishing House of the Bulgarian Academy of Sciences.

Georgiev, Vladimir I. 1983. Thrakisch und Dakisch. In: Wolfgang Haase (ed.). *Aufstieg und Niedergang der römischen Welt: Geschichte und Kultur Roms im Spiegel der neueren Forschung: II*: 1148–1194. Berlin / New York: Walter de Gruyter.

Graur, Alexandru. 1927. Le suffixe roumain *-escu* et le suffixe thrace *-isk*. *Romania* 53: 539–552.

Greenberg, Marc L. 2017. Slavic. In: Mate Kapović (ed.). *The Indo-European languages. Second edition*: 519–551. London / New York: Routledge.

Hamp, Eric P. 1976. Some Romanian areal etymologies. *Studii și cercetări lingvistice* 27(1): 33–36.

Hill, Peter M. 2011. Language contact in South-Eastern Europe. In: Bernd Kortmann, Johan van der Auwera (eds.). *The languages and linguistics of Europe. A comprehensive guide*: 409–428. Berlin / Boston: Mouton De Gruyter.

Hubschmid, Johannes. 1983. Zum Substrat und zur Vorgeschichte des Rumänischen (Probleme der Balkanlinguistik). *Zeitschrift für romanische Philologie* 99(5–6): 497–511. DOI: https://doi.org/10.1515/zrph.1983.99.5-6.497

Huld, Martin E. 1986. Accentual stratification of Ancient Greek loanwords in Albanian. *Zeitschrift für vergleichende Sprachforschung* 99(2): 245–253.

Hyllested, Adam. 2016. Proto-Indo-European reconstruction and Albanian phonotactics. In: Stephanie W. Jamison, H. Craig Melchert, Brent Vine, Angelo Mercado (eds.). *Proceedings of the 26th Annual UCLA Indo-European Conference: Los Angeles, October 24th and 25th, 2014*: 63–81. Bremen: Hempen.

Ivănescu, Gheorghe. 1980. *Istoria limbii române*. Iași: Junimea.

Kálmán, Béla. 1972. Hungarian historical phonology. In: Loránd Benkő, Samu Imre (eds.). *The Hungarian language*: 49–84. The Hague / Paris: Mouton.

Kaluzhskaya, Irina A. 1993. Paleobalkan linguistic heritage and the languages of Balkan-Carpathian area. *ORPHEUS. Journal of Indo-European and Thracian Studies* 3: 69–72.

Katičić, Radoslav. 1976. *Ancient languages of the Balkans: Part one*. The Hague / Paris: Mouton.

Khoshsirat, Zia, Andrew Miles Byrd. 2023. The Indo-Iranian labial-extended causative suffix. *Indo-European Linguistics* 11(1): 64–95. DOI: https://doi.org/10.1163/22125892-bja10025

Klotz, Emanuel. 2017. *Urslawisches Wörterbuch*. Wien: Facultas.

Ködderitzsch, Rolf. 2001. Albanisch und Thrakisch. *Studia Albanica* 34(1): 79–87.

Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden / Boston: Brill.

Leschber, Corinna. 2021. Urslavische Lexik in Rumänischen. In: Anna Jouravel, Audrey Mathys (eds.). *Wort- und Formenvielfalt: Festschrift für Christoph Koch zum 80. Geburtstag*: 207–218. Berlin: Peter Lang.

LIV = Rix, Helmut, Martin Joachim Kümmel, Thomas Zehnder, Reiner Lipp, Brigitte Schirmer. 2001. *Lexikon der indogermanischen Verben*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Lubotsky, Alexander, Sergei Starostin. 2003. Turkic and Chinese loan words in Tocharian. In: Brigitte L. M. Bauer, Georges-Jean Pinault (eds.). *Language in time and space. A Festschrift for Werner Winter on the occasion of his 80th birthday*: 257–270. Berlin / New York: Mouton de Gruyter.

Maiden, Martin, Adina Dragomirescu, Oana Uță Bărbulescu, Gabriela Pană Dindelegan, Rodica Zafiu. 2020. *The Oxford history of Romanian morphology*. New York: Oxford University Press.

Matzinger, Joachim. 2006. *Der altalbanische Text Mbsuame e kreshtere (Dottrina cristiana) des Leke Matrenga von 1592. Eine Einführung in die albanische Sprachwissenschaft*. Dettelbach: Verlag J. H. Röll GmbH.

Matzinger, Joachim. 2017. L'elemento latino della lingua albanese — un impatto della Via Egnatia? *Palaver* 6(2): 29–60.

Matzinger, Joachim. 2018. 98. The lexicon of Albanian. In: Jared Klein, Brian D. Joseph, Matthias Fritz, Mark Wenthe (eds.). *Handbook of comparative and historical Indo-European linguistics. HSK 41.3*: 1788–1800. Berlin / Boston: Walter de Gruyter GmbH.

MDA = Academia Română. 2001–2003. *Micul dicționar academic*. 4 vols. București: Editura Univers Enciclopedic.

Meyer, Gustav. 1891. *Etymologisches Wörterbuch der albanesischen Sprache*. Strassburg: Verlag von Karl J. Trübner.

Mihaylova, Bilyana, Albena Mircheva. 2022. The Thracian glosses revisited. In: Peter Delev et al. (eds.). *Ancient Thrace: Myth and reality. The proceedings of the Thirteenth International Congress of Thracology. Kazanlak, September 3 — 7, 2017*. Vol. 2: 209–217. Sofia: St. Kliment Ohridski University Press.

Miklosich, Franz. 1861. *Die slavischen Elemente im Rumunischen*. Wien: Kaiserlich-königliche Hof- und Staatsdruckerei.

Mircheva, Albena, Bilyana Mihaylova. 2020. The reflexes of the Indo-European laryngeals in the Paleo-Balkan languages. *Linguistique balkanique* 59(1): 46–61.

Nagy, Levente. 2003. Balcanistica hungarica rediviva. *Zeitschrift für Balkanologie* 39(2): 191–200.

Neroznak, V. P. 1978. *Paleobalkanskie jazyki*. Moskva: Nauka.

Orel, Vladimir E. 1998. *Albanian etymological dictionary*. Leiden / Boston / Köln: Brill.

Paliga, Sorin. 2022. Thracian v. Dacian (Daco-Moesian). How many 'Thracian languages' were there spoken in the Antiquity? In: Peter Delev et al. (eds.). *Ancient Thrace: myth and reality. The proceedings of the thirteenth interna-*

tional congress of Thracology. Kazanlak, September 3 — 7, 2017. *Vol. 2*: 231–238. Sofia: St. Kliment Ohridski University Press.

Paliga, Sorin. 2024. *An etymological dictionary of the Romanian language*. New York: Peter Lang.

Petrovici, Emil. 1948. Note slavo-române. III. *Dacoromania. Buletinul „Muzeului Limbei Române"* 11: 184–193.

Philippide, Alexandru. 1927. *Originea românilor: Volumul al doilea: Ce spun limbile romînă şi albaneză*. Iași: Tipografiea "Viaţa Romînească" S. A.

Poghirc, Cicerone. 1969. Influenţa autohtonă. In: Ion Coteanu, Gh. Bolocan, Matilda Caragiu Marioţeanu (eds.). *Istoria limbii române. 2*: 313–365. București: Editura Academiei Republicii Populare Române.

Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. *Vol. 1*. Bern / München: Francke Verlag.

Pușcariu, Sextil, Matteo Bartoli, Alois Belulovici, Arthur Byhan. 1926. *Studii istroromâne. 2. Introducere. Gramatică. Caracterizarea dialectului istroromân*. București: Cultura Naţională.

Ramstedt, Gustaf John. 1951. Alte türkische und mongolische Titel. *Journal de la Société Finno-ougrienne* 105(2): 59–82.

Rusakov, Alexander. 2017. Albanian. In: Mate Kapović (ed.). *The Indo-European languages. Second edition*: 552–608. London / New York: Routledge.

Rusakov, Alexander. 2018. 94. The documentation of Albanian. In: Jared Klein, Brian D. Joseph, Matthias Fritz, Mark Wenthe (eds.). *Handbook of comparative and historical Indo-European linguistics. HSK 41.3*: 1716–1731. Berlin / Boston: Walter de Gruyter GmbH.

Russu, Ioan I. 1981. *Etnogeneza românilor. Fondul autohton traco-dacic și componenta latino-romanică*. București: Editura Ştiinţifică şi Enciclopedică.

Șăineanu, Lazăr. 1929. *Dicţionar universal al limbei române*. București: Editura "Scrisul Românesc".

Saramandu, Nicolae. 1987. Concordanțe etimologice și lexicale româno-albaneze (bearcă, buș, 'pumn', urdori, cocobarză, corcoduș). *Studii și cercetări lingvistice* 38(2): 126–141.

Schade, Oskar. 1872–1882. *Altdeutches Wörterbuch: 1*. Halle: Verlag der Buchhandlung des Waisenhauses zu Halle.

Schumacher, Stefan. 2009. Lehnbeziehungen zwischen Protoalbanisch und balkanischem Latein bzw. Romanisch. In: Oliver Jens Schmitt, Eva Anne Frantz (eds.). *Albanische Geschichte: Stand und Perspektiven der Forschung*: 37–59. München: Oldenbourg Verlag.

Schumacher, Stefan. 2013. Historische Phonologie. In: Stefan Schumacher, Joachim Matzinger (eds.). *Die Verben des Altalbanischen: Belegwörterbuch, Vorgeschichte und Etymologie*. Wiesbaden: Harrassowitz Verlag.

Schütz, István. 2002. *Fehér foltok a Balkánon: Bevezetés az albanológiába és a balkanisztikába*. Budapest: Balassi Kiadó.

Scriban, August. 1939. *Dicționaru limbii românești. Etimologii, înţelesuri, exemple, citaţiuni, arhaizme, neologizme, provincializme*. Iași: Institutu de Arte Grafice "Presa Bună".

Slușanschi, Dan. 1981. Substratul daco-moesian al limbii române. *Limba Română* 30(4): 339–342.

Sutherland, Mhairi A., Cassandra B. Tucker. 2011. The long and short of it: A review of tail docking in farm animals. *Applied Animal Behaviour Science* 135(2): 179–191. DOI: https://doi.org/10.1016/j.applanim.2011.10.015

Tiktin, Heimann Hariton. 1903–1925. *Rumänisch-deutsches Wörterbuch*. *3 vols*. București: Imprimeria Statului.

Ungureanu, Dan. 2016. *Româna şi dialectele italiene*. București: Editura Academiei Române.

Vinereanu, Mihai. 2023. *Dicționarul etimologic al limbii române*. *2 vols*. București: Editura Uranus.

Witczak, Krzysztof Tomasz. 1995. Were the Bessans ancestors of the Albanians? A new opinion on the ethnogenesis of the Albanian nation. *Thraco-dacica* 16(1–2): 309–312.

Yanakieva, Svetlana. 2009. *Trakiyska hidronimiya*. Sofiya: Akademichno izdatelstvo "Prof. Marin Drinov".

Yanakieva, Svetlana. 2018. The Thracian language. *ORPHEUS. Journal of Indo-European and Thracian Studies* 25: 26–68.

Zaicz, Gábor. 2006. *Etimológiai szótár. Magyar szavak és toldalékok eredete*. Budapest: TINTA Könyvkiadó.

Zhan, Zhang. 2014. Secular Khotanese documents and the administrative system in Khotan. *Bulletin of the Asia Institute* 28: 57–98.

Андрей Ф. Стойка, Сюэ К. Р. Стойка. О нескольких румынских субстратных словах с албанскими соответствиями: этимология и импликации

В статье предлагается новый взгляд на этимологии некоторых румынских субстратных слов, имеющих близкие албанские соответствия, с упором на определение языка-источника для каждого слова. Выявлены два новых албанских заимствования: *berc* 'с ко-

ротким хвостом; бесхвостый' и *drizǎ* 'мохнатый пырей', тогда как *rangǎ* 'лом' и *urdǎ* 'вид сыра, похожий на рикотту' определяются как фракийские заимствования. Помимо этого, для *groapǎ* 'яма' (< *$g^{h}rob^{h}$-eh_2) предполагается происхождение от субстрата, сохранившего праиндоевропейское **o* (пеонийского или бригийского). Наконец, *mărar* 'укроп' выводится в конечном счете из древнегреческого **maráthrion* (диминутив от *márathron* 'фенхель'), через посредничество как фракийского, так и албанского.

Ключевые слова: румынский язык; румынская этимология; палеобалканские языки; албанский язык; фракийский язык.

А. В. Савельев
Институт языкознания РАН / НИУ ВШЭ, Москва; Институт истории и археологии УрО РАН, Екатеринбург;
a.savelyev@iling-ran.ru

О критериях выявления и стратификации булгаро-чувашских заимствований в пермских языках II[1]

Статья представляет собой вторую и заключительную часть исследования, посвящённого анализу тюркских заимствований булгаро-чувашского типа в пермских языках. В этой части предлагаются дополнительные фонетические критерии, характеризующие в первую очередь поздние (XIV–XV вв.) булгаризмы, прежде всего, в удмуртском языке, источник которых уместно определять как среднечувашский язык. С учётом того, что в первой части исследования был выделен слой ранних (VIII–IX вв.) заимствований из древнечувашского языка в прапермский, обсуждается механизм обновления заимствований в булгаро-пермской контактной лексике: рассматриваются, в частности, случаи сохранения старого заимствования в коми и замещения его новым чувашизмом в удмуртском, а также сохранения старого заимствования в удмуртском и замещения его заимствованием из русского языка в коми. Далее отдельно рассматривается вопрос о зонах поздних узколокальных контактов между среднечувашскими говорами и уже разошедшимися диалектами языков пермской группы и, в частности, о возможном присутствии чувашеязычного населения в бассейне р. Чепцы вплоть до XVII в. В этом контексте обсуждаются возможные следы позднего сепаратного чувашского влияния в бесермянском наречии, северных удмуртских, южных коми-зырянских, а также чепецко-татарских говорах. В финальной части исследования, после обсуждения прочих идентифицируемых пластов булгаризмов в пермских, рассматривается особая группа заимствований, широко распространённых как в удмуртских, так и в коми диалектах (в т. ч. за пределами южнокоми ареала) и демонстрирующих нетривиальные коми-удмуртские соответствия. Этот пласт заимствований предлагается датировать примерно X–XIII вв., его источник может быть определён как поздний древнечувашский, а на роль реципиента претендует находившийся в процессе распада общепермский диалектный континуум. В целом предлагаемая реконструкция указывает на практически непрерывные булгаро-пермские языковые связи с начала переселения булгар в Волго-Камский регион в самом конце VII в. и до XVI–XVII вв., когда чувашское население Арской стороны постепенно ассимилировалось среди татар и вероятный чувашский этноязыковой компонент на Чепце также растворился среди окружающих народов.

Ключевые слова: пермские языки; удмуртский язык; коми язык; булгаро-чувашские заимствования; чувашский язык; татарский язык; языковые контакты; стратификация заимствований; историческая фонетика; этимология.

[1] Статья написана в рамках работы по гранту Российского научного фонда № 24-18-00148 «Этимологический словарь удмуртского языка (славянские, тюркские и среднеиранские заимствования)» (https://rscf.ru/project/24-18-00148/). Support from the Russian Science Foundation under grant No. 24-18-00148 (https://rscf.ru/en/project/24-18-00148/) is gratefully acknowledged. Автор выражает признательность А. В. Дыбо, М. А. Живлову, П. С. Злобину, Р. И. Идрисову, К. Калверу, А. Л. Мусихину, Н. С. Мясникову, В. В. Напольских за ценные замечания как общего, так и частного характера.

1. Ещё раз о формальных признаках булгаризмов в пермских языках: адаптация среднечувашских заимствований в удмуртском и смежные вопросы

Некоторые звуковые переходы, отражающиеся в волжско-булгарском ареале, подействовали заведомо после начала булгаро-пермских (точнее, древнечувашско-прапермских) контактов, наиболее ранний пласт которых датируется VIII–IX вв. (см. первую статью цикла — Савельев 2025: 154–161). Таким образом, в отличие от 15 критериев, рассмотренных в первой части (там же: 139–147), эти признаки не всегда являются диагностическими с точки зрения строгого разграничения булгарских и кыпчакских заимствований (булгарское состояние до действия перехода может совпадать с кыпчакским). Однако они позволяют идентифицировать в корпусе булгаризмов пермских языков (прежде всего, в удмуртском) поздние заимствования. Предварительно датировать основной пласт этих заимствований можно XIV–XV вв., а их источник уместно называть среднечувашским языком; речь идёт о состоянии, промежуточном между древнечувашским (примерно до XIII в.) и современным чувашским (с XVII–XVIII вв. и далее). Из этого же среднечувашского состояния было заимствовано огромное количество лексики (более 1700 слов) в марийский язык, материал которого для наглядности также включается в нижеследующую роспись этимологий.

Консонантизм (дополнительные критерии)

(xvi)[2] ПТ **ć* > булг. **ć* > др.-чув. **ć* > позднедр.-чув., ср.-чув. **ś*[3] [→ удм. *ś*]. Критерий № 10 по Ю. Вихманну (Wichmann 1903: 3–4) и А. Рона-Ташу (Róna-Tas 1988: 764). Внутри корпуса булгаризмов пермских языков случаи с такой рефлексацией противопоставлены случаю др.-чув. **ćömälə* → ППерм (**ćomal'ʲ* >) **ćumal'ʲ* ‘копна’ (см. Савельев 2025: 151, 158) и рассмотренным ниже в секции 2 случаям др.-чув. **ćårla* → ППерм **ćårla* ‘серп’, др.-чув. **bårćə̂n* → ППерм **bår(t)ćin* ‘шёлк’, др.-чув. **keći* ~ **kećə* → ППерм **kɛ/ȩć* ‘коза’. В татаризмах удмуртского языка предковый **ć* стандартно отражается как *ć* (хотя среди локальных заимствований из чепецко-татарского отмечаются и заимствования с *ś* — см. их обсуждение ниже под критерием xviii). Ср. отражение **ć* > *ś* в следующих булгаризмах удмуртского языка.

- Удм. (диал.) *juśi* ‘печаль, скорбь, грусть; жалостно’ ← ср.-чув. **jɔśś-i(j)* > чув. *jüśś-i* ‘горькое, горечь’, (метафорически) ‘неудача’. Чувашская форма — субстантивация от *jüźə* ‘кислый, горький’, перен. (!) ‘огорчительный’ < ПТ **iāćuγ* ‘горький; кислый’, откуда и тат. *ećə̂*, *äćə* ‘тж.’. Чувашское происхождение принято для удмуртского слова уже в Wichmann 1903: 66 — правда, источником у Ю. Вихманна считается несубстантивированная форма. Из среднечувашского источника также мар. **jɔśə̂* ‘трудный, тяжелый, тягостный’ (Itkonen 1970: 262).
- Удм. *śaldar* ‘края (у сковороды)’, (бес., см. Кузнецова и др. 2013: 412) ‘открытая полка (без стенок, крепится к стене и используется на кухне для хранения утвари)’ — к чув. *śat-ma* ‘сковорода’, откуда также и мар. **śälmä* > луг. *salma*, горн. *selmä*, СЗ *sälmä* ‘сковорода’. Насколько известно, для удмуртского слова чувашская этимология предлагается

[2] Ради целостности презентации фонетических критериев булгаризмов в пермских языках нумерация признаков продолжает начатую в (Савельев 2025: 139–147).

[3] Развитие **ć* > **ś* само по себе не является критерием разграничения древнечувашского и среднечувашского состояний: фрикативизация **ć* могла произойти уже в позднедревнечувашский период. Это уточнение важно с точки зрения интерпретации пласта заимствований, обсуждаемого ниже в секции 4.

впервые. Основа чувашского слова восходит к ср.-чув. **śärt-*, а заимствованные из чувашского финно-угорские формы — к, по-видимому, вторичному ср.-чув. варианту **śält-*. Как полагал уже М. Рясянен (VEWT: 105), в чувашском отражается ПТ **ćert-* ‘обрезать края, делать зарубки’. С этим можно согласиться: семантически сравнение беспроблемно, а аналогичное по распределению сосуществование вариантов **śVrt-* (> *śVt-*) на чувашской почве и **śVlt-mA* (> *sVlmA*) в чувашизме марийского языка находим в случае чув. *śĭrt-ma*, *śĭt-ma*, *śĭd-ak* ‘детские штанишки с прорехой’, (диал.: СЗ) *śə̑t-ma* ‘зимние штаны из толстого домашнего сукна’ и мар. (луг., горн.) *sə̑lma* ‘зимние штаны из домашнего сукна’[4]. Несомненного рефлекса ПТ **ćert-* в татарском нет; в EDAL: 435 таковыми считаются *ćirt-* ‘щёлкать’ и *ćärd-äk-lä-* ‘раздробить, размозжить’, но оба случая семантически не вполне убедительны. Ввиду невозможности объяснить фонетические субституции выглядит случайным сходство удмуртского слова с чув. *sändrä* ‘полати; полка’ и тат. *sändərä* ‘полати; нары’. Впрочем, в СЧЯ XI: 94 для чув. *sändrä* даётся в т. ч. значение ‘толстая доска… от *улчупи* [посудного шкафчика без дверок] к другой стене, куда ставят чашки’, которое практически тождественно значению *śaldar* в бесермянском. В этом наречии можно предполагать контаминацию двух чувашских по происхождению основ.

- Удм. (+бес.) *śaśka* ‘цветок’ ← ср.-чув. **śäśkä* (откуда регулярно чув. диал. *śaś-ka* при вторично упереднённой литер. форме *śäś-kä*) ‘цвет (растений); ботва’ (< ПТ **ćeć-* ‘распускаться, расцветать’). Ср. Wichmann 1903: 97; Róna-Tas 1988: 764. Из чувашского источника также мар. **śäśkä* ‘ягоды, плоды; цветы; ботва; ростки’. Ср. УТ **ćeć-ek* > тат. *ćäćäk* → чув. *ćăʒăk* ‘цветок’.
- Удм. *śetĭ* ‘украшение для косы, девичий накосник (из косточек)’ ← ср.-чув. **śẹw(ə)tə* (чув. диал. *śidə* ‘коса’ при литер. *śivət* ‘тж.’) < булг. **ćęgdü* ~ УТ **ćügde* ‘заушная кость ~ затылок’ (без рефлекса в татарском). Ср. Wichmann 1903: 99.
- Удм. (+бес.) *śulo* ‘кнут, бич’ ← ср.-чув. **śɔl-a* (чув. *śol-a* ‘плеть, кнут’), диминутив от **śɔlə̑* < ПТ **ćal-uγ*. Из чувашского источника также мар. **śɔla* ‘тж.’. Дальнейшие этимологические комментарии см. в (Савельев 2025: 143).
- Удм. *ukśo* ‘деньги’ ← ср.-чув. **ɔkśa* (чув. *okśa* ‘тж.’), что является продолжением ПТ **āq-ća* (белый-COLL) ‘деньги, монеты’ (как бы “белое”, под которым подразумеваются серебряные монеты) напрямую (в таком случае это единственное чувашское слово, отражающее корневую основу **āq*; развитие **q* > чув. *k* рядом с *ś* ожидаемо) или, может быть, как достаточно раннее заимствование из узкотюркской ветви. Из чувашского источника также мар. **ə̑kśa* ‘деньги’. Ср. тат. *aqća* ‘деньги; монета’. Чувашская этимология удмуртского слова принята уже в Wichmann 1903: 122.
- Ср. также аналогичное развитие в удмуртском слове, продолжающем раннее заимствование в волжско-булгарский из арабского: удм. (диал.) *kẹ̑rś*, *körś* ‘подать’ ← ср.-чув. **xə̑r(ə̑)ś* (чув. *xĭrźə̑*, диал. *xĭrə̑ś* ‘тж.’). Далее к др.-чув. **xə̑r(ə̑)ć* ← араб. *ḫarǰ* ‘расходы, издержки’. Из того же арабского источника тат. (диал.) *xarə̑ʒ*, *xarə̑ć*, *qärəʒ* ‘деньги’; отсюда и рус. *харч*, (PL) *харчи́* ‘съестные припасы, продовольствие’. Ср. Wichmann 1903: 73–74; Róna-Tas 1988: 762.

[4] Это слово может быть возведено к чув. *śĭr-t-*, каузативно-интенсивной форме глагола *śĭr-* ‘царапать, чертить и т.п.; писать’ (< ПТ **jąɻ-*), среди значений которого есть и значение ‘проводить борозду’, хорошо совместимое с предполагаемой первичной семантикой ‘*штаны с прорехой’ у производного имени. Сокращение *j* > *ə̑* в северо-западном чувашском, отражаемое и в марийском, обусловлено, очевидно, позицией перед очень не характерным для чувашского языка трёхконсонантным сочетанием.

Развития др.-чув. **ć* > ср.-чув. **ś* не происходило (т. е. фрикативизация блокировалась) в особом фонетическом окружении. В частности, это касается старого сочетания **jć*, которое отражается в следующих чувашских заимствованиях удмуртского языка.

- Удм. *beźi*, (диал.) *böći*, *bići* и т. п. ‘жук (в целом или — по диалектам — отдельные разновидности: хрущ, навозник, светлячок); личинка овода’ ← ср.-чув. **bö(j)ći(j)* (без рефлекса в современном чувашском) < др.-чув. **böjćä-j* (DIM) < ПТ **bög-ćäk* (~ *-ə̄-*) ‘насекомое’, не имеющее рефлекса в татарском. Дальнейшие детали этимологии см. в Савельев 2025: 142. Ср. также Wichmann 1903: 44–45.
- Удм. *beće* ‘родственник’, (диал.) ‘сосед; (старший) брат; ровесница, подруга, спутница’, (бес.) ‘родня (родственники, живущие отдельно друг от друга, имеющие разных родителей, но связанные кровным родством)’ (Кузнецова и др. 2013: 56–57) ← ср.-чув. **bę(j)ćä* (чув. *piććä* ‘старший брат; дядя по отцу’) < др.-чув. **bęj-ćä* < ПТ **bęg-ćä*, диминутивное образование от **bęg* ‘старейшина рода, бег’. В узкотюркской ветви рефлексы этой формы употребляются с гендерным сдвигом, ср. параллельный чувашскому переход в терминологию родства в горноалтайских диалектах (*pićä* ‘ältere Schwester’) и обращения к высокородным дамам в западнокыпчакских языках (кумык. *bi:če* ‘Frau’, балк. *bijce* ‘Frau, Herrin, Fürstin’, караим. *bičä* ‘Königin’; VEWT 67)[5]. В татарском находим *bićä* ‘жена, замужняя женщина’, к которому в принципе могут восходить удмуртские формы с женским денотатом (если только это не внутриудмуртское развитие), но формы с мужским денотатом — явно чувашского происхождения. Ввиду совпадения значений и узколокального распространения тат. (диал.: чеп.) *bić-i* ‘родственник’ лучше считать заимствованием из удмуртского, как это и делается в (ТТДС: 89). Ср. Wichmann 1903: 44.

Кроме того, фрикативизация др.-чув. **ć* блокировалась в позиции т.н. булгарских палатализаций, т. е. перед ПТ **I*, **Ï* или *i*-образным элементом восходящего дифтонга. Об очень ранней первой булгарской палатализации см. Савельев 2025: 141, о т.н. второй палатализации см. ниже критерий (xvii); некоторые случаи с сохранением старого **ć* в булгарской ветви рассмотрены в контексте их отражения в булгаризмах венгерского и южнославянских языков в (Дыбо 2011). В целом надо сказать, что из всех булгарских палатализаций случаи с ПТ **ć* — наиболее проблемные; вне позиции перед старым неогубленным узким гласным для объяснения чув. *ć* в каждом конкретном случае приходится выбирать, вводить ли в реконструируемую пратюркскую форму дифтонг или же считать чувашское слово заимствованным из узкотюркской ветви. Только если налицо серьёзные аргументы против такого заимствования, пратюркские последовательности вида **ćiV-* могут реконструироваться более-менее уверенно.

Что касается отражения данной особенности в чувашизмах удмуртского языка, ср. случай удм. (диал.) *ćęk*, *ćek*, *ćök* ‘вот тебе [жертва]! вот, на!’, *ćęk-an* ‘жертвоприношение’ ←

[5] Ранее я предложил считать чувашское слово заимствованием из какого-то кыпчакского рефлекса ПТ **bęg-će* (Савельев 2023: 35), имея в виду, прежде всего, несовпадение в вокализме между чув. *piććä* ‘старший брат’ и чув. *pü* ‘князь’ (только в сложениях), где последнее слово отражает чистую пратюркскую основу. Однако развитие ПТ **bęg* > чув. *pü* в любом случае нерегулярно (ожидалось бы чув. ***pi*); может быть, стоит реконструировать чистую основу как ПТ **bęgö*, предполагая в булг. **bög(ö)* > чув. *pü* наведение огубленности по второму слогу, а в ПТ **bęg-će* ожидаемое выпадение инлаутного неширокого гласного. При этом с точки зрения семантики возведение чув. ‘старший брат; дядя по отцу’ к предковому ‘*князюшка’ достаточно тривиально (и в части развития значения чистой ПТ основы **bęg* находит параллели в сибирско-тюркском ареале; см. ЭСТЯ 2: 99), в то время как предположение о кыпчакском посредстве потребовало бы постулирования двойного (!) гендерного сдвига (мужской денотат в пратюркском > женский в кыпчакском > мужской в чувашском).

ср.-чув. **ćök* (чув. *ćük* ‘жертвоприношение’) (Wichmann 1903: 114–115). Вопреки Itkonen 1970: 263, чувашское слово не следует считать заимствованием из тат. (диал.) *ćök*, *ćük*. Татарское слово специфично для кряшенских говоров (значение в христианизированном контексте — ‘праздник заговения в Троицын день’, см. Әхмәтьянов 2015, II: 460), а культовая лексика кряшен в значительной степени имеет чувашское происхождение; естественно предполагать обычное направление заимствования и в этом случае. Чув. *ćük* находит прямые параллели в сибирско-тюркском материале; ср., в частности, алт. *ćök* ‘возлияние при жертвоприношении’ (Diószegi 1970). На этом основании может быть реконструировано ПТ **ćo̱ök* ‘жертвоприношение’. Косвенно на раннее присутствие рефлекса этой основы в булгарской ветви указывает вероятный булгаризм в венгерском *csök* ‘гуляние по поводу крестин’. На узкотюркском материале реконструируется также заднерядный вариант той же основы — ПТ **ćo̱oq* (Дыбо 2011). Заимствование из среднечувашского **ćök* отражается в мар. горн. *cök* ‘языческий обряд’, *cök-lə-* ‘приносить жертву’, заимствование из некоторого рефлекса заднерядного варианта тюркской основы — в мар. луг. *ćok-lə̑-*, СЗ *cok-lə̑-* ‘приносить жертву’.

С другой стороны, не может считаться исконным источник удм. *ćibor* ‘пёстрый’ — ср.-чув. **ćɤbar* (чув. *ćə̑bar* ‘пёстрый, рябой’). На этимологически неогубленный узкий гласный указывает не только заимствование в удмуртский, но и материал старописьменных памятников: так, в записях Н. И. Золотницкого, хорошо различавшего чувашские огубленные и неогубленные редуцированные гласные, данная основа записана как ‹чыбар› (КЧРС: 105). В таком случае в чувашском находит продолжение пратюркский вариант **ćïpar* (> узб. *ćipɔr*, уйг. *ćipar*, караим. *ćïbar* и т. д.). При этом в случае исконного рефлекса ожидалось бы чув. ***ćəbär*, так что *ćə̑bar* следует считать заимствованием из кыпч. **ćïbar* (> тат. *ćə̑bar*, башк. *sə̑bar* ‘пёстрый, рябой’). На фоне других вариантов этого пратюркского слова (**ćopur*, **ćupar*, **ćuβar*; см. EDAL: 454, 1344) вариант **ćïpar* может быть объяснён контаминацией с глаголом **ćïp-ar-* ‘покрываться сыпью, волдырями и т. п.’, связанным с именем **ćïp-γan* (ср., например, в гагаузском: *ćïbar-* ‘покрываться чирьями, фурункулами’, *ćïban* ‘чирей, прыщ’). Тат. *ćûwar* ‘пёстрый, рябой’ восходит к ПТ **ćuβar*, а тат. (диал.) *ćŏbar* ‘тж.’ — к ПТ **ćupar*, причём из последнего татарского варианта заимствованы удм. (диал.) *ćubor* ‘пёстрый, рябой’ и мар. **ćʊβar* ‘пёстрый, рябой; сыпь, прыщ, волдырь’. Соотношение форм чувашского и татарского происхождения в удмуртском (но без дальнейших этимологических деталей) было верно установлено уже в Wichmann 1903: 115.

Не является исконным и источник удм. *ćulka*, (диал.) *ćulko* ‘чулки, носки’ — ср.-чув. **ćʊlxa* (чув. *ćə̑°lγa* ‘чулки’). Чувашское слово выглядит как нативизация заимствованного тат. *ćŏlγaw* ‘портянки, онучи’ (с заменой характерного аффикса татарских девербативов *-Aw* на характерный аффикс чувашских девербативов *-A*). Татарское имя образовано от глагола *ćŏlγa-* ‘обвёртывать, обматывать; опутывать’, продолжающего ПТ **ć(i̱)ulγa-* ~ **ć(i̱)uγ-la-* < **ć(i̱)uγ* ‘связка’ (см. VEWT: 119; EDAL: 453); из этого татарского глагола заимствовано чув. *ćə̑°lγa-* ‘запутать (нитки)’. Такая схема значительно лучше объясняет соотношение между значениями чувашского глагола и чувашского имени, чем предположение о прямой производности чув. именной основы *ćə̑°lγa* от глагольной *ćə̑°lγa-* (через присоединение девербативного *-A* к глагольной *-A*-основе, что формально неотличимо от конверсии). На чувашское происхождение удмуртского слова указывают как его значение, так и облик ауслаута (в случае татаризма ожидалось бы *-au*). Из ср.-чув. **ćʊlxa* ‘чулки’ заимствовано, помимо удмуртского слова, и мар. **ćʊlka* ‘тж.’. Ср. на этом фоне коми *ćulki*, *ćuvki* ‘тж.’ ← рус. *чулки* ← тюрк. (Wichmann 1903: 118).

Требует специального обсуждения случай удм. *ćugl-es*, (диал.) *ćulg-es* ‘носки, чулки (из домотканного сукна или холста)’ ← некоторый рефлекс упомянутого ПТ **ć(i̱)uγ-la-* ~

ć(i̯)ulγa-* ‘обматывать’. В удмуртских формах практически лобовым образом отражаются как первоначальный, так и метатезированный варианты тюркского глагола. В части отражения ПТ **γ* эти формы находят явную параллель в основе удм. (диал.: ю., сред.) *bugro*, (сев.) *bugra*, (бес.) *bə̂gra* ‘сруб’, неотделимой от ср.-чув. **bʊwra* (> чув. *pûra*, диал. *pə̂°vra* ‘тж.’, заимствовано также в марийский как **pûwra*) < ПТ **boγra-* ‘делать надрезы, пропилы’ (VEWT: 78). Ср. также тат., башк. *bûra* ‘сруб’, но с учётом того, что за пределами Волго-Уральского региона имя с такой семантикой в узкотюркской ветви не встречается, очень вероятно, что татаро-башкирское слово является булгаризмом. В случае ‘сруба’ собственно удмуртский и бесермянский варианты основы с конечным *-a* следует считать результатом межъязыковой интерференции — перестройки удмуртского слова под влиянием татарского в зоне тесных контактов (в старом заимствовании ожидалось бы **-a* непервого слога > *-o*, что и находим в варианте *bugro*). Удм. *g* в рассматриваемых основах не может быть прямым продолжением ПТ **γ*, поскольку в обеих тюркских ветвях, контактировавших с пермскими языками, — булгарской и кыпчакской — поствокальный **γ* очень рано (заведомо до указанных контактов) переходил в **w*. В связи с этим нельзя согласиться с принятой в Идрисов 2013: 49 интерпретацией удм. *bugro* как «несомненного кыпчакизма». При этом против раннего заимствования по крайней мере удм. *ćugles ~ ćulges* из булгарской ветви свидетельствует отражение гласного *u* на месте ПТ **u*: в случае принадлежности к древнейшему слою булгаризмов ожидалось бы удм. *i̯* (Савельев 2025: 159). Косвенным аргументом против раннего заимствования является документация этих двух основ только в удмуртском, но не в коми. На этом фоне обращает на себя внимание то обстоятельство, что в обоих случаях рефлекс ПТ **γ* находится в кластере с сонантом. По-видимому, весь этот комплекс фактов объясним, если предполагать в удмуртском уже постпрапермские заимствования из тюркского идиома, в котором **w* развивал шумную артикуляцию в консонантном кластере (γ^{w}l*, **γ^{w}r*?) и поэтому на удмуртской почве имела место адаптация соответствующих кластеров как **gl*, **gr*. Заметим, что Б. Мункачи и Ю. Вихманн не считали возможным уверенно предпочитать чувашский либо татарский источник для удм. *ćugles ~ ćulges* (Munkácsi 1896: 327; Wichmann 1903: 117–118). Всё же, поскольку, во-первых, прошедшим через кыпчакское посредство выглядит паронимическое *ćulka* (см. выше) и, во-вторых, в параллельном к *ćugles* случае *bugro* соответствующая татарская форма сама выглядит как булгаризм, вывод именно о булгарском — позднем древнечувашском? — источнике удм. *ćugles* (соответственно, и о единой булгарской интерпретации обоих случаев) представляется наиболее обоснованным.

(xvii) Вторая булгарская палатализация: ПТ **t(I)*, в постсонантных контекстах также **d(I)* > булг., др.-чув. **t(I)* > ср.-чув. **ć* с полузвонким аллофоном **-ʒ́-* в интервокальной позиции (распределение сохраняется в современном чувашском) [→ удм. *ć*, *ʒ́*, но позиционно *č*, *ǯ*]. Как критерий булгаро-чувашских заимствований этот переход отсутствует у Ю. Вихманна и А. Рона-Таша. Развитие зубных взрывных в аффрикату не отражается в булгаризмах ни венгерского, ни южнославянских языков. С учётом того, что в волжско-булгарских эпитафиях данный тип палатализации отражается непоследовательно, в работе Дыбо 2011 обосновывается точка зрения, согласно которой он подействовал «только в волжско-булгарском, причём не ранее XIV в.». Аффрикату как результат булгарской палатализации находим в случае удм. *tužko*, (диал.) *turgičo*, *turkkǰičo* ‘кочерга’ — ср. чув. *torʒ́ə̂ga*, *torćka*, (диал.) *torə̂ćka* ‘тж.’ (Wichmann 1903: 110–111). И в удмуртском, и в чувашском отмечаются метатезированные варианты, но ранний облик основы представлен в чув. *torʒ́ə̂ga* (← ср.-чув. **tɔrʒ́ə̂Ga*). Это старое производное от ПТ **tart-* ‘тянуть’, ср. другие названия кочерги в чувашском (*tort-ma*, *tort-mala*), а также в соседних тюркских языках

(тат. диал., башк. *tärtəškä* ← предковая форма недокументированного чув. ***tortə̂ška* → удм. диал. *turttuško*[6]); с семантическим сдвигом чувашский источник отражается и в мар. (луг.) *torćak* ‘костыль, посох’, *torćaka* ‘гребок (грабли без зубьев для сгребания зерна; швабра)’. Ср.-чув. **tɔrźə̂Ga* восходит к более раннему **tårt-IK-a*. В удмуртском довольно неожиданно отражение твёрдых аффрикат *-ǯ-*, *-č-*; надо думать, это развитие из этимологических *-ʒ́-*, *-ć-* в сочетании с *-r-*.

(xviii) Потенциально критерием стратификации булгаро-чувашских заимствований в пермских могло бы служить и наличие или отсутствие **w*-протезы при развитии пратюркских начальных огубленных гласных (ср. Напольских 2018: 81). Эта протеза составляет критерий № 5 по Ю. Вихманну (Wichmann 1903: 3) и А. Рона-Ташу (Róna-Tas 1988: 763). Установление её хронологии, особенно ввиду возможной неодновременности наведения **w*-протезы при различных огубленных гласных, требует дополнительных разысканий. Во всяком случае материал булгаризмов в венгерском и славянских языках (Дыбо 2010; Дыбо 2021) как будто указывает на то, что долгие огубленные в булгарской ветви дифтонгизировались заметно раньше, чем краткие. Протеза на долгих последовательно отражается в различных пластах заимствований (ср. венг. *váj-* ‘копать’ ← булг. < ПТ **ōj-* ‘долбить’; венг. *vályú* ‘кормушка’ ← булг. < ПТ **ōluq* ‘жёлоб’; ст.-слав. *верига* ‘цепь’ ← булг. < ПТ **ōr-ük* ‘тж.’; вост.-слав. *ватага* ‘шатер’ ← булг. < ПТ **ōtaγ* ‘тж.’). Краткие же огубленные по крайней мере на начальной стадии булгаро-венгерских контактов ещё сохраняли своё историческое качество (ср. случаи типа венг. *ökör* ‘бык’ при чув. *və̂°gə̂°r* < ПТ **ökülɣ* ‘тж.’). Хотя бы относительная датировка дифтонгизации кратких огубленных на этом материале остаётся неопределённой[7]. При этом в корпусе волжско-булгарских эпитафий последней трети XIII — середины XIV в. в основах с анлаутом на старый огубленный (в т.ч. краткий) уже последовательно выписывается و *wāв*: وَلتِ **wẹl-t-ı* (умирать-PST-3) ‘он умер’ ~ чув. *vil-ʒ́-ə* < ПТ **öl-t-i* ‘тж.’; وطر **w.t.r* ‘тридцать’ ~ чув. *və̂°də̂°r* < ПТ **otuɣ* ‘тж.’; وُرُمْ **wurum* ‘длинный, высокий’ (в составе личных имён) ~ чув. *və̂°rə̂°m* < ПТ **uɣïn* ‘тж.’ и т. п. Заманчиво интерпретировать данную орфографическую особенность как свидетельство того, что к этому времени **w*-протеза была системно наведена. На материале чувашских заимствований в марийский язык, массовое проникновение которых предварительно датируется XIV–XV вв. (то есть синхронно основному пласту чувашско-удмуртских контактов), завершённое наведение **w*-протезы в идиоме-источнике подтверждается множеством случаев с анлаутным мар. **β-*. Что же касается булгаро-пермской контактной лексики, то ср. следующие случаи с отражением **w*-протезы.

- Удм. *veme* ‘толока, по́мочь’ ← ср.-чув. **wẹmä* (чув. диал. *vimä*, в литер. языке вторично *nimä*) ‘тж.’ < ? ПТ **ömä(g)* ~ **ümä(g)*, откуда и тат. *ömä* ‘тж.’. С тюркской стороны есть этимологические проблемы ввиду слабой представленности этого слова в узкотюркской ветви (см. ЭСТЯ 1: 633–634), противоречивых вокалических рефлексов и неясных отношений с монг. **öme* ~ **ömüg* ‘помощь’. При этом ввиду специфической для чуваш-

[6] Ср. чув. *torda* ‘оглобля’ (< ср.-чув. **tɔrta* → мар. **tɔrta*, к.-перм. *torta*, удм. диал. *turto*), производное от *tort-* ‘тянуть’ < ПТ **tart-* > тат., башк. *tart-* ‘тж.’, но переднерядный вокализм в тат., башк. *tärtä* ‘оглобля’ — по-видимому, тоже из булгаро-чувашского источника. Такой же неожиданный *ä* находим в тат., башк. *tärə* ‘икона’, которое считается в Rédei & Róna-Tas 1983: 31 надёжным заимствованием из чув. *torə̂* ‘Бог; икона’ (< ПТ **taŋrï*).

[7] Случай волжско-булгарского заимствования в восточнославянском *воръ* (ср. чув. *və̂°rə̂* ‘вор’ < ПТ **oγrï*) нельзя считать примером на отражение «чистой» **w*-протезы, поскольку глайд имеет здесь этимологический характер (булг. **owrï* > **worə̂* по метатезе).

ского языка **w*-протезы конкретный источник удмуртского слова не вызывает сомнений; из среднечувашского источника также мар. **βîmä* ‘пóмочь’. Ср. Wichmann 1903: 128; Róna-Tas 1988: 763.

- Коми *vid-* ‘ругать, бранить’ ← др.-чув. **wit-* ‘говорить просительно, жалобно, с неудовольствием’ < ПТ **öt-*; дальнейшие этимологические детали см. в Савельев 2025: 153–154.

Следующие пермские слова ранее в различных исследованиях предлагалось считать заимствованиями из булгарской ветви в период, предшествовавший развитию **w*-протезы.

- Удм. *uśse*, (бес.) *uśśe* ‘послезавтра’ уже в Wichmann 1903: 125 рассматривается как булгаризм, ср. булг. **öćə* > чув. *viźə* ‘три’, *viźəm* (уст.) ‘третий’, *viśm-iń-ä* (третий-SUBST-DAT.ACC) ‘послезавтра’, букв. ‘на третий [день]’. Принятие данной этимологии наталкивалось бы на серьёзные хронологические противоречия. Отражаемое здесь развитие **ć* > *ś* во всяком случае исключает принадлежность этого слова к древнейшему слою пермских булгаризмов (см. критерий xvi). Более того, данная основа не только не имеет коми параллели, но даже в удмуртском распространена ограниченно: по данным Wichmann 1987: 301 она была известна только у бесермян, северных и западных удмуртов, а в литературный язык это слово вошло лишь во второй половине XX в. из северной диалектной зоны (Максимов 2014: 94). Это указывает на позднее заимствование (во всяком случае послемонгольского времени). При этом неотражение **w*-протезы — архаичный признак, требующий датировать данное заимствование периодом до появления традиции волжско-булгарской монументальной эпиграфики и относить его к более старому слою контактной лексики, чем отражающее **w*-протезу удм. *veme* ‘пóмочь’. Однако булгарская этимология вовсе не выглядит предпочтительной на фоне другого вероятного источника, ср. кыпч. **üć* > тат. *ŭć* ‘три’ и конструкцию *ŭć-ənćə kŭn-gä* ‘послезавтра’ (букв. ‘на третий день’) в татарских диалектах. В казанско-татарском наречии и, в частности, в чепецко-татарских говорах известна реализация предкового **ć*, приближенная к *ś*, ср. описание в Бурганова 1962: 28: «слабовыраженная аффриката», в которой «взрывной элемент весьма незначителен»; в ряде позиций этот звук прямо описывается как «чистый щелевой мягкий спирант *ш*» (т. е. *ś*). Похоже, что рассматриваемое удмуртское слово — заимствование из татарского говора с аналогичным чувашскому развитием **ć* > *ś*, ср. другие подобные случаи: тат. *mäćə*, (диал.) *pəśäj* → удм. (бес.) *piśaj* ‘кошка’ (Кузнецова и др. 2013: 349; Идрисов 2013: 50), тат. *qunəć*, (диал.) *qunəś* → удм. (диал.) *kuńiś* ‘голенище’ (Wichmann 1903: 82), тат. *üć*, (диал.) *üś* ‘месть, злоба’ → удм. *ȩć*, (диал.) *ȩś*, (бес.) *eś* ‘тж.’. Последний случай показателен с той точки зрения, что в плане подъема корневого гласного удмуртские диалекты отражают не современное татарское состояние, а предшествующее известному «поволжскому перебою гласных» (ср. кыпч. **öć* ‘месть, злоба’). По-видимому, и *uś-* в удм. *uś-s-e* ‘послезавтра’ отражает вариант типа **üś* ‘три’ в татарском говоре-источнике, усложнённый уже на удмуртской почве посессивным показателем 3-го лица единственного числа *-s-* и аккузативным *-e*.
- Удм. *ar-a(l)-* ‘жать хлеба’ традиционно (см. Wichmann 1903: 41–42 и приведённые там ссылки на более раннюю литературу) считается заимствованием из предковой формы чув. *vȳr-* ‘тж.’ (< ПТ **or-*). Исторический вид основы типа *aral-* реально зафиксирован у Б. Мункачи (Munkácsi 1896: 27), о выпадении удм. *-l-* конца глагольной основы вне превокалической позиции см. Серебренников 1963: 251. С учётом известного соответствия между частотными наращениями к глагольной основе коми *-av-* ~ удм. *-a(l)-* на роль чистой основы претендует сегмент *ar-*. Если это старое слово, то на прапермском уровне следует реконструировать **är-* (см. об этом Савельев 2025: 154–155); уральского источника ППерм **ä* (традиционно **a*) не имеет, но появляется в заимствованной лексике

в соответствии с переднерядным **ä* или среднерядным **a* языка-источника. Удм. *ar-* < **är-*, очевидно, не может отражать напрямую предковую форму чув. *vi̯r-* (др.-чув., ср.-чув. **wɜ̑r-*); речь в принципе может идти только об отражении ещё более архаичной формы, близкой к ПТ **or-*. Поскольку других примеров на субституцию ПТ **o-* таким нетривиальным образом в корпусе булгаро-пермской контактной лексики нет, данное сближение остаётся недоказуемым. На этом фоне ср. венг. *arat-* (*ar-at-*?) ‘жать хлеба’, которое традиционно считается булгаризмом (WOT I: 70–71), причём предполагаемая субституция булг. *o* → венг. *a* может быть сопоставлена с аналогичной субституцией в венгерской лексике славянского происхождения, и заимствование *arat-* считается в таком сценарии единовременным славянским заимствованиям («в период, хронологически близкий к завоеванию венграми Среднего Подунавья — 895–900 гг.»; см. Хелимский 2000: 415; Дыбо 2010: 83–84). Если этот сценарий верен, удмуртское слово оказывается изолированным и не находящим правдоподобного объяснения. Но есть и другая возможность этимологизации венг. *arat-*, а именно возведение его к ПУ **šurV-* ‘резать’ (UEW: 503–504), хотя и сама реконструкция прауральского этимона в этом случае не вполне надёжна, и особенно спорна принадлежность к нему венгерского слова (венгерский гласный нерегулярно соответствует обско-угорскому и пермскому, а тем более самодийскому). Недавно С. Холопайнен составил перечень случаев, в которых венг. *a* всё же может отражать ПУ **u*, и, хотя венг. *arat-* он в этот корпус в конечном счёте не включает, решение это во многом обусловлено наличием альтернативной булгарской этимологии (Holopainen 2023: 128). Если венгерское слово всё же имеет уральское происхождение, не следует ли предполагать здесь венгерское заимствование в прапермском? При допущении прапермской древности данного слова отсутствие его рефлекса в коми можно было бы объяснить замещением в результате расширения семантики другого глагола, ср. удм. *vund-* ‘резать’ ~ коми *vand-* ‘резать; жать хлеба’. В связи с хлеборобной семантикой ср. с большой вероятностью контактное происхождение пары венг. *kenyér* ‘хлеб’ ~ ППерм **keŋV-* ‘жито, зерно, хлеба́’; правда, трактуется этот случай обычно, наоборот, как прапермское заимствование в венгерском (ср. Csúcs 2007: 97–98). Последнее обращение к проблеме венгерско-пермских контактов см. в Напольских 2026.

- Коми *usur* (диал. *usi̯r*, *usir*), выступающее в роли именного компонента в конструкциях с семантикой ‘надсадиться, надорваться; радоваться до изнеможения’ и в качестве производящей основы для глагола *usur-m-* ‘обессилеть, изнемочь; заснуть (обессилев)’, в Geisler 2002: 241–242 предложено считать древнечувашским заимствованием, ср. чув. *və̑j-zə̑r* [сила-CAR] ‘слабый, бессильный’. Учитывая своеобразие облика сравниваемых слов, извлекаемое из коми форм прототипическое значение, аналогичное чувашскому ‘бессильный’, а также отсутствие альтернативных объяснений[8], указанная этимологическая связь лежит на поверхности, однако дальнейшие детали требуют обсуждения. Прежде всего, дискуссионно происхождение чув. *və̑j*, (диал., вторично) *vûj*, *voj*, *vij* ‘сила’. Заимствование из ППерм **vij* ‘сила’ либо мар. **βîj* ‘тж.’ (< ПУ **wäki*; Zhivlov 2023: 167), которое может быть предложено в этой связи (ср. ЭСЧЯ 1: 104), фактически исключает возможность проникновения в коми чувашской основы с вокалическим анлаутом. Впрочем, с учётом типологии языковых контактов и этноязыковой динамики в Волго-Камском регионе в историческое время заимствование такого абстрактного понятия из субстратных языков выглядит очень сомнительно: все остальные надёжные

[8] О предложенном в *Лыткин, Гуляев 1975: 42 неудачном сравнении коми слова с лугово-восточным марийским именем (диал.) *üšär* ‘непроизводительный труд, напрасная надежда’ (Ӱпӹмарий 1926: 235) и глаголом *üšə̑rγə̑-* ‘стать вялым’ см. упомянутый пассаж в работе М. Гайслера.

заимствования из пермской и марийской ветвей в чувашский имеют явно субстратную семантику (термины природного окружения, названия инструментов для ведения хозяйства в лесной зоне и т. п.), в то время как для многочисленных синонимов чув. *və̑j* может быть продемонстрировано, наоборот, суперстратное происхождение (ср. *xə̑°vat*, *xal*, *xaval* и т. д. — татаризмы в конечном счёте арабского происхождения). Заметим также, что неогубленный *ə̑* в чув. *və̑j*, восстанавливаемый по данным различающих качество редуцированного диалектных словарей и старописьменных источников, может быть вторичным развитием из огубленного *ə̑°* в позиции перед *j*; ср. результат того же процесса в чув. *və̑jə̑* ‘игра’ (< ПТ **ojun*). В связи с этим ср. старую этимологию чувашского слова, а именно возведение его к ПТ **u-* ‘мочь’ (> др.-тюрк. *u-* и т. д., см. Benzing 1959: 707; VEWT: 510). Семантически выведение чувашской ‘силы’ из некоторого предкового ‘*умения, способности’ вполне правдоподобно, но словообразовательная модель остаётся неясной; по крайней мере о продуктивном характере аффиксации типа **u-δ*, **u-j* речь идти не может. На этом фоне, вероятно, стоит обратиться к ещё одной возможности этимологизации чув. *və̑j*, которая опирается на то обстоятельство, что основа чувашского слова *vi̯l’ə̑x*, (диал.: СЗ) *vol’ə̑x*, *vol’ə̑k* ‘скотина’, являющегося прозрачным стяжением из *⁺və̑j-lə̑x*, с одной стороны, может быть отождествлена с *və̑j* ‘сила’ (ЭСЧЯ 1: 135), а с другой стороны, признаётся в Мудрак 1993: 47; СИГТЯ 2006: 35, 173 соответствием УТ **uδ ~ *oδ* ‘бык’ (ЭСТЯ 1: 572–573). Полагаю, возможно объединить эти соображения, возводя чув. *və̑j* ‘сила’ к предковому **oj* < **oδ* ‘тягловой скот, тягловая сила’. Итак, для некоторого раннего волжско-булгарского состояния в принципе мыслимо восстанавливать форму типа **oj-sə̑r* ‘бессильный’; при её заимствовании в прапермский гласный первого слога адаптировался бы как **o* и далее регулярно развивался бы в **u* > коми *u* (Савельев 2025: 160). В связи с упрощением предкового **-js-* в коми *-s-* М. Гайслер приводит исконные коми основы, в которых ПУ **-j-* выпадает в позиции перед согласным. Таким образом, булгаро-пермское сравнение беспроблемно семантически и почти идеально фонетически, но всё же его трудно принять безоговорочно ввиду того, что после отказа от раннебулгарских этимологий для удм. *uśse* ‘послезавтра’ и *ara(l)-* ‘жать хлеба’ данное слово остаётся единственным правдоподобным примером булгаризма в пермских, отражающего стадию до **w*-протезы.

Попытка обобщить рассмотренные — очень немногочисленные — примеры наталкивается на некоторые трудности. Вообще говоря, чувашская форма, послужившая источником удм. *véme* ‘по́мочь’, как и другие тюркские формы облика *ömä ~ ümä*, вполне может объясняться как заимствование из среднемонгольского рефлекса монг. **öme* ‘помощь’. Заметим, что такая возможность в литературе обычно не рассматривается; в WOT I: 320–324 только сибирско-тюркские формы считаются монголизмами, а остальной тюркский материал возводится к ПТ **ün* ‘звук, голос’, что очень неочевидно. Если чувашское слово является монголизмом, несколько проблематично то, что датировать **w*-протезу на **ö-* в таком случае приходится очень узким промежутком в XIII в. (после монгольского завоевания Волжской Булгарии в 1230-х гг., которое создало условия для проникновения монголизмов в языки Волго-Камского региона, и до конца XIII — начала XIV вв., когда **w*-протеза уже надёжно отражается в памятниках волжско-булгарской эпиграфики). Такая датировка вступает в противоречие с анализом коми *vid-* ‘ругать’ как раннего волжско-булгарского (древнечувашского) заимствования; впрочем, эта последняя основа ввиду её краткости и неточного семантического совпадения с булгарским материалом (в чувашском — ‘просить, умолять’, в достоверных чувашизмах марийского языка — ‘доносить, докладывать’, ‘жаловаться, выражать неудовольствие’) принадлежит

к числу менее надёжных случаев из представленного в Савельев 2025: 149–154 списка древнечувашских заимствований в прапермском — здесь нельзя исключать и возможность случайного сходства. Что касается пермских слов, которые ранее предлагалось считать булгаризмами, не отражающими **w*-протезу, то булгарские этимологии удм. *uśse* ‘послезавтра’ и *ara(l)-* ‘жать (хлеба)’, как показано выше, нельзя считать обоснованными. При этом коми *usur* ‘*бессильный’ как будто действительно может быть заимствованием из некоторой ранней булгарской стадии, где **w*-протеза ещё не развилась, но изолированность этого случая не позволяет считать такую стратификацию надёжной.

Вокализм (дополнительные критерии)

(xix) Как было показано в Савельев 2025: 156, 160, рефлекс ПТ **ạ*, **ā* (далее обобщённо **Ạ*) в древнечувашском должен восстанавливаться как **ɤ̂* — напряжённый неогубленный заднерядный гласный верхне-среднего подъёма. В рамках ранних булгаро-пермских контактов этот гласный адаптировался в прапермском как **o*, после чего промежуточный ППерм **o* любого происхождения переходил в **u*. Качество др.-чув. **ɤ̂* сохранялось при переходе к среднечувашскому этапу, и при контактах послемонгольского времени ср.-чув. **ɤ̂* адаптировался как удм. **o̰*[9] (> *e̯* ~ *e̯°* ~ *ö* ~ *o̤* ~ *o* по говорам, см. об этом ряде соответствий Кельмаков 1998: 55–56). Ср. следующие формы.

- Удм. (диал.) *se̯l*, *söw* ‘здоровый’ ← ср.-чув. **sɤ̂wə̂* (чув. *si̯və̂* ‘тж.’) < ПТ **sạγə*, откуда и тат. *saw* ‘тж.’. Также в удмуртском отражается производное абстрактное имя: удм. (диал.) *se̯lli̯k*, *söwli̯k* ‘здоровье’ ← ср.-чув. **sɤ̂w(ə̂)-lə̂x* (чув. *si̯vlə̂x* ‘тж.’) < ПТ **sạγə-luqə*, откуда и тат. *sawlə̂q* ‘тж.’. Наконец, в удмуртский было заимствовано чувашское каритивное образование от той же основы: удм. *se̯se̯ji̯r*, (бес.) *se̯sə̂r* ‘калека’ ← ср.-чув. **sɤ̂wə̂-sə̂r* букв. ‘нездоровый’ (чув. *si̯vzə̂r*, *sûzə̂r* ‘калека, больной, раненый’) < ПТ **sạγə-suɻ*, откуда и тат. *sawsə̂z* ‘больной’. Ср. Wichmann 1903: 92–93.
- Удм. *śe̯li̯k*, (бес.) *śelə̂k* ‘грех’ ← ср.-чув. **śɤ̂w-lə̂x* (чув. *śi̯lə̂x*, диал. *śi̯vlə̂x* ‘тж.’) из **śɤ̂w* < ПТ **jąβ-* ‘плохой’. В узкотюркской ветви эта основа отражается в различных производных (**jąβ-ïz*, **jąβ-γan*, **jąβ-an* > **jam-an*). Не имеет к ней отношения чув. *jabə̂x* ‘плохой; худой, тощий’, заимствованное из тат. *jab-ə̂q* ‘худой, тощий’ < ПТ **jap-* ‘делать низким, придавливать, подавлять’. Из ср.-чув. **śɤ̂w-lə̂x* заимствовано также мар. **śûwlə̂k* ‘грех’. Ср. Wichmann 1903: 99.
- Ср. аналогичное развитие в чувашизме арабского происхождения: удм. (диал.) *ke̯rś*, *körś* ‘подать’ ← ср.-чув. **xɤ̂r(ə̂)ś* (чув. *xi̯rźə̂*, диал. *xi̯rə̂ś* ‘тж.’), дальнейшие детали представлены под критерием (xvi).

Ạ для сравнения рефлексом **Ạ* в татарском является *a* [*å*], который в раннем слое заимствований в удмуртский адаптировался как **å* > удм. *u* (см. об этом критерий xxiii), а в поздних отражается тривиально как удм. *a*.

[9] Такая реконструкция предварительно принимается здесь вслед за Itkonen 1970: 264; Róna-Tas 1988: 762–763. В первой части исследования праудмуртский гласный, реконструируемый для этого ряда соответствий, цитировался как **e̯*, однако реконструкция **o̰* значительно лучше объясняет тот факт, что при помощи этого гласного адаптировался не только ср.-чув. **ɤ̂*, но и ср.-чув. **ö* (см. ниже критерий xx). В связи с этим пассаж «В послемонгольских чувашизмах удмуртского языка в качестве отражения **ɤ̂* находим артикуляционно близкий **e̯*» из первой части исследования (Савельев 2025: 156) должен быть скорректирован следующим образом: «В послемонгольских чувашизмах удмуртского языка в качестве отражения **ɤ̂* находим артикуляционно близкий **o̰*».

(xx) В Савельев 2025: 156, 160 также показано, что др.-чув. **ӧ* (< ПТ **ӧ*, **ӧ̄*) в рамках ранних булгаро-пермских контактов адаптировался в прапермском как **o*, после чего промежуточный ППерм **o* любого происхождения переходил в **u*. Качество др.-чув. **ӧ* сохранялось при переходе к среднечувашскому этапу, и при контактах послемонгольского времени ср.-чув. **ӧ* адаптировался, как и ср.-чув. **ɤ̂*, при помощи удм. **o̭̊*. Помимо упомянутых выше (см. критерий xvi) случаев удм. *bẹćį*, (диал.) *böći* и т. п. ‘жук; личинка овода’ ← ср.-чув. **bö(j)ći(j)* и удм. *ćẹk*, (диал.) *ćek*, *ćök* ‘вот тебе [жертва]!’ ← ср.-чув. **ćök*, ср. аналогичную адаптацию в следующем монголизме: удм. *bẹd’ono*, (диал.) *böd’ono*, *bod’ono* и т. п. ‘перепёлка’ ← ср.-чув. **bödänä* (чув. *püdänä*, диал. *pod’ana* и т. п. ‘тж.’) ← ср.-монг. *bödene* → тат. *büdänä* ‘тж.’. Уже Ю. Вихманн (Wichmann 1903: 47–48) предпочитал для удмуртского слова чувашский источник, указывая на характерное для чувашизмов междиалектное соответствие удмуртских гласных (*ẹ ~ ӧ ~ o*), в то время как в заимствованиях из татарского, отражающих старое **ӧ*, в удмуртском наблюдается иное междиалектное соответствие: *u ~ ӥ ~ ẹ ~ ӧ* (см. представительное число примеров в Wichmann 1915: 37–38). При этом замечание Ю. Вихманна о том, что при заимствовании ‘перепёлки’ из чувашского более ожидаем был бы *-u-* в первом слоге, исходит из ошибочного предположения о первичности заднерядных форм в чувашском. Реально ср.-чув. **ӧ* лобовым образом отражает качество монгольского оригинала. В Róna-Tas 1988: 762–763 для монголизма в чувашском предполагается татарское посредство; это допустимо, но вовсе не обязательно.

(xxi*) ПТ **e*, **ē* (далее обобщённо **E*) > др.-чув., ср.-чув. **ӓ* [→ удм., коми **ӓ* > *a*] > чув. *a*. Критерий булгаризмов № 2 по Ю. Вихманну (Wichmann 1903: 2) и А. Рона-Ташу (Róna-Tas 1988: 763). В татарском в общем случае **E* > *i*, но в особых случаях — *ä*, так что использование этого критерия для разграничения булгаризмов и кыпчакизмов требует осторожности (поэтому номер критерия даётся под звёздочкой). Ср. случаи типа ПТ **teke* ‘баран’ > др.-чув. **täkä* > ср.-чув. **täkä ~ *tägä* > чув. *taga* ‘тж.’, но также и тат. *täkä* ‘тж.’. Надёжным заимствованием из среднечувашского является мар. **tägä* ‘баран, козёл’; из того же источника, судя по всему, удм. (диал.) *taga* ‘тж’. При этом источником удм. *taka*, (бес.) *t’aka* формально может быть как чувашский в стадии до озвончения интервокальных шумных, так и татарский. Такое же удмуртско-бесермянское соответствие *t’- ~ t-* находим в удм. (диал.) *tanka* ‘рубль; монета’, (бес.) *t’an’ka* ‘старинная монета’, что является заимствованием из тат. *täŋkä* ‘монета; рубль’ напрямую либо через посредство ср.-чув. **täŋkä* (> чув. *tängə* ‘серебряная монета; рубль’ с редукцией гласного второго слога). Поскольку данными примерами список возможных среднечувашских заимствований в удмуртский с начальным **tä-*, по-видимому, ограничивается, точнее определить источник этих слов затруднительно.

Природа *ä* первого слога в татарском дискуссионна. Хотя некоторые слова с *ä* в соответствии с ПТ **E* могут быть булгаро-чувашскими заимствованиями, таких случаев в татарском, пожалуй, слишком много, чтобы возможно было целиком объяснить их чувашским влиянием. О. А. Мудрак одним из основных источников тат. *ä* предлагает считать **E* в позиции перед кыпч. **ӧ* второго слога (Мудрак, Хисамитдинова 2018: 101–105).

Отдельную проблему составляет интерпретация примеров, где колебание *i ~ ä* в соответствии с **E* наблюдается уже внутри массива поволжско-кыпчакских диалектов, причём в литературе из-за опоры только на варианты с *i* (часто представленные и в стандартном татарском) соответствующие удмуртские формы с **ӓ* > *a* интерпретируются как несомненные чувашизмы. Ср. удм. *kaban* ‘стог сена’, (бес.) *kaban* ‘скирда’ при чув. *kaban* ‘стог; скирда’ < ср.-чув. **käbän* < ПТ **kEpön* (?), откуда и тат. *kibän* ‘стог’. Удмуртское слово традиционно считается чувашским по происхождению (Wichmann 1903: 66–67; Róna-Tas 1988: 763); ср., однако, татарский диалектный вариант *käbän* (находящий параллель и в

башк. *käbän* ‘тж.’) и распространение его по татарским говорам: мензелинский, бирский, пермский, (!) глазовский (ТТДС 1969: 266). Таким образом, формально удмуртское слово можно было бы возводить и к чепецко-татарскому источнику. Но в целом чувашская этимология выглядит всё же предпочтительнее, учитывая более широкий ареальный контекст: из среднечувашского заимствовано мар. **käβän* ‘стог, кладь, скирда’; кроме того, в вятских говорах русского языка известно (по *Далю*) *кабáн* ‘стог, скирда, кладь продолговатой кладки’, которое Ю. Вихманн, а вслед за ним и М. Фасмер (ЭСРЯ II: 149) также считали чувашизмом (возможно, проникшим в русский через удмуртское посредство). Историко-культурные соображения указывают скорее на чувашское заимствование в удмуртском (булгаро-чувашский является более вероятным источником удмуртской земледельческой лексики, чем татарский).

(xxii**) Как показано в Савельев 2025: 159, др.-чув. **ӳ*, **u* (< ПТ **ü*, **ӣ*, **u*, **ū*) в рамках ранних булгаро-пермских контактов адаптировались в прапермском как **u*, после чего промежуточный ППерм **u* любого происхождения переходил в **i̯*. Далее в волжско-булгарском ареале др.-чув. **ӳ*, **u*, а в части контекстов и огубленные гласные среднего подъёма развивались в ср.-чув. **ʏ*, **ʊ*[10], которые при контактах послемонгольского времени адаптировались в удмуртском как *u*, (бес.) *u ~ ə̂* (с пока не вполне ясным распределением). Аналогичным образом удмуртский адаптировал и поволжско-кыпчакские **ʏ*, **ʊ* (> тат. *ӧ̆*, *ӧ*), так что сам по себе этот критерий не может использоваться для разграничения чувашских и ранних татарских заимствований (даю его номер под двумя звёздочками). Впрочем, чувашское происхождение в таких случаях нередко может быть установлено по дополнительным признакам (например, если слово вовсе не представлено в татарском). Ср. следующие этимологии.

- Удм. (+бес.) *buko* ‘дуга’ ← ср.-чув. **bʏkə* (чув. *pə°gə°* ‘тж.’), производное от **bʏk-* (чув. *pə°k-* ‘гнуть’) < ПТ **bük-*. Хотя производящий глагол представлен и в тат. *bө̆k-* ‘гнуть’, аналогичное отглагольное имя в поволжско-кыпчакском, по-видимому, отсутствует. В Wichmann 1903: 48 среди тюркских девербативов такого облика приводится, помимо чувашского, только осм. *bükü*, *buki* ‘biegung, krümmung’. Из чувашского источника также мар. **pӱgə* ‘дуга’. Ср. ЭСЧЯ 1: 418.
- Удм. (диал.) *surem* в составе *surem šukkon* ‘битьё *сурем* — название обряда, когда молодые люди объезжали все дворы, изгоняя палками из деревни зло, болезни и несчастья’ ← ср.-чув. **sʏräm ~ *sʏrän* (чув. *sə°räm*, *sə°rän* ‘обряд изгнания нечисти’), производное от **sʏr-* ‘гнать’ < ПТ **sӣr-* > тат. *sө̆r-* ‘тж.’. В современном чувашском чистая глагольная основа не сохранилась, при этом предковое др.-чув. **sür-* было заимствовано в ППерм **sür-* > **zi̯r-* ‘гнать’ (Савельев 2025: 154). Из чувашского заимствованы также тат. *sө̆rän* в составе *sө̆rän suγû* ‘битьё *sörän*’ ~ *sө̆rän čabû* ‘скачки *sörän*’ (Напольских 2021: 41, 47, 50–51), мар. **sʏrem* ‘праздник в июле с жертвоприношением богам и затем изгнанием злых духов’.
- Удм. (диал.) *tuluk*, (вторично) *tulup* ‘сирота’ ← ср.-чув. **tʊlə̂x* (чув. *tə̂°lə̂°x* ‘сирота; вдовый’) < ПТ **dulq*, откуда и тат. *tө̆l* ‘вдовый’. Несомненно из чувашского по отражению ПТ **-lq* > *-lə̂x* (Савельев 2025: 145).

[10] Рефлексы пратюркских узких гласных в среднечувашском реконструируются как ненапряженные (расслабленные) гласные верхне-среднего подъёма **ɪ*, **ɤ*, **ʏ*, **ʊ* исходя из того, что, во-первых, в современном чувашском их продолжения реализуются как редуцированные гласные (в верховом и средненизовом диалектах с сохранением противопоставления не только по ряду, но и по огубленности: *ə*, *ə̂*, *ə°*, *ə̂°*), а во-вторых, такое же качество восстанавливается для субституций этих среднечувашских гласных при заимствовании в общемарийский.

- Удм. (+бес.) *итоj* ‘хороший, правильный’ ← ср.-чув. **υmaj* (чув. *maj* ‘лад, порядок, толк’) < ПТ **oŋaj* > тат. *иŋaj* ‘удобный, хороший; лад’. Несомненно из чувашского по отражению ПТ **ŋ* > *m*; дальнейшие детали этимологии см. в Савельев 2025: 143.
- Удм. *vudor* ‘выдра’ < удм. *vu* ‘вода; водный’ + ***udor* ← ср.-чув. **υdə̂r* (чув. *ə̂°də̂°r* ‘выдра’). Традиционная этимология удмуртского слова — народно-этимологическая перестройка общепермского названия выдры (ср. коми *vurd*) под влиянием сложения *vudor* ‘берег реки’ (КЭСК: 70) — выглядит крайне натянутой. Чув. *ə̂°də̂°r* ‘выдра; водяная крыса; хомяк; крот’ (значения распределены по диалектам), по всей видимости, отражает ранние контакты булгар с какой-то индоевропейской ветвью (ПИЕ **ud-r-o-*, **ud-r-eh*$_2$ ‘выдра’). Из чувашского источника также мар. **υdə̂r* ‘крот’[11]; в татарском это слово отсутствует. Удмуртская конструкция с предполагаемым буквальным значением ‘водяная выдра’, на первый взгляд, кажется неправдоподобной ввиду избыточности определения (выдра — по умолчанию водное животное), однако все сомнения снимаются наличием точно такой же конструкции и в соседнем марийском языке: мар. (луг.) *βüδ=koma* ‘выдра’ (*βüδ* ‘вода; водный’ + *koma* ‘выдра’). По-видимому, уже для среднечувашского состояния была характерна представленная в современном чувашском полисемия, в рамках которой основой **υdə̂r* обозначались как водные, так и наземные зверьки, и для удмуртского можно предполагать полукальку сложения типа ср.-чув. **šə̂w υdə̂r-ə* букв. ‘водяной **υdə̂r*’.
- Ср. аналогичное развитие в следующем монголизме: удм. *kudo*, (бес.) *kə̂do* ‘сват; друг; обращение взрослых мужчин друг к другу’ ← ср.-чув. **xυda* (чув. *xə̂°da* ‘сват; друг, приятель; обращение взрослых мужчин друг к другу; адвокат’) ← ср.-монг. *quda* (Róna-Tas 1988: 763) → тат. *qŏda* ‘сват’. Уже в Wichmann 1903: 77–78 было справедливо отмечено, что, хотя с точки зрения фонетических субституций для удмуртского столь же успешно можно было бы предполагать татарское посредство, в своеобразной полисемии удмуртский материал полностью совпадает с чувашским, и на этом фоне именно чувашское происхождение удмуртского слова оказывается несомненным. Косвенно на это указывает и то, пишет Ю. Вихманн, что многие другие свадебные термины в удмуртском также имеют чувашское, а не татарское происхождение.

(xxiii***) Потенциально на булгаро-чувашский источник могут указывать также случаи, где удм. *и* соотносится с чув. (верх.) *о* ~ (низ.) *и* (либо, в особых контекстах, с *û* в обоих диалектах) в соответствии с ПТ **a*, **ā* (далее **A*). Это критерий № 1 по Ю. Вихманну и А. Рона-Ташу (Wichmann 1903: 1; Róna-Tas 1988: 763). Согласно традиционной интерпретации, в случае заимствования татарского рефлекса тех же пратюркских гласных, а именно тат. *а* [*å*], в удмуртском отражается *а*. Однако такой сценарий неоправданно упрощает ситуацию многослойных удмуртско-чувашско-татарских контактов (в связи с этим номер данного критерия даётся под тремя звёздочками). Несомненной является цепочка ПТ **A* > др.-чув. **å* → ППерм **å* > удм. *и*, коми *а* (Савельев 2025: 155, 159); то есть булгарский источник может считаться надёжным, если у удмуртского слова есть соот-

[11] Конечно, с учётом раннего индоевропейского происхождения слова заимствование именно из чувашского в финно-угорские языки Волго-Камского региона, а не наоборот, не выглядит очевидным. Однако этот случай – лишь один пример из обширного корпуса чувашско-удмуртско-марийских параллелей, где формы трёх ветвей связаны друг с другом практически лобовыми фонетическими субституциями. Во многих других таких случаях может быть легко показано, что речь идёт о параллельных заимствованиях из среднечувашского языка послемонгольского времени в ранний удмуртский и общемарийский. По аналогии (поскольку нет явных аргументов в пользу иного сценария) принимаем такую схему и при этимологизации обозначения выдры.

ветствие в коми. Также несомненно, что продолжение др.-чув. **å* в среднечувашском (с некоторой условностью реконструируемое как **ɔ*) адаптировалось в раннем удмуртском лобовым образом как **å* (~ **ɔ*) и затем в результате регулярного сужения заднерядных широких гласных на удмуртской почве давало *и*. Однако в рамках контактов послемонгольского времени аналогичным способом могли адаптироваться и заимствования с тат. *a* ~ *å* — частично прошедшие через чувашское посредничество (собственно, на такую возможность уже указывал А. Рона-Таш), но в части случаев, вполне вероятно, и прямые[12].

Во многих случаях с указанным развитием иные фонетические маркеры всё же позволяют говорить о чувашском источнике удмуртского слова. Ср. примеры типа удм. *bultʒir* ‘свояченица; вторая жена’ ← ср.-чув. **bɔltə̂r* (чув. *poldə̂r* ‘младший свойственник, младшая свойственница’) < ПТ **bāldïʒ* > тат. *baldə̂z* ‘младшая свояченица’ (первый булгарский ротацизм = критерий i по Савельев 2025: 140–141); удм. *buśon(o)* ‘свояк’ ← ср.-чув. **bɔśana* (чув. *poźana* ‘тж.’) < ПТ **bāća-naq*, удм. *busʒ* ‘поле’ ← ср.-чув. **bɔsə̂* (чув. *pozə̂* ‘тж.’) < ПТ **pas-uγ* и удм. *śulo* ‘кнут’ ← ср.-чув. **śɔl-a* (чув. *śola* ‘тж.’) < ПТ **ćal-uγ* (отпадение старых ауслаутных велярных = критерий iv по Савельев 2025: 141–143, а в первом и последнем случаях также развитие **ć* > *ś*, см. критерий xvi выше); удм. *śulʒik* ‘платок-покрывало’ ← ср.-чув. **śɔwlə̂x* (чув. *śûlə̂k* ‘тж.’) < ПТ **jaγ-luqə* > тат. *jawlə̂q* ‘тж.’ и удм. *śulvor-* ‘упрашивать, умолять’ ← ср.-чув. **śɔlwar-* (без продолжения в современном чувашском) < ПТ **jalβar-* > тат. *jalwar-* ‘тж.’ (развитие ПТ **j-* = критерий vi по Савельев 2025: 143–144); удм. *jum* ‘исцеление, лекарство’ ← ср.-чув. **jɔm* (чув. *jom* ‘колдовство; лекарство’) < ПТ **Kʲiam* (сохранение глайдового элемента в старом восходящем дифтонге = критерий xiv по Савельев 2025: 146); удм. *turto* ‘оглобля’ (эквивалент в татарском тоже является булгаро-чувашским заимствованием, см. выше сноску 6) и *tužko* ‘кочерга’ (вторая булгарская палатализация, см. выше критерий xvii); удм. *burtćin* ‘шёлк’ (в поволжско-кыпчакском основа отсутствует, см. секцию 2 ниже); удм. *śurlo* ‘серп’ (развитие **ć* > *ś* на фоне отсутствия основы в поволжско-кыпчакском, см. секцию 2 ниже) и т. д.

С другой стороны, ср. следующие тюркизмы послемонгольского времени в удмуртском языке, в случае которых однозначно определить источник как татарский или чувашский не удаётся: удм. (диал.) *kulə̂m* ‘выкуп за невесту’ ← тат. *qalə̂m* ‘тж.’ либо ср.-чув. **xɔlə̂m* (чув. *xolə̂m* ‘тж.’) (ср. новый татаризм в удмуртском *kalʒim* ‘тж.’); удм. (диал.) *kurok* ‘разбойник, вор’ ← тат. *qaraq* ‘тж.’ либо ср.-чув. **xɔrax* (чув. *xorax* ‘тж.’); удм. *kuromo* ‘вяз; вязка (копыльев саней, салазок)’ ← тат. *qarama* ‘тж.’ либо ср.-чув. **xɔrama* (чув. *xorama* ‘тж.’); удм. (диал.) *uslom* ‘польза, прибыль, выгода’ ← тат. *aslam* ‘торговля наездом; барыш, процент’ либо ср.-чув. **ɔslam* (чув. *oslam* ‘прибыль, барыш’). Эти и подобные им примеры рассматриваются в Wichmann 1903 как надёжные чувашизмы, при этом в позднейшей литературе (ср. Напольских 2018: 82, 510) уже высказывалась мысль о допустимости в таких случаях и татарской этимологии. Важно, что во всех приведённых примерах чувашские параллели сами могут быть ранними татаризмами. А в следующих случаях кыпчакское происхождение чувашского слова и вовсе не вызывает сомнения, и вопрос лишь в том, предполагать ли в удмуртском прямой татаризм или прошедший через чувашское посредство: удм. *bukro* ‘куколь’ ← [ср.-чув. **bɔxra* (чув. *poxra*, *pokra* ‘тж.’) ←] тат. *baqra* ‘тж.’, производное от *baqə̂r* ‘медь, медный’ (по окраске цветков куколя); удм. (диал.) *sundʒik*

[12] Ареально-типологической параллелью здесь может служить контактная ситуация, в которой оказался в послемонгольское время марийский язык (географически прилегавший к удмуртскому). В корпусе общемарийской лексики кыпчакского происхождения надёжно выделяются как прямые татаризмы, так и несущие в себе очевидные признаки чувашского посредничества. В частности, это противопоставление отчётливо наблюдается при адаптации гласных типа *o*: тат. **o* → мар. **ô* (**jôrtə̂-* ‘бежать рысью’), в то время как тат. **o* → ср.-чув. **o* (~ **ɔ*) → мар. **ɔ* (**jɔltaš* ‘товарищ; попутчик’).

‘сундук’ ← [ср.-чув. **sɔndə̂x* (чув. *sondə̂x* ‘тж.’) ←] тат. *sandə̂q* ‘тж.’ (ср. Räsänen 1920: 196); удм. *Kuzon* ‘Казань’ ← [ср.-чув. **Xɔzan* (чув. *Xozan* ‘тж.’) ←] тат. *Qazan* ‘тж.’ (ср. Räsänen 1920: 173).

Таким образом, следует полагать, что удм. *и* появляется в качестве опосредованного продолжения ПТ **A* в старозаимствованной лексике как чувашского, так и татарского происхождения в результате более общего процесса сужения раннего **å* (~ **ɔ*) на удмуртской почве. Утверждение о субституции тат. *а* ~ *å* → удм. *а* верно только в отношении более поздних татаризмов (очевидно, уже не попадавших под действие указанного сужения). Отметим, что такая картина находит точную параллель в стратификации кыпчакизмов чувашского языка, где, как видно из примеров выше, старые татарские заимствования также были охвачены процессом сужения ср.-чув. **ɔ* (> верх. *о*, низ. *и*; последнее развитие аналогично удмуртскому и может составлять вместе с ним единую ареальную изоглоссу), в то время как в новых заимствованиях тат. *а* ~ *å* → чув. *а*.

2. Обновление заимствований в булгаро-пермской контактной зоне

Итак, в корпусе булгаризмов пермских языков надёжно выделяются два наиболее отчётливо контрастирующих друг с другом пласта: датируемые VIII–IX вв. древнечувашские заимствования в прапермский язык (несколько десятков; Савельев 2025: 160–161) и предварительно датируемые XIV–XV вв. среднечувашские заимствования в удмуртский (многочисленные примеры даны выше; общее их число во всяком случае превышает сотню). На этом фоне необходимо обратиться к проблеме обновления заимствований в рамках булгаро-пермских контактов. Речь идёт о таких случаях, когда ранний булгаризм после распада прапермского сохраняется в одном пермском языке, но замещается более новым заимствованием в другом. Если архаизм представлен в коми, то в удмуртском наблюдаем, как правило, замещение за счёт среднечувашской лексики, ср. следующие этимологии.

- Коми *ćarla* ‘серп’ < ППерм **ćårla* ← др.-чув. **ćårla* < булг. **ćarla-γ*, производное от **ćarla-* ‘точить’ (в других тюркских языках такое название серпа не представлено). Удм. (+бес.) *śurlo* ‘серп’ отражает более новое заимствование из ср.-чув. **śɔrla* (чув. *śorla* ‘тж.’). В КЭСК: 302 удмуртская и коми формы верно выводятся из чувашского источника, но при этом ошибочно обе возведены к общепермскому состоянию. В Róna-Tas 1988: 761 несоответствие между коми и удмуртским начальным согласным объясняется междиалектным соответствием **ćarla* ~ **śarla* уже на древнечувашском уровне. Предлагаемая здесь схема с обновлением заимствования в удмуртском более экономна, поскольку оперирует двумя пластами булгаризмов в пермских, существование каждого из которых не вызывает сомнений. Из среднечувашского источника заимствовано также мар. **śɔrla* ‘серп’. Ср. Wichmann 1903: 102–103.
- Коми *bi̯d-m-* ‘расти, вырастать’, *bi̯d-t-* ‘растить’ < ППерм **bi̯d-* < **bud-* ← др.-чув. **büt-* ‘расти, вырастать’. В удм. *bud-* ‘расти’ отражается более новое заимствование из ср.-чув. **bʏt-* (в такой семантике — без продолжения в современном чувашском). В КЭСК: 44 критикуется мнение Wichmann 1903: 45, согласно которому это слово этимологически тождественно ППерм **bi̯d-* < **bud-* ‘кончаться’ (← др.-чув. **büt-* ‘тж.’, см. Савельев 2025: 151). В ЭСТЯ 2: 152–155 тюркские формы типа *büt-* ~ *bit-* со значениями типа ‘кончаться’ и ‘вырастать’ возводятся к двум разным этимонам (и возможная связь между ними даже не обсуждается), а в EDAL: 957 считаются одним. В общем обе группы значений вполне сводятся к ПТ **büt-* ‘быть законченным, завершённым’, так что в конечном счё-

те, пожалуй, речь действительно идёт о единой этимологии. Однако, поскольку две группы значений рано разошлись, можно предполагать, что ко времени булгаро-пермских контактов **büt-* ‘кончаться’ и **büt-* ‘расти, вырастать’ сосуществовали в древнечувашском как отдельные слова без осознаваемой взаимосвязи. Ввиду того, что прапермская и удмуртская форма в разные периоды были заимствованы из булгаро-чувашского именно в значении ‘расти’, очень вероятно, что и в булгаро-чувашском VIII–XV вв. это слово функционировало как основной глагол с такой семантикой. У рефлекса ПТ **ə̄n-* ‘расти’ в чувашском (*ə̂°n-*) документированы только значения типа ‘удаваться, спориться’ и ‘хорошо уродиться (о хлебе и т. п.), быть обильным (об урожае)’, в то время как современное чув. *üs-* ‘расти’, по-видимому, заимствовано из тат. **ös-* > *üs-* ‘тж.’ (Мудрак 1993: 55) и в конечном счёте имеет, возможно, монгольское происхождение (EDT: 241, 251).

- Коми *ṯiļi̯m* ‘волокно, прядь волокна; былинка (о сене); клок’, (диал.) ‘прядь из пеньки (при витье дратвы)’, (диал.) *ṯiļe̯m* ‘прядь (верёвки, косы)’ < ППерм **ṯiļi̯m* < **tuli̯m* ← др.-чув. **tulə̂m* ‘прядь, клок’ < ПТ **tuluŋ* ‘виски; волосы на висках’ (Дыбо 2013: 325–326). В удм. *tulm-et* ‘третья прядь (в верёвке)’, *tulm-a(l)-* ‘ввить третью прядь (в верёвку); вбить клин’ отражается более новое заимствование из ср.-чув. **tʊlə̂m* (чув. *tə̂°lə̂°m*, низ. *tə̂lə̂m* ‘прядь, клочок шерсти, оставленный на спине овцы при стрижке, чтобы по нему узнавать её в стаде’). При этом значение ‘вбить клин’ у глагола *tulm-a(l)-* можно объяснять контаминацией с удм. *tul* ‘клин’. Ср. тат. *tŏlŏm* ‘коса’, (диал.) (ТТЗДС: 670) *tə̂lə̂m* ‘женское украшение для волос’ — по-видимому, исконное продолжение приведённой пратюркской основы, но (диал.: камско-устьинский говор) *tə̂lə̂m* ‘прядь, клок’, судя по семантике и распространению (на Горной стороне близ границы с ареалом буинско-тетюшских чувашей), является заимствованием из низовой чувашской формы. Из чувашского заимствовано также мар. (горн.) *tə̂lə̂m* ‘клок нестриженной шерсти на овце (оставленный как примета)’. Ср. КЭСК: 293; Лыткин 1967: 135–136[13].
- К поверхностному соответствию коми *i̯* ~ удм. *u* как маркеру сохранения древнечувашского заимствования в коми и замещения его среднечувашским заимствованием в удмуртском ср. также рассмотренный выше (секция 1, критерий xxii) случай коми *zi̯r-* ‘гнать’ ~ удм. *surem* ‘*изгнание’ (но здесь речь не идёт о повторном заимствовании глагольной основы; глагол ‘гнать’ в удмуртском замещался за счёт инноваций на собственной почве).

С другой стороны, в следующих случаях удмуртское слово выглядит как прапермский архаизм, в то время как в коми представлены замещающие старое слово заимствования из русского языка (а в первом случае — ещё и внутренняя инновация, обусловленная сдвигом семантики иной общепермской основы).

[13] Заимствование из древнечувашского в прапермский в принципе может быть поставлено под сомнение с учётом возможных возражений против исконности чувашского слова. Действительно, в заднерядных контекстах, восходящих к ПТ **-VlUN* (где *U* – узкий гласный, а *N* – носовой), плавный сонант в чувашском может выпадать, ср. в особенности квазиминимальную пару к случаю ‘пряди’: ПТ **quluN* ‘жеребёнок’ > чув. *xə̂°m* ‘плод, зародыш (у лошади)’. Однако ср., с другой стороны, и примеры типа чув. *śolə̂m* ~ УТ **jalïN* ‘пламя’ и т.п. По мнению А. В. Дыбо (личное сообщение), устранение **l* в подобных случаях может быть объяснено последовательностью закономерного выпадения узкого гласного во втором слоге основы в посессивных контекстах (**-VlUN* > **-VlNi*), упрощения **lN* > *N* (ср. примеры типа ПТ **jąlŋuq* > чув. *śi̯nə̂* ‘человек’) и аналогического распространения данного развития на непосессивные контексты. Однако в таком случае речь идёт о феномене парадигматического выравнивания, по умолчанию не предполагающего регулярности, так что считать отражение *-l-* в таких чувашских основах аргументом против их исконности нет оснований.

- Удм. *burtćin* ‘шёлк’ < ППерм **bår(t)ćin* ← др.-чув. **bårćə̑n* < ПТ **bārćun* (в конечном счёте иранского происхождения). Далее в волжско-булгарском ареале ср.-чув. **bɔrśə̑n* > чув. *porźə̑n* ‘тж.’. Булгаро-чувашское происхождение удмуртского слова было принято уже в Wichmann 1903: 49–50; при этом была отмечена и нетривиальность удмуртской аффрикаты (вместо ожидаемого сибилянта *ś*; см. секцию 1, критерий xvi). Сам Ю. Вихманн полагал, что удм. *burtćin* заимствовано из недокументированного собственно чувашского диалекта, сохранявшего *-ć-*. Однако никаких спекулятивных допущений не требуется, если предположить заимствование из раннего волжско-булгарского (древнечувашского) источника (где ПТ **ć* сохранялся, см. Савельев 2025: 158) в прапермский, а после распада праязыка — закономерное развитие этого слова на удмуртской почве и замещение его иными основами в коми. Ср., во-первых, коми *jermȩg* ‘шёлк’, которое, как указывает В. В. Напольских (личное сообщение), следует сравнивать с удм. *jurmeg* ‘ласка, горностай’. В таком случае для коми обозначения шёлка можно восстанавливать старую семантику типа ‘*горностай (мех)’. Во-вторых, в коми имеется и слово *šȩvk* ← рус. *шёлк*. Булгаризм в удмуртском в любом случае несомненен, поскольку рефлексы ПТ **bārćun*, хотя и известны в различных древне- и среднетюркских памятниках, в поволжско-кыпчакских языках отсутствуют (тем более не обоснована идентификация чувашского слова как кыпчакизма в СИГТЯ 2006: 45). Из среднечувашского источника заимствовано мар. **pɔrśə̑n* ‘шёлк’.
- Удм. *keć* ‘коза’ < ППерм **kɛć* (в реконструкции В. И. Лыткина) / **kȩć* (в реконструкции В. В. Понарядова) < **kɛ/ȩći* ‘коза’← др.-чув. **keći* ~ **kećə* < ПТ **geći* ~ **gećə* ‘тж.’. В коми коза обозначается русским заимствованием *kȩza*. В волжско-булгарском ареале др.-чув. **keći* > ср.-чув. **käćə* > чув. *kaś-aga* ‘коза’, в то время как др.-чув. **kećə* > ср.-чув. **käśə* > чув. *kaś=ťaga* ‘козёл’; последний вариант основы зафиксирован также в старописьменном чув. (XVIII в., словник Миллера; см. Савельев 2023: 31–32) ‹KáÍa› ‘Ziege = коза’, где отражается вариант типа ***kaźə̑*, а также как среднечувашское заимствование в мар. **käśə* ‘коза’. Булгаро-чувашское происхождение удм. *keć* широко принимается (ср. Wichmann 1903: 73; Rédei, Róna-Tas 1983: 24); восходящая к Munkácsi 1883: 82 альтернативная этимология, усматривающая здесь отражение тат. *käźä* ‘коза’, не выдерживает критики: как показано уже Ю. Вихманном, в таком случае в удмуртском ожидалась бы форма типа ***kaza*. При этом удм. *keć* ‘коза’ очень трудно отделять в этимологическом отношении от коми *kȩć* ‘заяц’: соответствие удм. *e* ~ коми *ȩ* в пермистике хорошо известно (Лыткин 1964: 153–162; Понарядов 2014: 18–19), а с точки зрения семантики связующим звеном является удмуртское сложение *lud=keć* ‘заяц’, букв. ‘полевая/дикая коза’ (удм. диал. *keć* ‘заяц’, известное в кукморских говорах, может быть результатом эллипсиса этого сочетания). Таким образом, данная основа правдоподобно выводится на прапермский уровень, но остаётся вопрос относительно её исходной семантики. Основанная на сопоставлении с лив. *kõps* ‘заяц’ точка зрения, согласно которой в коми слове находит продолжение старое пермское название зайца (см. Viitso 1993: 89–91; Напольских 2018: 503), не может быть принята, поскольку: (a) коми *ȩ* в соответствии с удм. *e* в исконной лексике продолжает ПУ **ä*, что несовместимо с ливским вокализмом; (b) ливское название зайца предпочтительнее считать локальной инновацией, связывая его с эстонскими экспрессивными глаголами *kapsata* ‘прыгнуть (в т.ч. о зайце)’ ~ *kepselda*, *kepsutada* ‘скакать, резвиться (в т.ч. о зайце)’ ~ *kopsata* ‘ударить’ ~ *kõpsutada* ‘производить стучащие звуки при ходьбе и т. п.’, ср. также эст. *kõps* как именное обозначение подобных звуков (возможность объяснения ливского слова на такой почве допускал и сам Т.-Р. Виитсо); (c) значительно более подходящим кандидатом на роль архаичного пермского обозначения зайца является ППерм **ńimVl* > к.-язьв. *ńimɵl* ‘заяц’,

имеющее надёжные уральские истоки (ПУ **ńoma-la* ‘тж.’; Zhivlov 2023: 162). Интересную ареальную параллель составляет тат. (диал.) *qə̂r-taj* ‘заяц’, где *qə̂r* ‘поле(вой); дикий’ функционально тождественно удм. *lud* в составе *lud=keć* ‘заяц’, а *taj* имеет значение ‘жеребёнок’. Данный прецедент осмысления зайца как дикой формы некоторого домашнего копытного (склонного прыгать, резвиться и т. п.) является дополнительным аргументом в пользу первичности значения ‘коза’ у рассматриваемого пермского слова. Таким образом, если именно это значение реконструируется для прапермского состояния, то коми значение ‘заяц’ следует объяснять постпрапермским семантическим сдвигом. По всей видимости, на фоне этого сдвига в коми и возникла необходимость вновь заимствовать название козы, на этот раз из русского. Проблему при рассмотренном сценарии составляет только то обстоятельство, что ‘коза’ оказывается, возможно, единственным булгаризмом, где рефлекс ПТ **E* имеет в пермских субституцию, отличную от удм. *a* ~ коми *a* (см. Савельев 2025: 154–155). Естественно, этот феномен можно связывать, например, с диалектным варьированием на древнечувашском уровне, но ввиду отсутствия хороших параллелей на такую же субституцию приходится признать *ad hoc* характер такого объяснения.[14]

[14] С учётом сказанного требует дополнительного обсуждения предложенное в работе Напольских 2025: 144 возведение удм. *besmen* ‘межа, граница (полей, покосов); конец звена в изгороди’ к волжско-булгарскому прототипу вида **bäsmän* ‘мера веса > ? площадь поля, засеваемая определённым количеством зерна > ? границы поля’ (ср. чув. *pasman* ‘мера зерна’). В общем нельзя исключать, что удм. *besmen* является ещё одним, наряду с удм. *keć* ‘коза’, примером на удм. *e* как продолжение прапермской субституции др.-чув. (диал.) **e*, но такая интерпретация не является единственно возможной. Дело в том, что высказанные В. В. Напольских семантические и фонетические возражения против непосредственной связи удм. *besmen* с рус. *безме́н* и чув. *pismän* ‘весы-безмен’, как кажется, могут быть сняты, если предполагать заимствование в удмуртский из (др.-рус. →) ср.-чув. **bẹsmän*. Фонетические субституции в таком случае тривиальны, а синхронное удмуртское значение, как и в первом сценарии, вполне выводимо через правдоподобную цепочку семантических изменений из значения типа ‘мера веса’ в идиоме-источнике – нужно лишь допустить, что среднечувашская полисемия была аналогична сосуществованию значений ‘мера веса’ ~ ‘род весов’ в истории русского слова (Напольских 2025: 139). Таким образом, удм. *besmen* может быть объяснено как позднее сепаратное заимствование без необходимости постулировать необычную субституцию древнечувашского гласного (ср. и отсутствие параллели в коми, что снижает вероятность прапермской древности этого слова). По фонетическим основаниям от данного этимона, судя по всему, следует отделять (признавая активную контаминацию по крайней мере в удмуртском) следующий набор форм: чув. *pəśmän* – компонент топонимов (*Pəśmän śi̯rmi* ‘овраг *Пĕçмен*’); удм. (диал.: ю.) *biśmen*, (пер.-ю.) *bićmen* – также компонент топонимов, а в апеллятивном употреблении – слово, семантически тождественное удм. *besmen*; мар. **piśmän* ‘межа между полосами пахоты’. Удмуртское и марийское слово являются чувашскими заимствованиями, а чувашское анализируемо как имя по продуктивной модели на **-mAn* от ср.-чув. **pɨ́ś-* < ПТ **bḯć-* ~ **bić-* ‘(от)резать’. Кроме того, в татарском находим *pəćmän* – компонент топонимов, а в апеллятивном употреблении ‘коллективный выход в начале лета на поля с ритуальными молениями’ (< метонимически из ‘*участок поля; межа’; Әхмәтьянов 2015, II: 109). Формально это имя выглядит производным от татарского рефлекса (*pəć-* ~ *pə̂ć-* ‘обрезать и т.п.’ с не совсем ясным – экспрессивным? – глухим *p-*) того же пратюркского глагола и является, видимо, нативизированным заимствованием / словообразовательной калькой по булгаро-чувашскому образцу. На ср.-чув. **pɨ́ś-* ‘резать’, рефлекс которого в современном языке не сохранился, указывает также мар. луг. *piskĕ*, СЗ *piskə* ‘пила’ ← ср.-чув. **pɨ́ś-kə* ‘тж.’ при чув. *pə̂ćkə̂* ‘пила’ ← тат. *pə̂ćkə̂* ‘тж.’. К семантической производности типа ‘участок земли’ < ‘резать’ ср. рус. *нарезать* в значении ‘размежевать участок земли’, *нарез* ‘надел, выделенный в собственность участок земли’. Аффриката *ć* в периферийно-южной удмуртской форме носит вторичный характер и потому не должна рассматриваться как отражение старого булг. **ć*. Ср. хорошее изложение синхронного материала по этому последнему набору форм, но сомнительную этимологическую интерпретацию в Culver & Metsäranta 2025: 337–341. Предлагаемые в этой работе производность от предковой формы чув. *piś-* ‘вариться, печься; калиться (на солнце); обжигаться и т.п.’ и реконструкция исходного для имени значения ‘*подсека, росчисть’ проблематичны как фонетически (соответствия гласно-

Следует подчеркнуть, что учёт феномена обновления заимствований позволяет не только адекватно разобрать ряд трудных булгаро-пермских этимологий, но и расширить представленный в Савельев 2025: 147–154 корпус вероятных древнечувашских заимствований в прапермский: к нему также могут быть отнесены ППерм **ćårla* ‘серп’, **bud-* > **bi̯d-* ‘расти’, **tuli̯m* > **ti̯li̯m* ‘прядь’, **bår(t)ćin* ‘шёлк’, **kɛ/e̯ći* > **kɛ/e̯ć* ‘коза’.

3. Узколокальные зоны поздних булгаро-пермских контактов: Нижняя Вятка и Нижняя Чепца

Отдельную, до сих пор очень слабо разработанную проблему составляет возможность заимствований в уже разошедшиеся диалекты языков пермской группы из (поздне)среднечувашского состояния.

Некоторая географическая неравномерность в насыщенности удмуртских диалектов чувашизмами заметна уже по классическим материалам Б. Мункачи и Ю. Вихманна (Munkácsi 1896; Wichmann 1903). Однако текущее состояние вопроса не позволяет сделать на этом основании системные выводы: наблюдаемая картина неизбежно содержит значительный «шум», связанный, в частности, с неравномерностью самой документации удмуртских диалектов; кроме того, узколокальные чувашизмы не выделяются какими-либо фонетическими признаками на фоне несколько более ранних чувашизмов, широко распространённых в удмуртском диалектном массиве, так что поздние контакты могут быть выявлены только на основании дистрибуционного критерия. На данном этапе можно лишь выделить несколько локальных зон, где поиск поздних сепаратных чувашизмов представляется особенно перспективным.

Знаковой для этого направления исследований является работа Напольских 2021, в которой рассматривается генезис названия весеннего обрядового комплекса удм. (диал.: завятские говоры) *akaška*, (бес.) *akajaška* и связанной с ним терминологии (удм. диал. *surem* ‘*обряд изгнания злых духов’); см. формальную сторону данных этимологий в Савельев 2025: 146 и под критерием xxii в секции 1 настоящей статьи. По реконструкции В. В. Напольских, не позднее конца XV — начала XVI в. предки бесермян заимствовали этот ритуальный комплекс и его название от чувашей, населявших правобережье Нижней Вятки, и затем перенесли его на север в низовья Чепцы, а не ранее второй половины XVI в. этот термин и видоизменённый обряд были заимствованы юго-западными (завятскими) удмуртами от того же чувашского населения или же оставшейся на Нижней Вятке части бесермян. Следует отдельно остановиться на двух аспектах указанной реконструкции.

Во-первых, принципиально важна чёткая формулировка позиции, согласно которой ещё несколько столетий назад одним из ареалов расселения чувашского народа была Арская сторона вплоть до Нижней Вятки, ср.: «…факты, свидетельствующие о присутствии в это время здесь этнической группы с чувашским языком, чувашским традиционным костюмом и комплексом весенних обрядов, происхождение которых связано в том числе и с чувашской традицией, отрицать невозможно» (Напольских 2021: 52). Факт заимствования завятскими удмуртами и бесермянами культурной лексики из языка этой группы указывает на её заметный социолингвистический вес (даже если сам по себе и не позволяет судить об абсолютной численности местного чувашского населения). При этом русские документы XVI–XVII вв. с территории Свияжского и Казанского уез-

го первого слога трудно объяснимы), так и семантически (ни чув. *piś-*, ни его когнаты в других тюркских языках не употребляются в значении типа ‘быть сожжённым, выжженным’, которое требовалось бы в качестве производящего для значения ‘подсека’).

дов содержат множество упоминаний т. н. *ясачной чюваши* в Предкамье. Согласно распространённой в татарской исторической школе трактовке, этот термин имел характер соционима, применявшегося в отношении татарских крестьян. Работа В. В. Напольских стала важной вехой в прямой критике этой концепции; развитие его аргументации с привлечением широкого круга лингвистических и исторических свидетельств см. в Савельев & Мясников 2026: 7–9, 46.

Во-вторых, предложенное объяснение бес. *akajaška* как термина, заимствованного предками бесермян на Нижней Вятке и затем перенесённого ими на Чепцу, ставит более общий вопрос о природе чувашско-бесермянских контактов. В частности, необходимо выяснить, могут ли все их лингвистические следы объясняться контактами на прежней территории расселения бесермян или же часть соответствующего материала предполагает непосредственные контакты с чувашеязычным населением уже в бассейне Чепцы.

С точки зрения письменной истории последний сценарий не лишён оснований, поскольку именно с бассейном Чепцы связаны наиболее ранние упоминания чувашей в русских документах XVI–XVII вв. (*Л*уппов 1958; Гришкина 1988; Гришкина 1989; Мусихин 2013; Савельев & Мясников 2026: 6–7). Уже в грамотах 1510, 1522, 1546 и 1548 гг. чуваши упоминаются среди населения Чепецкого края. Особый интерес представляет грамота Ивана Грозного от 1542 г., в которой каринским князьям разрешалось призывать в свои вотчины удмуртов и *чюваш* «из казанских мест»; при этом отдельно оговорены случаи, когда удмурты и чуваши сами желали воссоединиться со своими жёнами и детьми, уже находившимися в «великого князя вотчине в полону на Вятке». Эти данные хорошо согласуются с реконструкцией позднесредневекового миграционного коридора из района Нижней Камы и Нижней Вятки на Чепцу[15].

Проблема интерпретации этноязыковой реальности, стоявшей за термином *чуваши* применительно к Чепецкому краю, осложняется не вполне ясным контекстом сосуществования этого обозначения с термином *бесермяне*. В документах XVI в. с территории Вятской земли в бассейне Чепцы упоминаются только *чуваши*; с начала XVII в. появляется название *бесермяне*, после чего оба обозначения некоторое время употребляются параллельно, а в отдельных случаях — как будто взаимозаменяемо. Начиная с переписной книги 1678 г. слово *чуваш* из документов Вятской земли исчезает, и в дальнейшем в них фигурируют уже только *бесермяне*. Т. И. Тепляшина (1968: 179–181) интерпретировала эту картину таким образом, что *чувашами* до некоторых пор назывались и поныне населяющие бассейн Нижней Чепцы бесермяне. Однако проблема в действительности распадается по меньшей мере на три связанных между собой, но требующих раздельного рассмотрения вопроса. Первый касается этноязыковой принадлежности чепецкой *чюваши* письменных источников; второй — наличия чувашского субстрата в бесермянском наречии и, в случае признания такого субстрата, места его формирования; третий — возможных элементов чепецко-чувашского происхождения в наречии бесермян, а также в иных идиомах Чепецкого края и сопредельных районов бассейна Вятки.

[15] Следует отметить, что маршрут, связывавший ядерные булгарские территории с бассейном Чепцы, был освоен, судя по всему, задолго до позднесредневековой миграционной волны XV–XVI вв. Булгарские укреплённые фактории существовали в этом ареале уже с IX–X вв. (Напольских 2018: 73–75, 99–100); с существованием такой фактории близ Гурьякара Т. И. Останина (2021) связывает установку здесь в 1323 г. т. н. «Гординского камня» – надгробной плиты на булгаро-чувашском языке с арабскими формульными вставками. Таким образом, к началу XIV в. бассейн Чепцы уже находился в зоне устойчивых связей с булгарской метрополией. Это обстоятельство существенно как свидетельство исторической глубины данной линии контактов, но само по себе, естественно, не указывает на прямую преемственность между булгарским присутствием на Чепце в начале XIV в. и чепецкой *чювашей* позднего Средневековья.

Что касается первого вопроса, есть серьёзные основания сомневаться в простом отождествлении чепецких *чювашей* и бесермян. Дело в том, что на протяжении письменной истории Волго-Камского региона этноним *чуваш* в целом устойчиво связан с группами, являющимися чувашеязычными или, по крайней мере, бывшими чувашеязычными в недавнем прошлом[16]. Собственно, на это обстоятельство уже обращал внимание В. В. Напольских (2021: 51), предполагая переселение части чувашского населения из региона Нижней Вятки — Нижней Камы вместе с предками бесермян, чепецких удмуртов и каринских татар. Ещё определённее высказывается по этому поводу С. А. Максимов (2008: 108), прямо рассматривающий историческую *чювашу* бассейна Чепцы как группу собственно чувашского населения, в конечном счёте ассимилированную бесермянами. Антропонимический материал может служить дополнительным, пусть и пока ограниченным (поскольку системная его обработка остаётся делом будущего) аргументом в пользу чувашской этноязыковой принадлежности чепецких *чювашей*. Хотя их имена, как и имена бесермян, в основном носят выраженно татарский облик (что в целом характерно для традиционного антропонимикона чувашей и наиболее тюркизированных финно-угорских народов Поволжья), некоторые из них прозрачно этимологизируются на чувашской почве и не имеют очевидных альтернативных объяснений; ср. в особенности имя *Хоза Сабетов* из переписной книги 1646 г. ~ чув. *хоźа* ‘господин, хозяин’.

Из ответа на первый вопрос следует, что идентификация чепецких *чювашей* с бесермянами была бы в принципе допустима, если предположить, что бесермяне до некоторых пор говорили на чувашском языке и лишь впоследствии были удмуртизированы. Это — гипотеза чувашского субстрата в языке и культуре бесермян — составляет суть второго вопроса. Формулировки данной гипотезы варьируются в зависимости от локализации и датировки предполагаемого субстратно-суперстратного взаимодействия. Вариант субстратной концепции, согласно которому ещё в Нижнем Прикамье некоторые «древнебулгарские» группы послужили основой для формирующегося бесермянского народа, был предложен Т. И. Тепляшиной (1970: 243) и на десятилетия стал *received wisdom* бесермяноведения (ср. Насипов 2010: 17). С другой стороны, С. А. Максимов (2008: 108), признающий, как было сказано, чувашский этноязыковой профиль чепецкой *чюваши*, предполагает, что в конечном счёте это население было ассимилировано бесермянами. Вообще говоря, вхождение некоторых чепецко-чувашских родов в состав бесермянского народа выглядит вполне вероятным с учётом того, что смена идентификации *чуваш* > *бесермянин* прослеживается по документальным свидетельствам порой внутри одних и тех же семейных линий[17]. Однако приходится оговориться, что этнические маркировки в переписях XVII в. в целом исключительно нестабильны, что ограничивает применимость этого корпуса данных в этноисторических и историко-лингвистических построениях. Может быть, ассимиляция в данном случае имела только точечный характер, поскольку — и это критично для всей «гипотезы древнебулгарского/чувашского субстрата» — предположение о наличии в бесермянском наречии специфи-

[16] Единственный известный мне достоверный пример на последнюю ситуацию – молькеевские (подберезинские) кряшены, жители Кайбицкого района Татарстана на границе с Чувашией, которые в этнографическом отношении очень мало отличаются от низовых чувашей и говорят на мишарско-татарском диалекте с большим количеством чувашских заимствований, в т.ч. выраженно субстратного (судя по отражаемым семантическим полям) характера. В течение XIX в. в среде кряшен этой группы неоднократно фиксировалось самоназвание *чуваш* (История и культура татар-кряшен: 225–226).

[17] На вероятную ассимиляцию отдельных чувашских родов не только среди бесермян, но и среди удмуртов Каринского стана указывают также личное имя *Чюваш* и фамилия *Чювашев*, документированные в переписях XVII в. среди бесермянского и «отяцкого» населения.

чески субстратных элементов булгаро-чувашского происхождения не подтверждается анализом фактического языкового материала (Кельмаков 1985; Идрисов 2013). Само по себе исчезновение чувашей как отдельной этнической группы бассейна Чепцы вовсе не свидетельствует автоматически о преимущественной или исключительной ассимиляции среди бесермян — об этом будет подробнее сказано ниже.

На этом фоне важно подчеркнуть, что известный факт исключительного этнографического сходства между чувашами и бесермянами, а равно и наличие чувашских заимствований в бесермянском наречии (которые, однако, в массе своей не являются сепаратными, присутствуя и в других удмуртских диалектах; см., например, Идрисов 2013: 27–29) вовсе не требуют предположения именно о субстратно-суперстратном характере связей. Совокупность соответствующих параллелей вполне могла возникнуть в результате относительно длительного и тесного сосуществования двух групп — т. е. адстратного взаимодействия. И если упомянутая работа В. В. Напольских показывает, что такое взаимодействие началось, по-видимому, ещё на Нижней Вятке — Нижней Каме, завершиться оно могло уже на Чепце.

Переходя, таким образом, к третьему из поставленных вопросов — о возможных элементах чепецко-чувашского происхождения в языках Чепецкого края и сопредельных территорий, — следует сразу отметить крайнюю разреженность соответствующего материала.

В бесермянском наиболее серьёзным кандидатом на статус позднего сепаратного чувашизма является, пожалуй, случай *gugama* ‘повитуха’, *gugama abi* ‘тж.’, *gugam-a(l)-* ‘принимать роды’ (Wichmann 1987: 59, 62) ← ср.-чув. **kogama* (чув. *kogama-j* VOC ‘бабушка по матери; обращение к старухе’). Чувашская этимология предложена для бесермянского слова в Максимов 2018b: 35–36. Чув. *kogama-j* является сложением основы **koga*, с эмфатической геминацией представленной в чув. *kokka* ‘дядя по матери’ (? ср. казах. *köke* ‘обращение к старшему брату ~ отцу ~ дяде по отцу’) и чув. *ama* ‘мать’. Таким образом, речь здесь идёт о ‘бабушке по матери’ как о ‘матери брата матери’; ср. ту же модель, например, в удм. *čužanaj* ‘бабушка по матери’ и см. специально посвящённую ей статью Максимов 2019. Консонантный костяк *g-g-* в бесермянском обусловлен, по-видимому, контаминацией с *gogi* ‘пупок’. Дальнейшее развитие этой народной этимологии, т. е. окончательную перестройку чувашизма под прозрачное сложение удмуртских основ, находим в удм. (сев.: слободской говор) *gogi=mumi* ‘повитуха’, букв. ‘пупочная мать’. К полисемии ‘бабушка’ ~ ‘повитуха’ ср., помимо рус. *повивальная бабка*, например, чув. (диал.) *äbi* ‘бабушка; мать; повитуха’, а также *babuška* ‘повитуха; знахарка’ (← рус.) в современной документации бесермянского наречия (Усачёва и др. 2017: 247–248). Узколокальное распространение термина *gugama* в бесермянском наречии и, по-видимому, народно-этимологически перестроенного варианта того же термина в слободском удмуртском позволяет с большой вероятностью связывать заимствование этого слова с поздним чувашским присутствием непосредственно на Чепце.

С другой стороны, с пребыванием чувашского населения на Нижней Чепце и прилегающих территориях бассейна Вятки могут быть связаны отдельные сепаратные чувашизмы в южных коми-зырянских диалектах. Некоторые такие случаи рассматриваются в работе Г. В. Федюневой (2014); большинство предложенных ею этимологий, впрочем, не может быть принято. Ср., однако, следующий случай, где чувашское заимствование конкретно в южные коми-зырянские говоры (без соответствий в других пермских идиомах) несомненно: коми (диал.: лузско-летские и нижневычегодские говоры) *kẹrša* в составе *śin kẹrša* ‘бровь’ (*śin* ‘глаз’), ср. чув. *xarža* в составе *koś xarž-i* ‘тж.’ (*koś* ‘глаз’). Ср. удм. *śin-kaš* ‘бровь’, где вторая часть заимствована из тат. *qaš* ‘бровь’. УТ **qāš* восходит

к ПТ **qāɬ*, откуда и ср.-чув. **xəl* → мар. (горн.) *xal* ‘бровь’. Наряду с этим можно восстанавливать диминутивную форму вида ПТ **qāɬ-aq* > др.-чув. **xə(r)ša* с развитием **ɬ* > **š* перед гласным по правилу О. А. Мудрака. Наконец, развитие *š* (*-ž-*) > *-rž-* в интервокале является среди тюркских специфически чувашским, причем довольно поздним, поскольку под него подпадают даже некоторые русские заимствования (например, рус. *сажень* → чув. *šaržan*). Существенно, что часть южнокоми-зырянского ареала относится, как и земли по Чепце, к бассейну Вятки (*Л*етка является её правым притоком), так что узколокальную дистрибуцию этого заимствования естественнее всего объяснять контактами на северной периферии предполагаемого чепецко-чувашского ареала.

Практически не изучен вопрос о возможном чувашском субстрате в диалекте чепецких (каринских) татар. По крайней мере, вызывает интерес информация Н. Б. Бургановой (1962: 27), согласно которой для этой группы говоров характерен финальный *-m* в соответствии с *-n* в других татарских диалектах. Как известно, появление *-m* в соответствии с УТ **-n* — характернейшая черта фонетики булгарской ветви (СИГТЯ 2006: 44–48). А ещё одним идиомом, для которого у Н. Б. Бургановой указана та же особенность, является говор молькеевских кряшен, сформировавшийся на явном чувашском субстрате. Такое совпадение не выглядит случайным и даже независимо от вопроса о механизмах ассимиляции представляет дополнительный аргумент в пользу былого чувашского языкового элемента на Чепце. К возможности ассимиляции чепецко-чувашского населения среди татар отсылает и отмеченная выше нестабильность этнических маркировок: в документах XVII в. она, пусть и окказионально, затрагивает также соотношение *каринские чуваши ~ каринские татары*. При всей ограниченности выводов, которые можно извлечь из подобных обозначений, этот материал служит ещё одним аргументом в пользу того, что исчезновение чепецкой *чюваши* не следует сводить исключительно к её растворению среди бесермян.

Наряду с этим в чепецко-татарском достаточно многочисленны слова, на первый взгляд напоминающие сепаратные (как правило, отсутствующие в других татарских диалектах) чувашизмы, ср.: *umaj* ‘удобный’ (при тат. литер. *uŋaj* ‘тж.’ и ср.-чув. **ʊmaj* ‘удобный; удобство, порядок, лад’, о котором см. Савельев 2025: 143; регулярное развитие **ŋ* > *m* чепецко-татарским идиомам неизвестно); *bići* ‘родственник’ (при тат. литер. *bićä* ‘жена, замужняя женщина’ и ср.-чув. **bẹj-ćä* ‘старший родственник по мужской линии’, см. выше секцию 1, критерий xvi); *sańik* ‘вилы’ (при тат. *sänäk* и ср.-чув. **säńək* > чув. *sänək* ‘тж.’); *bələm* ‘блин’ (при тат. литер. *bələn* ← др.-рус. *блинъ* → ср.-чув. **bıləm* с характернейшей меной **-n* на **-m* после этимологического узкого гласного > чув. *pələm* ‘тж.’). Однако эти чувашизмы распространены и в удмуртском языковом массиве (*umoj* ‘хороший, правильный’; *beće* ‘родственник’; *sańik* ‘вилы’; *bi̯l’i̯m* ‘блин’), так что во всех приведённых случаях речь может идти уже о бесермянских или северноудмуртских заимствованиях в чепецко-татарском.

Таким образом, следы возможного чепецко-чувашского влияния в языках региона не составляют плотного слоя свидетельств, однако изолированные факты такого влияния обнаруживаются сразу в бесермянском, северном удмуртском, южном коми-зырянском и каринско-татарском, так что в совокупности они складываются в достаточно выразительную картину позднего присутствия чувашеязычного населения на Чепце. Эта ареальная конфигурация хорошо согласуется с рассмотренными выше письменными свидетельствами о присутствии в регионе сравнительно немногочисленного *чювашского* населения, а также с данными топонимики (на территории бывшей Вятской губернии — прежде всего, по левому берегу Вятки и, в частности, в бассейне Чепцы — хорошо известна топонимическая основа *Чуваш-*, см. Тепляшина 1968: 182–183). Указанная немно-

гочисленность, по-видимому, способствовала ассимиляции чепецких чувашей среди соседних народов, причём данные письменных источников скорее указывают на бесермянское (и, возможно, северноудмуртское) её направление, а предполагаемые следы лингвистического субстрата — на татарское. Тем самым исчезновение *чюваши* Чепцы как особой документально фиксируемой группы во второй половине XVII в. едва ли следует понимать как результат её растворения в каком-либо одном соседнем народе.

4. Булгаро-чувашские заимствования в общепермский диалектный континуум

Итак, теперь можно говорить о нескольких относительно надёжно выделяемых пластах пермских булгаризмов: древнечувашские заимствования в прапермский (VIII–IX вв.; см. Савельев 2025), среднечувашские заимствования в (обще)удмуртский (XIV–XV вв., см. секции 1–2 настоящей статьи); (поздне)среднечувашские заимствования в уже разошедшиеся диалекты языков пермской группы — по крайней мере завятские и северные говоры удмуртского языка, бесермянское наречие, южные коми-зырянские говоры (XV–XVII вв., см. секцию 3 настоящей статьи)[18]. Однако остаётся рассмотреть ещё одну группу лексем контактного происхождения, которые характеризуются тем, что они (а) широко распространены как в удмуртском, так и в коми (за пределами южнокоми ареала),

[18] Заслуживает особого рассмотрения – и потому не обсуждается детально в настоящей статье – проблема своеобразной группы булгаро-чувашских заимствований в пермских языках, распространение которых ограничено удмуртским и коми-пермяцким ареалами (без соответствий в коми-зырянском либо только со спорадической представленностью у южных коми-зырян). Ср. такие показательные случаи, как удм., к.-перм. *kab* ‘колодка для обуви’ ~ чув. *kap* ‘форма’ < ПТ **gēp*; удм. *turto* ‘оглобля, дышло’, к.-перм. *torta* ‘палка с плоской дужкой на конце (для скучивания зерна на току)’ ~ чув. *torda* ‘оглобля’ (см. детали под сноской 6 выше); удм. *śur*, к.-перм. *śor* ‘неродной, приёмный’ ~ чув. *śor(ə̂)* ‘половина, полу-; неродной, приёмный’ < ПТ **jār-*; удм. *suso*, к.-перм. *susa* ‘ткацкий челнок’ ~ чув. *ə̂°za*, (XVIII в.) *sə̂°za* ‘тж.’ (< ПТ **sus-*) и др. Эта дистрибуционная группа была подробно рассмотрена К. Редеи и А. Рона-Ташем (см. Rédei & Róna-Tas 1972: 281; 1975: 32–38, а также англоязычное изложение в Rédei & Róna-Tas 1983: 27–34), которые относят соответствующие слова к постпрапермскому слою булгаризмов, при этом коми-пермяцкие формы интерпретируют не как результат прямых контактов с булгаро-чувашским, а как заимствования, опосредованные удмуртским. В ряде случаев такое направление передачи обосновывается фонетическими и морфологическими признаками; особенно показателен случай удм. *ulmo*, к.-перм. *ulmę* ‘яблоко’ ~ чув. *olma* ‘тж.’ < ПТ **alma*, где к.-перм. *u* (при обычном в аналогичных случаях *o*) указывает на поздний удмуртизм. В конечном счёте К. Редеи и А. Рона-Таш распространяют эту интерпретацию и на остальные случаи, отрицая возможность контактов между булгаро-чувашским и предком коми-пермяцкого после распада прапермского языка. Иной интерпретации этого вопроса придерживается В. В. Напольских, связывающий повышенную представленность булгаризмов в коми-пермяцком по сравнению с коми-зырянским с относительно более плотными и длительными контактами между булгарами и населением Верхнего Прикамья. Контекстом для таких контактов могла служить вовлечённость этого ареала в систему булгарских укреплённых факторий, оставивших памятники типа Анюшкара и Рождественского городища, где археологически прослеживается булгарское население. Ср.: «…их рост начинается в IX-X вв. одновременно со сложением булгарского государства, упадок наступает в XIII в. в период монгольского нашествия и первого разгрома Булгарии, а к концу XIV в. практически все эти городища окончательно прекращают своё функционирование – одновременно с гибелью Волжской Булгарии как самостоятельного политического и экономического образования и исчезновением вместе с ней экономической, социально-политической и военной базы этих центров» (Напольских 2018: 73–75, 99–100). Нет необходимости по умолчанию противопоставлять две приведённые интерпретации, поскольку возможные сепаратные булгаро-пермские контакты вплоть до XIV в. в Верхнем Прикамье вовсе не исключают идеи о том, что отдельные слова булгаро-чувашского происхождения могли проникнуть в коми-пермяцкий через удмуртское посредство. Разграничение этих ситуаций потребует специального исследования.

при этом (b) соответствия между коми и удмуртскими формами в части случаев носят нерегулярный характер, что не позволяет возводить эти слова к прапермскому состоянию, и (c) в некоторых случаях фонетические особенности на стороне лексического донора либо реципиента указывают на более позднюю датировку контакта по сравнению с булгаро-пермскими контактами VIII–IX вв. Сочетание этих признаков позволяет поставить вопрос об идентификации пласта заимствований, источник которых с некоторой условностью может быть определён как «позднедревнечувашский», а на роль реципиента претендует находившийся в процессе распада общепермский диалектный континуум. Предварительно эти контакты могут быть отнесены к X–XIII вв.; основания для верхней границы такой датировки будут рассмотрены ниже. Необходимость выделять данный пласт заимствований наиболее наглядно иллюстрируют следующие этимологии.

- Удм. *bus* ‘туман, дымка, пар’, (бес.) *bə̑s* ‘туман’, коми *bus* ‘пыль’, (диал.) ‘мелкий дождь, изморось’ ← позднедр.-чув. **bus* (чув. *pə̑°s* ‘пар; угар’) < ПТ **bus*. Слово неизвестно в татарском, но документировано как *bŏθ* ‘пар’ в башкирском, т. е. в общем в поволжско-кыпчакском ареале известно, и нельзя исключать, что ранее оно было распространено и в татарских говорах. Кроме того, ср. рус. (диал.: архангельские, вологодские, вятские, пермские, уральские, сибирские говоры) *бус* ‘мелкий дождь; туман, изморось; мучная пыль на мельницах, идущая на корм скоту’ (СРНГ 3: 302). Выяснение отношений между формами этого слова в разных языках давно вызывает трудности. Ю. Вихманн сначала считал удмуртское и коми слова непосредственно связанными друг с другом, но впоследствии поменял позицию и предложил объяснять коми слово как заимствование из русского, в то время как для удмуртского принималось чувашское происхождение (Wichmann 1903: 50). Между тем, согласно А. Е. Аникину (РЭС 5: 201–202), русское слово само заимствовано из пермских или напрямую из тюркских. Точечно (на Урале, в Сибири) тюркизм действительно не исключён, однако в целом тюркское происхождение выглядит сомнительным, учитывая явное тяготение ареала русского слова к Русскому Северу. Если его источником является коми, то сама коми форма (как и её удмуртская параллель) должна быть признана достаточно поздним чувашским заимствованием. Отвергаемые при этом альтернативы — считать её заимствованной из поволжско-кыпчакского (но тогда это был бы исключительный случай прямого татаризма с широкой дистрибуцией в коми, см. Савельев 2025: 149), либо унаследованным из прапермского состояния заимствованием из древнечувашского (ср. КЭСК: 42–43). В последнем случае, однако, совершенно неясна вокалическая субституция, поскольку ожидалось бы ППерм **bus* > **bi̯s* > коми, удм. ***bi̯s* (Савельев 2025: 159). Естественнее считать удм., коми *bus* продолжениями заимствования из позднедр.-чув. **bus*, уже не подпадавшего под действие ППерм развития **u* > **i̯*. Из среднечувашского источника заимствовано мар. **pʊs* ‘запах, угар; пар’.
- Удм. (+бес.) *iśke*, (диал.) *śike* ‘значит, стало быть, в таком случае’, коми *eśkẹ*, (диал.) *veśkẹ* ‘бы, же’ ← позднедр.-чув. **ɪś=ke* (чув. *-(ə)śkä* ‘ведь; оказывается, разумеется и т. п.’) (Wichmann 1903: 59–60). Чувашская частица является сложением самостоятельно употребительных дискурсивных частиц *əś*, *-ś* ‘ведь, действительно; оказывается’ и *kä* ‘частица, значение которой пока недостаточно ясно: иногда выражает побуждение (?); при уст<упительном> накл<онении> в протазисе и буд<ущем> изъяв<ительном> с вопр<осом> — «и», в аподозисе — как бы безразличное отношение вместе с предупреждением; в вопросах — живость и заинтересованность’ (СЧЯ VI: 187). Первый элемент сложения этимологически тождествен тат. *ić* ‘ведь, же’, которое считается в Әхмәтьянов 2015, I: 317 вариантом тат. *hić*, диал. *ić* ‘никак; ни-’ (ср. также чув. *əś* ‘совсем не’) ← персидская отрицательная частица *hić*, хотя детали семантического развития в таком слу-

чае нуждаются в прояснении[19]. Частица *-kä* не имеет этимологии. Чувашизм в удмуртском надёжен; при этом коми слово трудно отделять от удмуртского. Между тем формально в коми частице возможно выделять в качестве второго элемента *kę* ‘если; видимо, может быть’; в таком случае приходится предполагать здесь сложение на собственной почве, а вопрос стоит только о связи между чув. *əś*, *-ś* и коми *(v)eś-*. Семантических препятствий для этого нет; коми-удмуртское вокалическое соответствие нетривиально, но имеет параллели в общепермской лексике (ср. Понарядов 2014: 33, где примеры с коми *e* ~ удм. *i* составляют одну из рубрик категории «Неясные случаи [развития гласных в пермских языках]»). Возможно, особое соответствие обусловлено в этом и большинстве других похожих случаев позицией гласного рядом с палатальным согласным. Заимствование в пермские выглядит довольно старым, но отражение в пермском материале **ś* идиома-источника при этимологическом **ć* препятствует отнесению этого случая к древнейшему пласту булгаро-пермской контактной лексики (см. Савельев 2025: 158 и в настоящей статье секцию 1, критерий xvi).

- Удм. (+бес.) *uś-ja-* ‘открывать’, удм. *uś-t-* ‘открыть’, коми *voś-t-*, (диал.) *o̭ś-t-*, *veś-t-*, к.-перм. *oś-t-*, к.-язьв. *i̮ś-t-* ‘открыть’ ← позднедр.-чув. **åś-* ‘открывать, открыть’ (Wichmann 1903: 123–124; Rédei & Róna-Tas 1983: 14). Чув. < ПТ **ać-*, откуда и тат. *ać-* ‘тж.’ Пермский суффикс **-ja-* является мультипликативным (множество повторяющихся ситуаций), а **-t-* — семельфактивным (единичная ситуация) (Кузнецова и др. 2013: 536). Полностью аналогичное вокалическое соответствие наблюдается в исконной основе удм. *uj*, коми *voj*, к.-перм. *oj*, к.-язьв. *i̮j* ‘ночь’ < ПУ **e(x)ji ~ *ü(x)ji* ‘тж.’ (ср. Понарядов 2014: 13); т. е. заимствование в пермские выглядит довольно старым. При этом отражение в пермском материале **ś* идиома-источника в соответствии с ПТ **ć* препятствует отнесению этого заимствования к древнейшему пласту булгаро-пермской контактной лексики.
- Удм. *śerį* ‘шпулька, цевка’, (бес.) *śerə̑* ‘шпулька, катушка (для ниток)’, коми *śuri* ‘уток, намотанный на цевку’, (диал.) *śuri* ‘цевка, шпулька (для наматывания пряжи)’, к.-перм. *śuri* ‘цевка (катушка для пряжи)’ ← позднедр.-чув. **śürə* (чув. *śə°rə°* ‘кольцо; цевка для тканья’). Чув. < ПТ **jüɻük*, откуда и тат. *jə̑zə̑k* ‘перстень’. Из чувашского источника также тат. *šürə* ‘цевка’, откуда обратное заимствование — чув. (диал.) *šürə* ‘цевка, шпулька (на которую наматывают нитки)’. Соответствие коми *u* ~ удм. *e* проблематично (ср. категоризацию данной основы вместе с ещё двумя примерами на то же соответствие под рубрикой «Неясные случаи» в Понарядов 2014: 33). На основании этой нерегулярности, а также ввиду нехарактерного для старых булгаризмов сохранения ауслаутного узкого гласного в коми двусложной основе данное слово уже у Ю. Вихманна (Wichmann 1903: 98–99, 139) и далее у Д. Лако (Lakó 1935: 170) рассматривается как постпраперм-ское заимствование из источника чувашского типа. Иначе (как заимствование в общий предок коми и удмуртского) — в Лыткин 1964: 225; КЭСК: 274. Ю. Вихманн предполагал, что чувашское слово могло сначала проникнуть в коми-пермяцкий, а оттуда — в южные коми-зырянские говоры; эта точка зрения принята и в Rédei & Róna-Tas 1983: 25.
- Удм. *śart-ćį*, (бес.) *ćar-ćə̑* ‘репа’, коми *śork-ńi*, к.-перм. *śort-ńi* ‘тж.’ ← позднедр.-чув. **śäw(ə)r-ək* (чув. *śarə̑k* ‘репа’) (Wichmann 1903: 97; КЭСК: 253). Предковое чув. **śäw(ə)r-ək* букв. ‘кругляшок’ производно от **śäwər-* (чув. *śavə̑r-* ‘крутить, вертеть’; ср. по той же мотивационной модели чув. *śavr-aga* ‘круглый’, *śavr-ə̑m* ‘круг’, *śavr-aška* ‘кружок’). Современное чув. *śarə̑k* отражает стяжение комплекса *-avə̑-* в многосложных словах, ср. ана-

[19] В татарском есть образование *ićmasa* ‘хотя бы; даже’, являющееся стяжением из *(h)ić=bulmasa* ‘по крайней мере’ (букв. ‘как бы то ни было’). Может быть, *ić* ‘ведь, же’ стоит считать как бы сокращённым вариантом этой или иной конструкции, синхронно не имеющей негативного компонента в семантике, но этимологически включающей в качестве первого элемента *(h)ić* ‘ни-’.

логичное развитие, например, в *arman* ‘мельница’ (этимологически **avə̂r-man*, производное от *avə̂r-* ‘молоть’). Номинация ‘кругляшок’ > ‘корнеплод; растение со съедобным корнеплодом’ тривиальна. Изложенная в Rédei & Róna-Tas 1983: 39 критика гипотезы о булгарском происхождении пермского слова представляется необоснованной. Другие примеры на соответствие удм. *a* ~ коми *o* см. в Понарядов 2014: 10; не исключено, что в данном случае оно проистекает из сочетания «открытый гласный + **w*» в идиоме-источнике. При этом отражаемый в пермском материале **ś* на месте ПТ **ć* исключает возможность отнесения этого заимствования к древнейшему пласту булгаро-пермской контактной лексики. Иная этимология чувашского слова обосновывается в Напольских 2018: 501–506.[20]

Если нижняя хронологическая граница данного пласта определяется его непринадлежностью к древнейшему слою булгаро-пермской контактной лексики VIII–IX вв., то верхняя граница не может быть установлена без обращения к вопросу о среднемонгольских заимствованиях в пермских языках.

Проникновение среднемонгольской лексики в Волго-Камский регион надёжно связывается с монгольскими завоеваниями 1230-х гг. Монголизмы в большом количестве присутствуют в татарском языке (см. представительный набор примеров в Әхмәтьянов 2015), а через татарский они могли проникать в чувашский и соседние финно-угорские языки (включая удмуртский). Помимо этого, в чувашском языке имеются десятки таких монголизмов, которые синхронно отсутствуют или очень узко представлены в татарском. Н. Поппе (Poppe 1977) полагал, что и в таких случаях речь следует вести о былом татарском посредничестве при заимствовании из среднемонгольского источника (даже в отсутствие формальных признаков татаризмов), однако А. Рона-Таш обосновал возмож-

[20] Отдельно следует остановиться на двух основах, которые нередко безоговорочно включаются в список булгаризмов пермских языков, но достаточно плохо вписываются даже в выделяемый здесь пласт «позднедревнечувашских» заимствований в общепермский диалектный континуум. Если для большинства рассмотренных выше этимологий такая атрибуция позволяет дать приемлемое объяснение наблюдаемым фонетическим отклонениям, то в двух следующих случаях проблемы субституции остаются довольно принципиальными; в сочетании с не совсем точным семантическим соответствием между булгарским и пермским материалом даже сам статус булгарских заимствований здесь может быть поставлен под сомнение. Во-первых, ср. случай удм. (+бес.) *śil’* в составе удм. *śil’=tẹl*, (бес.) *śil’=tɔ̈l* ‘буря, ураган’, коми *śi̮v*, (диал.) *śiv*, *śil* ‘тж.’ — при чув. *śil* ‘ветер’ < ср.-чув., др.-чув. **śẹl* < ПТ **jẹl* ‘тж.’ (Wichmann 1903: 99–100). Во всех иных известных случаях – и в том числе в палатализующих контекстах – булг. **ẹ* отражается как удм. *e*, ср. этимологии удм. *śetị* ‘накосник’ и *beće* ‘родственник’ (секция 1, критерий xvi), а также удм. *veme* ‘по́мочь’ (секция 1, критерий xviii). Особая «сужающая» субституция через удм. *i* в принципе может быть обусловлена положением гласного не просто рядом с палатальным, а между двумя палатальными согласными. При этом удм. *śil’=tẹl* ‘буря’ выглядит как полукалька с чув. *śil=tə̂°və̂°l* ‘тж.’ (букв. ‘ветер-буря’; материальное сходство вторых элементов сложения здесь случайно). Таким образом, для удм. *śil’* ‘*ветер’ чувашское происхождение очень вероятно, но не следует ли отделять от него фонетически нерегулярно ему соответствующее коми *śi̮v* ‘буря’, связывая последнее напрямую с венг. *szél* ‘ветер’, традиционная булгарская этимология которого компрометируется практически уникальным появлением венг. *sz-* на месте ПТ **j-* (ср. Дыбо 2007: 31) при обычной субституции через венг. *gy-*? По крайней мере формально венг. *szél* и коми *śi̮v* могут быть возведены к ПУ **ćäli*. Конечно, возможность такой интерпретации, предполагающей только случайное сходство между удмуртским и коми словом, требует дальнейшего обоснования. Во-вторых, ср. случай удм. *śek-ɟt*, коми *śe̮k-ɟd* ‘тяжелый’ – при чув. *śə°k* ~ *śə̂°k* ‘воз, груз, ноша’ < ПТ **jük* ‘тж.’ (Wichmann 1903: 97–98; Rédei & Róna-Tas 1975: 32; 1983: 11–12). Уже у Ю. Вихманна сопоставление даётся под вопросом; К. Редеи и А. Рона-Таш также считают это сопоставление фонетически проблематичным из-за необычного соответствия широкого неогубленного гласного как в коми, так и в удмуртском этимологически узкому огубленному гласному в чувашском. По-видимому, такая субституция действительно уникальна. О самом соответствии удм. *e* ~ коми *e̮* см. выше секцию 2 под обсуждением удм. *keć* ‘коза’ ~ коми *ke̮ć* ‘заяц’.

ность идентификации прямых монгольских заимствований в чувашском (Róna-Tas 1975; 1982). Подобные монголизмы также попадали из чувашского в финно-угорские языки — прежде всего, в марийский и удмуртский — в рамках контактов послемонгольского времени. В контексте настоящего исследования важно подчеркнуть, что в удмуртском имеется достаточно представительный корпус заимствованной чувашской лексики монгольского происхождения. Выше в секции 1 уже были рассмотрены в этом ключе удм. *bẹd'ono* ‘перепёлка’ (под критерием xx), удм. *kudo* ‘сват’ (критерий xxii) и (?) удм. *veme* ‘пóмочь’ (критерий xviii). Однако этот список следует расширить, по крайней мере, за счёт следующих этимологий.

- Удм. *kandelem*, (диал.) *kemdel'e*, *kangelem* ‘свидетель’ ← ср.-чув. **kVndäläN* (чув. *kündälän*, диал. *kəndäläm*, *kəndälän*, *kəndäm* ‘свидетель’). В чувашском слово производно от глагола *kündälä-* ‘свидетельствовать’, который, в свою очередь, заимствован из ср.-монг. рефлекса монг. **kündü-le-* ‘уважать, почитать, чествовать’ < **kündü* ‘тяжелый; уважение, почёт’. Вопреки ЭСЧЯ 1: 322–323, не исконными соответствиями к чувашскому слову, а параллельными монголизмами являются похожие формы в тюркских языках Сибири (телеут., кюэрик. *kündülä-* ‘оказать почет, с почетом угостить’, якут. *kündüle:-* ‘угощать, почитать, оказывать гостеприимство’ и т. д.). Ср. Wichmann 1903: 71.
- Удм. *muglo*, (диал.) *mugilo*, *mugil'o*, *moklok*, *muglok* ‘комолый’ ← ср.-чув. **mʊGə̑la* и т. п. (чув. *mə̑°kla*, диал. *mə̑kka*, *mə̑°kl'ä*, *mə°kl'ä* ‘комолый’), ср. чув. *mə̑°ga* ~ *moga* ‘тупой, тупоконечный’, *mə̑°gə̑°l'*, *mə̑°kl'ä*, *mə°kl'ä* и т. п. ‘нарост, шишка’. Чувашские формы отражают заимствование из ср.-монг. рефлекса монг. **mokV-* ~ **mogV-* ‘тупой; комолый’ (Nugteren 2011: 445), ср. халх. *moxō* ‘тупой’, *moγo-tor* ‘комолый; тупой, тупоконечный’ и т. п. Вопреки Róna-Tas 1982: 103–104, нет оснований считать чувашское слово татаризмом (ср. тат. *mŏqŏt*, *mŏqə̑j* и т. п. ‘тупица’). Чувашским заимствованием является мар. **mʊgə̑la* ‘комолый’. Невероятны построения ЭСЧЯ 1: 347–348, где чув. *mə̑°ga* ‘тупой’ на основании сравнения с указанным монгольским этимоном и тунгусскими параллелями возводится к праалтайскому состоянию, а чув. *mə̑°gə̑°l'* ‘нарост’ считается заимствованием из марийского (хотя при этом и приводятся похожие монгольские формы).
- Удм. *tingil'i*, *tị̈ngil'i* ‘ось’ ← ср.-чув. **tɪŋ(g)ələ* (чув. *tənəl*, диал.: СЗ *tẹŋgəl*, *tiŋgəl* ‘ось’). В тюркологической литературе чувашское слово возводится к ПТ **deŋgil* (EDAL: 1365; СИГТЯ 2006: 69), однако практически все тюркские формы указывают на **i* в первом слоге, ср. тур. *dingil*, тат. (диал.) *təŋəl* и т. п. При этом к гипотетическому ПТ **diŋgil* чувашское слово восходить также не может, поскольку в таком случае ожидалось бы ***ćənəl* по второй палатализации (см. секцию 1, критерий xvii). Лучше усматривать в чувашском заимствование из ср.-монг. *teŋgeli(g)* ‘ось’, о котором см. Nugteren 2011: 518. Позднему происхождению чувашского слова как будто противоречит венг. *tengely* ‘ось’, которое традиционно считается булгаризмом, но А. В. Дыбо (2010: 97) допускает для него и раннекыпчакское происхождение.

Совершенно иначе обстоит дело с предполагаемыми монголизмами в коми. Согласно С. К. Белых (Belykh 2007; Белых 2009: 100–102), среди тюркских слов среднемонгольского происхождения выделяется исключительно компактная группа лексем (от 2 до 4), которые проникли не только в удмуртский язык, но и в коми. Поскольку среднемонгольская лексика не могла попасть в языки Волго-Камского региона ранее монгольских завоеваний 1230-х гг., С. К. Белых интерпретирует подобные коми-удмуртские параллели как свидетельство того, что процесс распада общепермского диалектного континуума в это время ещё не был завершён. Однако даже эти несколько этимологий выглядят очень проблематичными.

- Коми (+к.-перм.) *dugd-* ‘остановиться, перестать, прекратиться’ ~ удм. (+бес.) *dugd-* ‘тж.’ ← ? тюрк. ← ? ср.-монг. рефлекс монг. **togta-* ‘быть устойчивым; стоять спокойно; останавливаться’ (Belykh 2007: 33–34; Белых 2009: 100–101). Ср. бес. *tukta-* ‘тж.’ — надежное заимствование той же монгольской основы через кыпчакское (тат. *tuqta-*) или чувашское (*tə̂°xta-*) посредство. Фонетические субституции в случае «тюрко-монгольского» происхождения *dugd-* неясны (сомнительность этой этимологии признаёт и С. К. Белых). Вместе с тем следует отметить, что и уральская этимология пермского слова (КЭСК: 97; Sammallahti 1988: 554) не может быть принята (Kuokkala 2018: 31–32).
- Коми (диал.: верхневычегодские говоры) *koń=koń*, *końe=końe* ‘возглас, которым подзывают телят’ ~ ? удм. *kuńan* ‘телёнок’ ← ? тюрк. ← ср.-монг. рефлекс монг. **gunan* ‘трёхлетний (о животном)’ (Belykh 2007: 33–34; Белых 2009: 101). Среднемонгольский, действительно, послужил источником для тат. *qŏnan* ‘трёхлетняя лошадь (реже — корова)’ (Әхмәтьянов 2015, I: 436–437). Однако связь пермского слова с этим гнездом не выглядит очевидной (в части, касающейся коми, это признаёт и С. К. Белых). Инлаутный *-ń-* в коми и удмуртском может быть объяснён разве что на экспрессивной почве; кроме того, и семантический переход от ‘*трёхлетней лошади (реже коровы)’ к ‘телёнку’ выглядит достаточно нетривиальным. С. К. Белых приводит в связи с данным случаем тат. (диал.) *qŏnna-* ‘жеребиться’, что могло бы несколько упростить семантическую сторону сравнения (по крайней мере, объясняя переход к семантике ‘*детеныш копытного’ в пермских), однако это татарское слово принадлежит к другой этимологии (тат. *qŏlŏn*, диал. *qŏn* ‘жеребёнок’ < ПТ **quluN* ‘тж.’).
- К.-перм. *ćeber* ‘гордый, своенравный’, (диал.) *ćebe̮r* ‘хвастливый; красивый; гордый’ ~ удм. *ćeber* ‘красивый’ ← тюрк. ← ср.-монг. *ćeber* ‘красивый; приятный’ (Belykh 2007: 33–34; Белых 2009: 100–101). Согласно С. К. Белых, данная параллель «не вызывает особых сомнений», заметим, однако, что восстановление цепочки контактов именно в таком виде — от монгольского через тюркские к общепермскому диалектному континууму — не выглядит очевидным. Хотя тюркское слово (ср., в частности тат. *ćibär* ‘красивый’ → чув. *ćibär* ‘тж.’) традиционно считается монгольским по происхождению, Rédei & Róna-Tas 1983: 28 указывают на ненадёжность такой интерпретации, отмечая широкую дистрибуцию в тюркских, отсутствие внутренней монгольской этимологии и необъяснимость на монгольской почве форм типа туркм. *ćeper* (с глухим *-p-*!). Узкая дистрибуция слова в коми ареале (только коми-пермяцкий) вовсе не требует предположения о заимствовании в общепермское состояние, а скорее объясняется существованием в послемонгольское время особой контактной зоны в районе расселения коми-пермяков, которая могла затрагивать удмуртские, а также поволжско-кыпчакские и/или булгаро-чувашские идиомы (см. выше сноску 18).
- Коми ***ćid-: ćidje̮=vadje̮ voli̮ni̮* ‘быть в нужде, безвыходном положении’, к.-перм. *ćid-* ‘беспокоиться, сокрушаться’, к.-язьв. *ćide̮m* ‘злость, обида’ ~ удм. *ćid(-)a(l)-* ‘терпеть, выносить’ ← тюрк. ← ср.-монг. *ćida-* ‘мочь, быть в состоянии’ (Belykh 2007: 33–34; Белых 2009: 100–101). Монгольское происхождение чув. *ćə̂t-* ‘терпеть’ не вызывает сомнения; поволжско-кыпчакское посредство здесь допустимо, но вовсе не обязательно (основа *ćə̂da-* ~ *cə̂da-* не является общетатарской, будучи представленной только в мишарских и сибирско-татарских говорах, см. Әхмәтьянов 2015, II: 467). С учётом этого (средне)чувашский выглядит наиболее вероятным источником удмуртского слова. Приходится заметить, что в целом данная монгольская по происхождению основа имеет стабильное значение типа ‘терпеть, выносить, выдерживать’ во всех языках Волго-Камского региона (ср. также мар. **ćıtə-* ‘тж.’; вероятно, сюда же и эрз. *cid’-ard(o)-* ‘выдерживать’); на этом фоне значительно отличающаяся семантика коми форм вызывает некоторые сомнения в их

принадлежности к тому же этимологическому гнезду. Но развитие значений типа ‘беспокоиться; обижаться’ из ‘*терпеть’, конечно, не исключено, так что этот случай выглядит наименее проблематичным из всего представленного списка.

Таким образом, практически в каждом из приведённых случаев монгольская этимология для коми слова выглядит сомнительной; даже если какие-то из этих форм и являются, действительно, монгольскими по происхождению, общее число монголизмов в коми оказывается исчезающе малым. Особенно показателен этот факт на фоне достаточно многочисленных монголизмов в удмуртском, в том числе надёжных заимствований через чувашское посредство. Резкая асимметрия в распределении этого хронологически маркированного слоя служит сильным свидетельством в пользу того, что к середине XIII в. общепермский диалектный континуум уже фактически распался (что тем самым определяет эпоху монгольских завоеваний как верхнюю границу для рассматриваемого здесь пласта «позднедревнечувашских» заимствований), а связи коми языка с Волго-Камским языковым ареалом в целом сошли на нет. Отдельные исключения не противоречат этому выводу, поскольку имеют локальный характер и объяснимы сепаратными контактами послемонгольского и даже более позднего времени: это упомянутая зона соприкосновения коми-пермяцкого ареала с удмуртским и поволжско-тюркским ареалом, а также связи южных коми-зырян с Вятско-Чепецкой контактной зоной (см. выше секцию 3). Если известные «особые отношения» (Белых 2009: 91), связывающие, с одной стороны, южнокоми ареал и удмуртский язык, а с другой стороны — северноудмуртский ареал и коми язык, и восходят частично к эпохе общепермского диалектного континуума, то возникновение соответствующих изоглосс следует датировать периодом до монгольских завоеваний. Остальные же изоглоссы требуют объяснения в свете коми-удмуртских контактов более позднего времени (ср. Максимов 2016; 2018а; 2025).

* * *

Проведённый анализ позволяет существенно детализировать картину стратификации булгаро-чувашских заимствований в пермских языках. Наряду с уже идентифицированным в первой части исследования (Савельев 2025) пластом древнечувашских заимствований в прапермский язык VIII–IX вв. выделяются: позднедревнечувашские заимствования X–XIII вв., распространявшиеся в условиях уже распадавшегося общепермского диалектного континуума; многочисленный пласт среднечувашских заимствований в (обще)удмуртский XIV–XV вв.; наконец, отдельные зоны — Арская сторона, Чепецкий край — поздних (XV–XVII вв.) контактов среднечувашских говоров с уже обособившимися диалектами пермских языков. Представленная модель позволяет объяснить как нетривиальные коми-удмуртские соответствия, не укладывающиеся в реконструкцию единого прапермского прототипа, так и случаи последовательного обновления прапермских булгаризмов более поздними заимствованиями в отдельных пермских языках (как правило, за счёт среднечувашской лексики в удмуртском и за счёт русской — в коми).

Таким образом, можно говорить о практически непрерывных булгаро-пермских языковых связях с начала переселения булгар в Волго-Камский регион (самый конец VII в.) и до XVI–XVII вв., когда чувашское население Арской стороны постепенно ассимилировалось среди татар и одновременно вероятный чувашский этноязыковой компонент на Нижней Чепце также растворился среди окружающих народов. Ныне зоны компактного сельского расселения чувашей и удмуртов соприкасаются только в двух микроареалах: в Бавлинском районе Татарстана (где удмурты присутствуют с XVII в., чуваши — с XVIII в.),

а также в Граховском районе Удмуртии (куда чуваши переселились в связи со столыпинской реформой в 1911–1914 гг.). В остальном чувашский и удмуртский ареалы оказались разорваны, прежде всего, вследствие интенсивной кыпчакизации Волго-Камского региона в эпоху Казанского ханства и первых веков существования Среднего Поволжья в составе Российского государства, что в целом и обусловило прекращение длившихся почти тысячу лет булгаро-пермских контактов.

Сокращения

алт. – алтайский язык; араб. — арабский язык; балк. — балкарское наречие; башк. — башкирский язык; бес. — бесермянское наречие удмуртского языка; булг. — булгарская ветвь тюркских языков; венг. — венгерский язык; верх. — верховой диалект чувашского языка; вост.-слав. — восточнославянские языки; горн. — горномарийское наречие; диал. — диалектная форма; др.-рус. — древнерусский язык; др.-тюрк. — древнетюркский язык; др.-чув. — древнечувашский язык; казах. — казахский язык; караим. — караимский язык; к.-перм. — коми-пермяцкий язык; к.-язьв. — коми-язьвинское наречие; кумык. — кумыкский язык; кыпч. — кыпчакская группа тюркских языков; кюэрик. — кюэрикское наречие; лив. — ливский язык; литер. — литературная форма; луг. — луговомарийское наречие; мар. — марийский язык; монг. — монгольская семья языков; низ. — низовой диалект чувашского языка; осм. — османско-турецкий язык; пер.-ю. — периферийно-южные говоры удмуртского языка; ПИЕ — праиндоевропейский язык; ППерм — прапермский язык; ПТ — пратюркский язык; ПУ – прауральский язык; рус. — русский язык; СЗ — северо-западное наречие марийского языка / северо-западный диалект чувашского языка; сев. — северное наречие удмуртского языка; ср.-монг. — среднемонгольский язык; ср.-чув. — среднечувашский язык; сред. — срединные говоры удмуртского языка; ст.-слав. — старославянский язык; тат. — татарский язык; телеут. — телеутский язык; тур. — турецкий язык; тюрк. — тюркская семья языков; удм. — удмуртский язык; узб. — узбекский язык; уйг. — уйгурский язык; уст. — устаревшая форма; УТ — узкотюркская ветвь тюркских языков; халх. — халха-монгольский язык; чеп. — чепецкий диалект татарского; чув. — чувашский язык; ю. — южное наречие удмуртского языка; эрз. — эрзя-мордовский язык; эст. — эстонский язык; якут. — якутский язык; CAR — каритив (лишительная форма); COLL — собирательный показатель; DAT.ACC — датив-аккузатив (дательно-винительный падеж); DIM — диминутивная форма; PL — множественное число; PST — прошедшее время; SUBST — субстантиватор; VOC — вокатив (звательная форма); 3 — третье лицо.

Литература

Әхмәтьянов, Р. Г. 2015. *Татар теленең этимологик сүзлеге. Ике томда: I том (А–Л); II том (М–Я).* Казан: Мәгариф–Вакыт.

Белых, С. К. 2009. *Проблема распада прапермской этноязыковой общности*. Ижевск: Удмуртский госуниверситет.

Бурганова, Н. Б. 1962. Говор каринских и глазовских татар. В: *Материалы по татарской диалектологии* 2: 19–56. Казань: «Татполиграф».

Гришкина, М. В. 1988. Служилое землевладение арских князей в Удмуртии XVI — первой половины XVIII веков. В: В. Е. Владыкин и др. (ред.). *Проблемы аграрной истории Удмуртии*: 20–35. Ижевск: Удмуртский институт истории, языка и литературы.

Гришкина, М. В. 1989. Новые грамоты арским князьям на Вятке. *Советские архивы* 1989(4): 67–71.

Дыбо, А. В. 2007. *Лингвистические контакты ранних тюрков. Лексический фонд. Пратюркский период*. Москва: Восточная литература.

Дыбо, А. В. 2010. Вокализм раннетюркских заимствований в венгерском. В: Valentin Ju. Gusev, Anna Widmer, Annkatrin Klein (ред.). *Gedenkschrift für Eugen A. Helimski. Finnisch-Ugrische Mitteilungen, Bd. 32/33, Jahrgang 2008/2009*: 83–132.

Дыбо, А. В. 2011. Отражение булгарских палатализаций в булгаризмах венгерского языка. В: Кузнецов А. В. (ред.). *Чувашский язык и этнос в истории евразийской цивилизации. Материалы Международной тюркологической конференции*: 10–21. Чебоксары: ЧГИГН.

Дыбо, А. В. 2013. *Этимологический словарь тюркских языков. Том 9 (дополнительный)*. Астана: ТОО «Prosper Print».

Дыбо, А. В. 2021. Булгарские заимствования в славянских языках: особенности вокализма. *Русин* 66: 183–206.

Идрисов, Р. И. 2013. *Тюркские заимствования разных периодов в бесермянском диалекте удмуртского языка*. Москва: Издательство МГУ.

Кельмаков, В. К. 1985. К вопросу о булгарском субстрате в языке бесермян. В: *Вопросы диалектологии тюркских языков*: 114–122. Уфа: БФАН СССР.

Кельмаков, В. К. 1998. *Краткий курс удмуртской диалектологии: Введение. Фонетика. Морфология. Диалектные тексты. Библиография*. Ижевск: Изд-во Удм. ун-та.

Кузнецова, А. И., Н. В. Сердобольская, М. Н. Усачева, О. Л. Бирюк, Р. И. Идрисов. 2013. *Словарь бесермянского диалекта удмуртского языка*. Москва: ТЕЗАУРУС.

КЧРС = Золотницкий, Н. И. 1875. *Корневой чувашско-русский словарь*. Казань: Типография Императорского Университета.

КЭСК = Лыткин, В. И., Е. С. Гуляев. 1970. *Краткий этимологический словарь коми языка*. Москва: «Наука».

Луппов, П. Н. 1958. Удмурты в XV–XVII веках. Документы по истории Удмуртии XV–XVII веков. Ижевск: Удмуртское книжное издательство.

Лыткин, В. И. 1964. *Исторический вокализм пермских языков*. Москва: Наука.

Лыткин, В. И. 1967. О древнетюркских элементах в лексике пермских языков. *Вопросы финно-угорского языкознания* IV: 131–142.

Лыткин, В. И., Е. С. Гуляев. 1975. Дополнения к краткому этимологическому словарю коми языка. *Коми филология. Труды ИЯЛИ КФАН СССР* 18: 3–45.

Максимов, С. А. 2008. Самоназвания удмуртов и бесермян, их варианты, названия соседних народов в языке удмуртов и бесермян. *Linguistica Uralica* 44(2): 108–118.

Максимов, С. А. 2014. Возможности и пути обогащения литературного языка лексикой маргинальных диалектов (на примере удмуртского языка). *Ежегодник финно-угорских исследований* 2014(5): 92–101.

Максимов, С. А. 2016. О коми заимствованиях в удмуртском языке. *Linguistica Uralica* 52(2): 111–121.

Максимов, С. А. 2018a. Северноудмуртско-коми ареальные языковые параллели: лексика, фонетика, морфология. Ижевск: УИИЯЛ УдмФИЦ УрО РАН.

Максимов, С. А. 2018b. Тюркско-удмуртские лексические метаморфозы, или о некоторых словах со «странной» внутренней структурой в удмуртском языке. *Ежегодник финно-угорских исследований* 12(3): 30–46.

Максимов, С. А. 2019. О терминах родства по материнской линии в удмуртском языке: источник типологической модели. Вестник ЧГПУ им. И.Я. Яковлева 3(103): 126–134.

Максимов, С. А. 2025. Бесермянско-коми изоглоссы как свидетельство былых контактов между народами. *Linguistica Uralica* 61(2): 89–105.

Мудрак, О. А. 1993. *Исторические соответствия чувашских и тюркских гласных: опыт реконструкции и интерпретации*. Москва: Институт востоковедения РАН.

Мудрак, О. А., Ф. Г. Хисамитдинова. 2018. *Кыпчакские языки Урало-Поволжья*. Астана: Ғылым.

Мусихин, А. Л. 2013. Грамоты каринским арским князьям XVI в.: история публикации и изучения. *Ежегодник финно-угорских исследований* 2013(4): 82–90.

Напольских, В. В. 2018. *Очерки по этнической истории*. Казань: Институт археологии им. А. Х. Халикова Академии наук Республики Татарстан.

Напольских, В. В. 2021. Удмуртская *акашка* ~ бесермянская *акаяшка* и позднесредневековая этническая история Нижнего Прикамья. *Этнография* 4(14): 37–54.

Напольских, В. В. 2025. К этимологии русск. *безмéн*. *Вопросы языкознания* 2025(1): 136–147.

Напольских, В. В. 2026. О времени и исторических обстоятельствах пермской деназализации. *Томский журнал лингвистических и антропологических исследований* 51(1): 29–39.

Насипов, И. С. 2010. Финно-угорские заимствования в татарском языке: синопсис и таксономия: дисс. д. ф. н. Казань.

Останина, Т. И. 2021. Гординский камень 1323 г.: новые сведения. Поволжская археология 38(4): 179–190.

Понарядов, В. В. 2014. *Развитие гласных в пермских языках*. Сыктывкар: Коми научный центр УрО РАН.

РЭС 5 = Аникин, А. Е. 2017. *Русский этимологический словарь. Выпуск 5 (буба I — вакштаф)*. Москва: Языки славянской культуры.

Савельев, А. В. 2023. Комментарии к чувашским материалам Г. Ф. Миллера. *Российская тюркология* 40–41: 24–52.

Савельев, А. В. 2025. О критериях выявления и стратификации булгаро-чувашских заимствований в пермских языках I. *Вопросы языкового родства* 23(1–2): 137–167.

Савельев, А. В., Н. С. Мясников. 2026. *К происхождению этнонима чуваш: основные версии и их историческая интерпретация. (Научные доклады ЧГИГН 42.)* Чебоксары: Чувашский государственный институт гуманитарных наук.

Серебренников, Б. А. 1963. *Историческая морфология пермских языков*. Москва: Издательство АН СССР.

СИГТЯ 2006 = Тенишев, Э. Р., А. В. Дыбо (отв. ред.). 2006. *Сравнительно-историческая грамматика тюркских языков: Пратюркский язык-основа. Картина мира пратюркского этноса по данным языка*. Москва: Наука.

СРНГ 3 = Филин, Ф. П. (гл. ред.). 1968. *Словарь русских народных говоров. Вып. 3: Блазнишка — Бяшутка*. Москва / Ленинград: Наука.

СЧЯ = Ашмарин, Н. И. 1928–1950. *Словарь чувашского языка. Т. I–XVII*. Казань–Чебоксары.

Тепляшина, Т. И. 1968. О смешении терминов «чуваши» и «бесермяне» в письменных источниках. В: *Ученые записки ЧНИИЯЛиЭ. Вып. 40: История, этнография, социология*: 177–187. Чебоксары: Чувашкнигоиздат.

Тепляшина, Т. И. 1970. *Язык бесермян*. Москва: Наука.

ТТДС = Махмутова, Л. Т. (гл. ред.). 1969. *Татар теленең диалектологик сүзлеге*. Казан: Татар. кит. нәшр.

ТТЗДС = Баязитова, Ф. С., Д. Б. Рамазанова, З. Р. Садыйкова, Т. Х. Хәйретдинова (сост.). 2009. *Татар теленең зур диалектологик сүзлеге*. Казан: Татар. кит. нәшр.

Усачёва, М. Н., Т. А. Архангельский, О. Л. Бирюк, В. А. Иванов, Р. И. Идрисов. 2017. *Тезаурус бесермянского наречия: имена и служебные части речи (говор деревни Шамардан)*. Москва: Издательские решения.

Ӱпӹмарий (Васильев, В. М.). 1926. *Марий мутэр*. Моско: СССР калыӄ-влак рӱдӧ савыӄтыӹш.

Федюнева, Г. В. 2014. К вопросу о булгарско-пермских взаимоотношениях и их интерпретации по данным лексики. В: С. А. Мызников и др. (ред.). *Вопросы уралистики 2014. Научный Альманах*: 116–134. Санкт-Петербург: Нестор-История.

Хакимов, Р. С. и др. (ред.). 2017. *История и культура татар-кряшен (XVI–XX вв.)*. Казань: Институт истории им. Ш. Марджани АН РТ.

Хелимский, Е. А. 2000. *Компаративистика, уралистика. Лекции и статьи*. Москва: Языки русской культуры.

ЭСРЯ = Фасмер, Макс. 1986–1987. *Этимологический словарь русского языка. Т. I–IV*. Москва: Прогресс.

ЭСТЯ 1 = Севортян, Э. В. 1974. *Этимологический словарь тюркских языков: Общетюркские и межтюркские основы на гласные*. Москва: Наука.

ЭСТЯ 2 = Севортян, Э. В. 1978. *Этимологический словарь тюркских языков: Общетюркские и межтюркские основы на букву «Б»*. Москва: Наука.

ЭСЧЯ = Федотов, М. Р. 1996. *Этимологический словарь чувашского языка. Т. 1–2*. Чебоксары: Чувашский государственный институт гуманитарных наук.

References

Əkhmət'janov, R. G. 2015. *Tatar teleneŋ etimologik sүzlege. Ike tomda: I tom (A–L); II tom (M–Ja).* Kazan: Məgarif–Vakyt.

Belykh, Sergei. 2007. Words of the Middle Mongolian origin in the Permic languages. *Idnakar: Metody istoriko-kul'turnoj rekonstruktsii* 2(2): 33–39.

Belykh, S. K. 2009. *Problema raspada prapermskoj etnojazykovoj obshchnosti*. Izhevsk: Udmurtskij gosuniversitet.

Benzing, Johannes. 1959. *Das Tschuwaschische*. In: Jean Deny, Kaare Groenbech, Helmuth Scheel, Zeki Velidi Togan (eds.). *Philologiae Turcicae Fundamenta* 1: 695–751. Wiesbaden: Steiner.

Burganova, N. B. 1962. Govor karinskix i glazovskix tatar. In: *Materialy po tatarskoj dialektologii* 2: 19–56. Kazan': Tatpoligraf.

Csúcs, Sándor. 2007. Permi-magyar nyelvtörténeti párhuzamok. *Nyelvtudományi Közlemények* 104: 86–123.

Culver, Christopher, Niklas Metsäranta. 2025. Towards the Reconstruction of the Volga Bulgarian and Middle Chuvash Lexicon. *International Journal of Eurasian Linguistics* 7: 335–357. doi: 10.1163/25898833-20250078.

Diószegi, Vilmos. 1970. Libation Songs of the Altaic Turks. *Acta Ethnographica Academiae Scientiarum Hungaricae* 19: 95–106.

Dybo, A. V. 2007. *Lingvisticheskije kontakty rannix tjurkov. Leksicheskij fond. Prat'urkskij period*. Moskva: Vostochnaja literatura.

Dybo, A. V. 2010. Vokalizm rannet'urkskix zaimstvovanij v vengerskom. In: Valentin Ju. Gusev, Anna Widmer, Annkatrin Klein (eds.). *Gedenkschrift für Eugen A. Helimski* (= *Finnisch-Ugrische Mitteilungen, Bd. 32/33, Jahrgang 2008/2009*): 83–132.

Dybo, A. V. 2011. Otrazhenije bulgarskix palatalizatsij v bulgarizmax vengerskogo jazyka. In: A. V. Kuznetsov (ed.). *Chuvashskij jazyk i etnos v istorii jevrazijskoj tsivilizatsii. Materialy Mezhdunarodnoj t'urkologicheskoj konferentsii*: 10–21. Cheboksary: ChGIGN.

Dybo, A. V. 2013. *Etimologicheskij slovar' tjurkskix jazykov. Tom 9 (dopolnitel'nyj)*. Astana: TOO "Prosper Print".

Dybo, A. V. 2021. Bulgarskije zaimstvovanija v slav'anskix jazykax: osobennosti vokalizma. *Rusin* 66: 183–206.

EDAL = Starostin, Sergei A., Anna V. Dybo, Oleg A. Mudrak. 2003. *Etymological dictionary of Altaic languages*. Leiden: Brill.

EDT = Clauson, Gerard. 1972. *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford: Clarendon Press.

Erdal, Marcel. 1993. *Die Sprache der wolgabolgarischen Inschriften*. Wiesbaden: Harrassowitz.

ESChJa = Fedotov, M. R. 1996. *Etimologicheskij slovar' chuvashskogo jazyka. T. 1–2*. Cheboksary: Chuvashskij gosudarstvennyj institut gumanitarnyx nauk.

ESRJa = Vasmer, Max. 1986–1987. Etimologicheskij slovar' russkogo jazyka. T. I–IV. Moskva: "Progress".

ESTJa 1 = Sevortjan, E. V. 1974. *Etimologicheskij slovar' t'urkskix jazykov: Obshchet'urkskije i mezht'urkskije osnovy na glasnye*. Moskva: Nauka.

ESTJa 2 = Sevortjan, E. V. 1978. *Etimologicheskij slovar' t'urkskix jazykov: Obshchet'urkskije i mezht'urkskije osnovy na bukvu "B"*. Moskva: Nauka.

Fedjuneva, G. V. 2014. K voprosu o bulgarsko-permskix vzaimootnoshenijax i ix interpretatsii po dannym leksiki. In: S. A. Myznikov et al. (eds.). *Voprosy uralistiki 2014. Nauchnyj Al'manax*: 116–134. Sankt-Peterburg: Nestor-Istorija.

Geisler, Michael. 2002. Etymologien aus den permischen Sprachen. *Linguistica Uralica* 38(4): 241–244.

Grishkina, M. V. 1988. Sluzhiloje zemlevladenije arskix kn'azej v Udmurtii XVI — pervoj poloviny XVIII vekov. In: V. Ye. Vladykin et al. (eds.). *Problemy agrarnoj istorii Udmurtii*: 20–35. Izhevsk: Udmurtskij institut istorii, jazyka i literatury.

Grishkina, M. V. 1989. Novyje gramoty arskim kn'az'jam na V'atke. *Sovetskije arxivy* 4: 67–71.

Helimski, E. A. 2000. Komparativistika, uralistika. Lektsii i stat'i. Moskva: Jazyki russkoj kul'tury.

Holopainen, Sampsa. 2023. Notes on an old problem of Hungarian historical vocalism: the sporadic (?) change of Uralic **u* > Hungarian *a, á*. *Finnisch-Ugrische Forschungen* 68: 101–140.

Idrisov, R. I. 2013. *T'urkskije zaimstvovanija raznyx periodov v beserm'anskom dialekte udmurtskogo jazyka*. Moskva: Izdatel'stvo MGU.

Itkonen, Erkki. 1970. Bemerkungen über den Vokalismus der ersten Silbe von tschuwaschischen Lehnwörtern in einigen finnisch-ugrischen Sprachen. *Finnisch-Ugrische Forschungen* 38: 257–273.

KChRS = Zolotnitskij, N. I. 1875. *Kornevoj chuvashsko-russkij slovar'*. Kazan': Tipografija imperatorskogo universiteta.

Kel'makov, V. K. 1985. K voprosu o bulgarskom substrate v jazyke beserm'an. In: *Voprosy dialektologii t'urkskix jazykov*: 114–122. Ufa: BFAN SSSR.

Kel'makov, V. K. 1998. *Kratkij kurs udmurtskoj dialektologii: Vvedenije. Fonetika. Morfologija. Dialektnyje teksty. Bibliografija*. Izhevsk: Izd-vo Udm. un-ta.

KESK = Lytkin, V. I., Je. S. Guljaev. 1970. *Kratkij etimologicheskij slovar' komi jazyka*. Moskva: "Nauka".

Khakimov, R. S. et al. (eds.). 2017. *Istorija i kul'tura tatar-kr'ashen (XVI–XX vv.)*. Kazan': Institut istorii im. Sh. Mardzhani AN RT.

Kuokkala, Juha. 2018. Finnic-Saamic labial vowels of non-initial syllables: an etymological evaluation. In: Sampsa Holopainen, Janne Saarikivi (eds.). Περὶ ὀρθότητος ἐτύμων. *Uusiutuva uralilainen etymologia (Uralica Helsingiensia 11)*: 11–75. Helsinki: Suomalais-Ugrilainen Seura.

Kuznetsova, A. I., N. V. Serdobol'skaja, M. N. Usacheva, O. L. Birjuk, R. I. Idrisov. 2013. *Slovar' beserm'anskogo dialekta udmurtskogo jazyka*. Moskva: TEZAURUS.

Lakó, György. 1935. A permi nyelvek szóvégi magánhangzói. *Nyelvtudományi közlemények* 49: 141–171.

Luppov, P. N. 1958. *Udmurty v XV–XVII vekax. Dokumenty po istorii Udmurtii XV–XVII vekov*. Izhevsk: Udmurtskoje knizhnoje izdatel'stvo.

Lytkin, V. I. 1964. *Istoricheskij vokalizm permskix jazykov*. Moskva: Nauka.
Lytkin, V. I. 1967. O drevnet'urkskix elementax v leksike permskix jazykov. *Voprosy finno-ugorskogo jazykoznanija* IV: 131–142.
Lytkin, V. I., Je. S. Guljaev. 1975. Dopolnenija k kratkomu etimologicheskomu slovar'u komi jazyka. *Komi filologija. Trudy IJaLI KFAN SSSR* 18: 3–45.
Maksimov, S. A. 2008. Samonazvanija udmurtov i beserm'an, ix varianty, nazvanija sosednix narodov v jazyke udmurtov i beserm'an. *Linguistica Uralica* 44(2): 108–118.
Maksimov, S. A. 2014. Vozmozhnosti i puti obogashchenija literaturnogo jazyka leksikoj marginal'nyx dialektov (na primere udmurtskogo jazyka). *Jezhegodnik finno-ugorskix issledovanij* 2014(5): 92–101.
Maksimov, S. A. 2016. O komi zaimstvovanijax v udmurtskom jazyke. *Linguistica Uralica* 52(2): 111–121.
Maksimov, S. A. 2018a. *Severno-udmurtsko-komi areal'nyje jazykovyje paralleli: leksika, fonetika, morfologija*. Izhevsk: UIIJaL UdmFITs UrO RAN.
Maksimov, S. A. 2018b. T'urksko-udmurtskije leksicheskije metamorfozy, ili o nekotoryx slovax so "strannoj" vnutrennej strukturoj v udmurtskom jazyke. *Jezhegodnik finno-ugorskix issledovanij* 12(3): 30–46.
Maksimov, S. A. 2019. O terminax rodstva po materinskoj linii v udmurtskom jazyke: istochnik tipologicheskoj modeli. *Vestnik ChGPU im. I. Ja. Jakovleva* 3(103): 126–134.
Maksimov, S. A. 2025. Beserm'ansko-komi izoglossy kak svidetel'stvo bylyx kontaktov mezhdu narodami. *Linguistica Uralica* 61(2): 89–105.
Mudrak, O. A. 1993. *Istoricheskije sootvetstvija chuvashskix i t'urkskix glasnyx: opyt rekonstruktsii i interpretatsii*. Moskva: Institut vostokovedenija RAN.
Mudrak, O. A., F. G. Khisamitdinova. 2018. *Kypchakskije jazyki Uralo-Povolzhja*. Astana: Gylym.
Munkácsi, Bernát. 1883. Idegen elemek a votják nyelvben. *Nyelvtudományi Közlemények* 18: 55–155.
Munkácsi, Bernát. 1896. *A votják nyelv szótára*. Budapest: Magyar Tudományos Akadémia.
Musikhin, A. L. 2013. Gramoty karinskim arskim kn'az'jam XVI v.: istorija publikatsii i izuchenija. *Jezhegodnik finno-ugorskix issledovanij* 2013(4): 82–90.
Napol'skix, V. V. 2018. *Ocherki po etnicheskoj istorii*. Kazan': Institut arxeologii im. A. Kh. Khalikova Akademii nauk Respubliki Tatarstan.
Napol'skix, V. V. 2021. Udmurtskaja *akashka* ~ beserm'anskaja *akajashka* i pozdnesrednevekovaja etnicheskaja istorija Nizhnego Prikam'ja. *Etnografija* 4(14): 37–54.
Napol'skix, V. V. 2025. K etimologii russk. *bezmén*. *Voprosy jazykoznanija* 1: 136–147.
Napol'skix, V. V. 2026. O vremeni i istoricheskix obstojatel'stvax permskoj denazalizatsii. *Tomskij zhurnal lingvisticheskix i antropologicheskix issledovanij* 51(1): 29–39.
Nasipov, I. S. 2010. *Finno-ugorskije zaimstvovanija v tatarskom jazyke: sinopsis i taksonomija. Doctoral thesis*. Kazan'.
Nugteren, Hans. 2011. *Mongolic phonology and the Qinghai-Gansu languages*. Utrecht: LOT.
Ostanina, T. I. 2021. Gordinskij kamen' 1323 g.: novyje svedenija. *Povolzhskaja arxeologija* 38(4): 179–190.
Ponarjadov, V. V. 2014. *Razvitije glasnyx v permskix jazykax*. Syktyvkar: Komi nauchnyj tsentr UrO RAN.
Poppe, Nicholas. 1977. On Chuvash-Mongolian Linguistic Contacts. *Journal of the American Oriental Society* 97(2): 111–114.
Räsänen, Martti. 1920. *Die tschuwassischen Lehnwörter im Tscheremissischen*. Helsinki: Société finno-ougrienne.
Rédei, Károly, András Róna-Tas. 1972. A permi nyelvek őspermi kori bolgár-török jövevényszavai. *Nyelvtudományi közlemények* 74: 281–298.
Rédei, Károly, András Róna-Tas. 1975. A bolgár-török–permi érintkezések néhány kérdése. *Nyelvtudományi közlemények* 77: 31–44.
Rédei, Károly, András Róna-Tas. 1983. Early Bulgarian loanwords in the Permian languages. *Acta Orientalia Academiae Scientiarum Hungaricae* 37: 3–41.
RES 5 = Anikin, A. E. 2017. *Russkij etimologicheskij slovar'. Vypusk 5 (buba I — vakshtaf)*. Moskva: Jazyki slav'anskix kul'tur.
Róna-Tas, András. 1975. The Altaic theory and the history of a Middle Mongolian loan word in Chuvash. In: L. Ligeti (ed.). *Researches in Altaic languages*: 201–211. Budapest: Akadémiai Kiadó.
Róna-Tas, András. 1982. Loan-words of ultimate Middle Mongolian origin in Chuvash. In: András Róna-Tas (ed.). *Studies in Chuvash Etymology 1 (Studia Uralo-Altaica 17)*: 66–134. University of Szeged.
Róna-Tas, András. 1988. Turkic influence on the Uralic languages. In: Denis Sinor (ed.). *The Uralic languages: Description, History and Foreign Influences*: 742–780. Leiden: Brill.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages. In: Denis Sinor (ed.). *The Uralic languages*: *Description, History and Foreign Influences*: 478–554. Leiden: Brill.

Savelyev, A. V. 2023. Kommentarii k chuvashskim materialam G. F. Millera. *Rossijskaja t'urkologija* 40–41: 24–52.

Savelyev, A. V. 2025. *O kriterijax vyjavlenija i stratifikatsii bulgaro-chuvashskix zaimstvovanij v permskix jazykax I*. Voprosy jazykovogo rodstva 23(1–2): 137–167.

Savelyev, A. V., N. S. M'asnikov. 2026. *K proisxozhdeniju etnonima chuvash: osnovnyje versii i ix istoricheskaja interpretatsija. (Nauchnyje doklady ChGIGN 42.)* Cheboksary: Chuvashskij gosudarstvennyj institut gumanitarnyx nauk.

SChJa = Ashmarin, N. I. 1928–1950. *Slovar' chuvashskogo jazyka. T. I–XVII*. Kazan' / Cheboksary.

Serebrennikov, B. A. 1963. *Istoricheskaja morfologija permskix jazykov*. Moskva: Izdatel'stvo AN SSSR.

SIGTJa 2006 = Tenishev, E. R., A. V. Dybo (eds.). 2006. *Sravnitel'no-istoricheskaja grammatika t'urkskix jazykov: Prat'urkskij jazyk-osnova. Kartina mira prat'urkskogo etnosa po dannym jazyka*. Moskva: Nauka.

SRNG 3 = Filin, F. P. (ed.). 1968. *Slovar' russkix narodnyx govorov. Vyp. 3: Blaznishka — Bjashutka*. Moskva / Leningrad: Nauka.

Tepl'ashina, T. I. 1968. O smeshenii terminov "chuvashi" i "beserm'ane" v pis'mennyx istochnikax. In: *Uchenyje zapiski ChNIIJaLiE 40: Istorija, etnografija, sotsiologija*: 177–187. Cheboksary: Chuvashknigoizdat.

Tepl'ashina, T. I. 1970. *Jazyk beserm'an*. Moskva: Nauka.

TTDS = Makhmutova, L. T. (ed.). 1969. *Tatar teleneŋ dialektologik sүzlege*. Kazan: Tatar. kit. nəshr.

TTZDS = Bajazitova, F. S., D. B. Ramazanova, Z. R. Sadyjkova, T. Kh. Khəjretdinova (eds.). 2009. *Tatar teleneŋ zur dialektologik sүzlege*. Kazan: Tatar kit. nəshr.

UEW = Rédei, Károly. 1988–1991. *Uralisches Etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Usacheva, M. N., T. A. Arkhangel'skij, O. L. Birjuk, V. A. Ivanov, R. I. Idrisov. 2017. *Tezaurus beserm'anskogo narechija: imena i sluzhebnyje chasti rechi (govor derevni Shamardan)*. Moskva: Izdatel'skije reshenija.

Üpymarij (Vasil'ev, V. M.). 1926. *Marij muter*. Mosko: SSSR kalyk-vlak rüdö savyktysh.

VEWT = Räsänen, Martti. 1969. *Versuch eines etymologischen Wörterbuchs der Türksprachen*. Helsinki: Suomalais-Ugrilainen Seura.

Viitso, Tiit-Rein. 1993. Livonian kõps 'hare' and Finnic *korva 'ear'. *Linguistica Uralica* 29(2): 89–94.

Wichmann, Yrjö. 1903. *Die tschuwassischen Lehnwörter in den permischen Sprachen. (Mémoires de la Société Finno-ougrienne XXI.)* Helsinki: Druckerei der finnischen Litteraturgesellschaft.

Wichmann, Yrjö. 1915. *Zur Geschichte des Vokalismus der ersten Silbe im Wotjakischen mit Rücksicht auf das Syrjänische. (Mémoires de la Société Finno-ougrienne XXXVI.)* Helsinki: Druckerei der finnischen Litteraturgesellschaft.

Wichmann, Yrjö. 1987. *Wotjakischer Wortschatz. (Lexica Societatis Fenno-Ugricae 21.)* Helsinki: Suomalais-Ugrilainen Seura.

WOT = Róna-Tas, András, Árpád Berta. 2011. *West Old Turkic: Turkic Loanwords in Hungarian. Parts I–II*. Wiesbaden: Harrassowitz.

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In: Daniel Abondolo et al. (eds.). *The Uralic Languages (2nd edition)*: 117–175. NY, Abingdon: Routledge.

Alexander Savelyev. On the criteria for identification and stratification of Bulghar-Chuvash loanwords in Permic II

This article is the second and final part of a study of Bulghar-Chuvash loanwords in the Permic languages. It proposes additional phonetic criteria for identifying late (14th–15th-century) Bulgharic loanwords, primarily in Udmurt, whose source language is identified as Middle Chuvash. The first part of the study identified a layer of early (8th–9th-century) borrowings from Old Chuvash into Proto-Permic. Against this background, the article discusses the mechanism of loanword renewal in the Bulgharic-Permic contact lexicon. In particular, it examines cases in which an older borrowing was retained in Komi but replaced by a more recent Chuvash loanword in Udmurt, as well as cases in which an older borrowing was retained in Udmurt but replaced by a Russian loanword in Komi.

A separate section addresses areas of late, highly localized contact between Middle Chuvash dialects and the already differentiated dialects of the Permic languages, with particular attention to the possible presence of a Chuvash-speaking population in the Cheptsa River ba-

sin as late as the 17th century. In this context, the article discusses possible traces of late Chuvash influence specific to Beserman, Northern Udmurt, Southern Komi-Zyrian, and Cheptsa Tatar varieties.

Following the discussion of the other identifiable strata of Bulgharic loanwords in Permic, the final section examines a distinct group of borrowings widely distributed across both Udmurt and Komi dialects (including areas beyond the Southern Komi zone) that exhibit non-trivial Komi-Udmurt correspondences. These borrowings are tentatively dated to the 10th–13th centuries; the proposed source language is Late Old Chuvash, and the likely recipient is the Common Permic continuum undergoing disintegration. Overall, the proposed reconstruction points to virtually continuous Bulgharic-Permic linguistic contacts from the beginning of the Bulghar migration into the Volga-Kama region at the very end of the 7th century until the 16th–17th centuries, when the Chuvash population of *Arskaya Storona* was gradually assimilated into the Tatar population, while the presumed Chuvash ethnolinguistic component in the Cheptsa region was likewise absorbed into the surrounding populations.

Keywords: Permic languages; Udmurt language; Komi language; Bulghar-Chuvash loanwords; Chuvash language; Tatar language; language contact; stratification of loanwords; historical phonology; etymology.

Reports / Хроника

Anastasia Garnet Romantsevski
Institute of Linguistics of the Russian Academy of Sciences; Institute of Oriental and Classical Studies, RSUH (Moscow);
sicfaturlacrimanss@gmail.com

The 21st Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics Moscow, HSE University, March 23–24, 2026

The 21st Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics was held in Moscow at HSE University on March 23–24, 2026. Organized by the Centre for Comparative Studies and Phylogenetics of the Institute for Oriental and Classical Studies, HSE University, the conference typically brings together researchers working on languages from a wide range of families and geographical areas, and this time was no exception as the schedule covered traditional problems of comparative-historical linguistics — phonological and morphological reconstruction, language classification and etymology — alongside lexicostatistics, computational phylogenetics, semantic change, and language contact.

As usual, *Indo-European studies* occupy a prominent spot in the program. This time, Artyom Badeev (Institute of Linguistics of the Russian Academy of Sciences) opened the talks with **"On the origin of indefinite markers in Mitagi-Dzhalgan /Tat/ language."** Mitagi-Dzhalgan (Tat) developed for a long period in relative isolation from the Southwestern Iranian language group, resulting in a system of indefinite pronouns and determinatives distinct from that of Middle and New Persian. The absence of a postpositive indefinite article is compensated by the grammaticalization of the numeral *ye* 'one' as a prepositional indefinite marker: *ye ki* 'someone', *ye šæy* 'something', *ye ædæmi* 'some person'. A range of postpositive markers, some derived from the verbs 'to be' and 'to want; to love', points to an incomplete process of grammaticalization. At the same time, the Mitagi-Dzhalgan system shows parallels with the Northwestern Iranian Caspian languages and with Azerbaijani.

Sergei Bolotov (Institute of World Culture, Moscow State University) addressed the long-standing problem of Szemerényi's law in his paper **"Queen against fox and the Latin prosody in Sanskrit: 'anti-Szemerényi laws' with their sequence."** Rather than treating the forms traditionally associated with the law as the outcome of a single process, the paper considers a set of elementary rules operating in particular languages and morphological environments. The resulting system accounts both for forms displaying the apparent "Szemerényi effect" and for those in which it is absent. Greek γυνή 'woman', for instance, behaves differently from ἀλώπηξ 'fox', while different types of -ντ-participles likewise show distinct but regular developments. Sanskrit and Greek participial formations provide further examples of outcomes that are regular precisely because they are not reducible to a single general rule.

Alexandra Evdokimova (Institute of Linguistics of the Russian Academy of Sciences) turned to **accent curves of the second declension in Byzantine Greek**, drawing on the BGAT corpus. Byzantine grammatical works, including those associated with Theodore Gaza and Manuel Moschopoulos, transmit numerous examples whose accentuation varies from manuscript to manuscript. Comparison of the same examples across different grammatical texts shows that some scribes consistently preferred particular accentual patterns, while others reproduced the forms found in their sources. Such variation provides evidence for the emergence and transmission of accentuation patterns in Byzantine grammatical tradition.

Elena Efimova (Tyumen State University) explored **broad-meaning verbal vocabulary in Germanic languages and its trajectories of semantic development**. The analysis focuses on the verb 'to do' in Gothic, Old High German, Old Norse, and Old English, combining comparative-historical analysis with distributional semantics. Thus, Germanic **dōn-* continues Proto-Indo-European **d^heh_1-* 'to put, place': from meanings associated with placing and producing objects, as well as creating or establishing conditions and rules, the verb has developed spatial-temporal, institutional, and abstract-causative uses. The individual Germanic languages followed these trajectories in different ways.

Anastasia Garnet Romantsevski (Institute of Linguistics of the Russian Academy of Sciences; Institute of Oriental and Classical Studies, RSUH) discussed **"The Epitaph of Ursula: A Linguistic Commentary**

on Inscription CIL XIII 3909." The Latin-Gallic funerary inscription, found near Augusta Treverorum (modern Trier) and tentatively dated to the fourth or fifth century AD, is inscribed on a clay tile and provides valuable evidence for Gallo-Latin bilingualism and language contact in Late Roman Gaul. Particular attention was paid to the relationship between the daughter's Latin name *Ursula* ("Little She-Bear") and her mother's Gallic name *Artula*, interpreted as its semantic and phonetic correlate and derived from Gallic *artos* 'bear'. The paper also addressed alternative readings of the presumably Gallic final part of the inscription, *Kara matir titiulum po(suit)*. The linguistic commentary bears on both the reconstruction of the inscription and the linguistic competence of the local Treverian population, as well as on the representation of *Latinitas vulgata* in provincial epigraphy.

Artem Trofimov (School for Advanced Studies in the Humanities, RANEPA; Institute of Linguistics of the Russian Academy of Sciences) considered **"Anatolian and Tokharian verbs with secondary semantic developments compared to Proto-Indo-European."** Anatolian and Tocharian are generally regarded as the first and second branches to have split from the Proto-Indo-European language community. Their lexical evidence sometimes preserves earlier meanings, as in Hittite *mer-*/*mar-* 'to disappear' versus Narrow Indo-European **mer-* 'to die', and Hittite *pāš-* 'to swallow' versus Narrow Indo-European peh_3- 'to drink'. At the same time, both branches contain verbal stems with secondary semantic developments, often associated with morphological derivation. The paper examined such cases on the basis of etymological dictionaries and databases.

On the source data of Indo-European languages (mainly Celtic and Slavic), Tatyana Mikhailova (Institute of Linguistics of the Russian Academy of Sciences; Institute of Oriental and Classical Studies, RSUH) examined **"two-step semantic shifts"**, drawing on data from DatSemShift and related databases. The discussion centered on semantic developments associated with the concepts 'life' and 'food', including the shift *to live, to be alive → food* and subsequent developments toward 'meat', 'rice', and 'cereals'. Particular attention was given to the interpretation of Welsh *bwyd* 'food', the history of Slavonic *žirъ*, and the possibility of reconstructing an unrecorded intermediate meaning in the development *to live → food → cereal*. Insular Celtic evidence for a possible *to eat → cereal*shift was also considered. The broader issue is whether two- or more-step semantic shifts can be reconstructed on the basis of motivated intermediate stages, and to what extent such stages may help recover meanings not directly attested in the historical record.

Linguistic contacts between Indo-European and other language families (in this case, Semitic) were also touched upon by Alexey Lyavdansky (Institute for Classical and Oriental Studies, HSE University), who examined **"Iranian Lexical Borrowings in Eastern Middle Aramaic: A Dialectological Perspective."** Iranian–Aramaic contact represents a long and complex history of interaction, but Iranian borrowings in Jewish Babylonian Aramaic and Mandaic remain less thoroughly studied than those in Classical Syriac. Several Iranian loans shared by Jewish Babylonian Aramaic and Mandaic but absent from Syriac were considered, alongside loans occurring in all three dialects whose phonetic development displays features common to Jewish Babylonian Aramaic and Mandaic. Examples include MP *zyān* 'loss, damage' and MIr *frasaxa* 'parasang'. Such Mandaic–Jewish Babylonian Aramaic lexical and morpho-lexical isoglosses are relevant to the subgrouping of Eastern Aramaic and add to the evidence for a closer relationship between the two dialects.

Turning to other language families of Eurasia, comparative *Uralic studies* were represented by two presentations:

Valentin Gusev (Institute of Linguistics of the Russian Academy of Sciences) discussed **vowel harmony in Nganasan**, focusing on its history and present-day structure. Although the modern system resembles front-back harmony, the phonetic motivation for this opposition has been weakened by subsequent changes in the vowel system. Historical correspondences nevertheless point to two harmonic classes and make it possible to reconstruct phonetically natural front–back vowel pairs. The evidence suggests that Proto-Samoyedic non-first-syllable vocalism was not as impoverished as has sometimes been assumed. Low vowels and schwa appear particularly frequently, while mid vowels are unexpectedly more frequent than high vowels.

Leonid Zaitsev (Lomonosov Moscow State University) presented **"Stepwise reconstruction for morphophonology and Uralic consonant gradation."** The starting point is E. A. Helimski's reconstruction of a complex Proto-Uralic system of consonant alternations involving rhythmic and syllabic gradation. One difficulty of this reconstruction lies in comparing relatively recent material from a language long separated from its closest relatives with the proto-languages of other branches. Forest Nenets, which preserves a subsystem of cluster gradation, provides an additional intermediate stage. In comparison with Tundra Nenets

and Nganasan, these data allow consonant gradation to be reconstructed at the Proto-Nenets and Proto-(Northern) Samoyed levels. Stepwise reconstruction thus makes the intermediate systems comparable with other Uralic proto-languages and strengthens the case for Proto-Uralic consonant gradation.

The field of *Turkic studies* was represented by Alexander Savelyev (Institute of Linguistics of the Russian Academy of Sciences; Institute for Oriental and Classical Studies, HSE University; Institute for History and Archaeology, Ural Branch of the Russian Academy of Sciences), who gave a talk on **"Chuvash wC-type clusters in the light of Proto-Turkic reconstruction and Altaic comparison."** The paper systematically examines cases in which Chuvash *wC-type clusters correspond to a plain *C in Narrow Turkic. Given the binary structure of the Turkic family, possible Proto-Turkic reconstructions must also take Altaic parallels into account. The evidence argues against treating Chuvash **w* as an "intrusive" element: it primarily continues Proto-Turkic **b* and **g*, with a few other sources occurring more rarely. The analysis also points to the need for a more systematic incorporation of Bulgharic evidence into Proto-Turkic reconstruction.

Regarding the areal relationships of specific Turkic languages, Timofey Kuzmenko (Institute of Asian and African Studies, Lomonosov Moscow State University) discussed **"Possible Middle Chinese borrowings in the Salar language."** Phonetic patterns in loanwords from Northern Chinese dialects suggest that some items may have been borrowed from Middle Chinese. This would point to earlier and closer contacts between Salar and Middle Chinese than those implied by Ming historical chronicles and Salar folklore, which give 1370 as the date of the Salar "migration" to the territory of present-day China. The evidence may also be relevant to glottochronological estimates placing the separation of Salar and Turkmen in 1220. The paper seeks to identify a possible group of direct Middle Chinese loans and distinguish them from later borrowings from Northern Chinese dialects and Southern Mongolian languages.

Turning to the Caucasus and, specifically, the *Kartvelian family*, Roman Pavlov presented **"Principles for building an etymological database of the Kartvelian languages."** The discussion starts from the problems posed by the existing Kartvelian etymological dictionaries, particularly the quality and reproducibility of the input data used for external comparison. Some phonological correspondences do not have sufficiently well-established distributions; dialectal Svan, Megrelian, and Laz material is often presented without precise dialectal or source information; and the selection of verbal forms does not always adequately represent the distribution of ablaut grades. The proposed database structure is intended to address these issues and draws on the experience of such projects as *Lexikon der indogermanischen Verben*, *Nomina im Indogermanischen Lexikon*, and *A North Caucasian Etymological Dictionary*.

The adjacent (but not genetically related) *Nakh-Daghestanian family* was covered by two presentations. In one of them, Timur Maisak and Steven Kaye (Institute of Linguistics of the Russian Academy of Sciences; HSE University; University of Surrey) discussed **"Irregular presents of motion verbs in Andic languages: comparative data and a hypothesis on their origin."** Most Andic languages, with the exception of Akhvakh and Tindi, possess irregular present-tense forms of the motion verbs 'go (away)' and/or 'come'. The present forms, or derived imperfective stems, contain a special marker with a lateral *ƛ* or *ɬ* in place of the regular present-tense or imperfective marker. This element resembles a reduced form of the general motion verb 'go, walk', which is not always preserved as an independent lexeme. The authors propose that the irregular forms originated in combinations of the aorist (perfective stem) of a directed motion verb with the present tense of the general verb 'go, walk', the latter subsequently becoming grammaticalized as a tense suffix or stem marker.

Another talk, by Mikhail Zhivlov (Institute of Oriental and Classical Studies, RSUH; Institute for Oriental and Classical Studies, HSE University), addressed **"Laryngeal theory or consonant stems? A problem in Proto-Andic reconstruction."** The discussion centered on nouns with unusual correspondences involving resonants, vowel length, and nasalization. The word 'water', represented by forms such as Andi *ɬen* and Chamalal *ɬīⁿ*, provides a particularly clear example. Whereas the *North Caucasian Etymological Dictionary* reconstructs resonant–laryngeal clusters, an alternative analysis treats these forms as consonant stems ending in resonants. The two approaches lead to different interpretations of the Proto-Andic consonant system.

Several talks addressed contemporary issues in the classification and internal history of *Sino-Tibetan languages*. Anton Kogan (Institute of Oriental Studies of the Russian Academy of Sciences) proposed a classification of **Northwestern Tibetan dialects** based on glottochronological counts. The resulting dated genealogical tree provides a basis for hypotheses concerning the chronology of the spread of particular Tibetan varieties in the Upper Indus region and their contacts with the languages of the local pre-Tibetan population. The proposed chronology was compared with genetic data.

Sophia Khafizova (School of Linguistics, HSE University) gave a talk on the **"Classification of West-Gyalrongic languages based on the analysis of basic lexicon."** Her study focuses on two branches of West Gyalrongic: Horpa, including Eastern Geshiza and *rTa'u*, and Khroskyabs, including Lavrung (Yelong and Guanyinqiao) and Khroskyabs Wobzi. The classification is based on basic-lexicon wordlists, and the results are intended to provide a basis for further etymological analysis.

Additionally, Li Andong (Peoples' Friendship University of Russia) addressed some issues of Chinese (Sinitic) dialectology in **"On Some Phonetic Aspects of the Influence of Chinese Idioms on Mandarin."** The phonetic norm of Mandarin in the People's Republic of China is based on Beijing pronunciation, but a number of direct and indirect exceptions reflect other traditions. Orthoepic standardization affected both native "colloquial" and borrowed "literary" readings of Chinese characters, sometimes retaining a single reading and sometimes assigning different readings to specific domains of use. Southern Chinese varieties have also left their mark on phonetic loanwords, including proper names. Such forms may differ considerably from their original foreign pronunciation in Beijing Mandarin, while Southern traditions continue to survive within Putonghua. New borrowings may likewise show dialectal influence.

Comparative *Yeniseian studies* formed the principal topic of George Starostin's (Institute for Oriental and Classical Studies, HSE University; Santa Fe Institute, New Mexico) presentation **"From 'hyper-reconstruction' to 'optimisation' (on the example of comparative Yeniseian material)."** The paper considered the tendency for established proto-language reconstructions to become increasingly elaborate through the addition of controversial phonetic correspondences, poorly substantiated material, internal reconstruction, "word-family" arguments, or hypothetical contacts with neighboring language families. The result may be a "hyper-reconstruction" in which a systematically established core is combined with debatable or highly improbable elements. The current state of Proto-Yeniseian reconstruction was reviewed, beginning with S. A. Starostin's classic 1982 study and including subsequent alternative proposals. Rather than further complicating the reconstruction, the paper advocates "optimisation": typological oddities and suspicious redundancies can be removed by identifying additional positional distributions and intra-family borrowings. As one example, the velar and uvular fricatives **x* and **χ* in the 1982 reconstruction can be eliminated without seriously affecting the system as a whole.

Turning further East — to the field of comparative *Austronesian studies* — we find Albert Davletshin's (Institute of Oriental and Classical Studies, RSUH) talk on **"Laryngeal cooccurrence constraints in Proto-Polynesian."** Polynesian roots generally exclude adjacent homorganic consonants, while codas are absent from most Polynesian languages. Within the Proto-Polynesian system, the laryngeals **ʔ* and **h* do not normally cooccur within a root, except that **h* may occur next to the rounded vowels **o* and **u*; **ʔ* is likewise incompatible with **k*. These restrictions are considered both from an articulatory and from a historical perspective. They also bear on the reconstruction of the Proto-Polynesian lexicon and may help account for the double reflexes **s* and **h* of Proto-Central Pacific ***c*/***z*.

Evgeniya Korovina (Institute of Linguistics of the Russian Academy of Sciences) also used the Austronesian family as her source material while examining the role of the analytical method in **"Method as a Variable: Classifying the Languages of East Polynesia."** Different clustering procedures may produce substantially different results from the same dataset, making it important to distinguish genuinely stable nodes from those introduced by a particular model. East Polynesian provides a useful testing ground: although the subgroup itself is clearly defined, its internal classification remains disputed, especially with regard to Hawaiian and Rapa Nui. Using basic vocabulary lists of different lengths, the paper compares classical distance-based procedures with phylogenetic methods, including algorithms capable of accounting for homoplasy. The broader question is what evidence is sufficient to establish a genealogically meaningful subgroup.

The African and American continents were represented by one talk each this time. Anna Smirnitskaya (Institute of Oriental Studies of the Russian Academy of Sciences) explored **the correlation between the cognitive mechanism of semantic shift and its morphological manifestation in the EvoLex etymological dictionary of Khoe-Kwadi**. The various types of semantic-shift realization described in DatSemShift 3.0 link lexical meanings at different semantic distances. Such relationships may arise through borrowing, metaphor, external similarity, or metonymy; in Khoekhoe, for example, *mû* means both 'to see' and 'eye'. The greatest semantic distance is found among cognates displaying radial semantic development, as in the reflexes of PKK *g!uri:* 'fat' in Khoe 'fat' and Kora 'height', or PKK *ǂae:* 'page' in Naro and 'arrow tip' in Khoe. The analysis is based on the EvoLex comparative dictionary of Khoe-Kwadi languages.

Javier Carol (CONICET–Universidad Autónoma de Entre Ríos, Argentina) presented **"Person indices in Proto-Mataguayan."** The Mataguayan family comprises Wichí, Chorote, Nivaclé, and Maká. Four principal personal paradigms can be reconstructed for the proto-language: active indicative, inactive indicative, non-indicative, and possessive. Considering these paradigms together is essential, since reflexes of the same index frequently occur in different paradigms across the daughter languages. The comparison was extended to Proto-Guaycuruan, a closely related family: most, if not all, personal indices in the two families have cognates. This evidence modifies conclusions based on Mataguayan alone and also sheds new light on the internal diversification of Proto-Mataguayan, which appears somewhat different when viewed through personal morphology rather than lexical data.

Finally, the Conference would be incomplete without at least one significant talk on the general methodology of historical linguistics, and on this occasion Mikhail Vasilyev, Arseniy Manusov, and Mikhail Saenko (Institute of Oriental Studies of the Russian Academy of Sciences; School of Linguistics, HSE University; Institute of Slavic Studies of the Russian Academy of Sciences) presented **"New linguistic software for constructing, dating and analyzing lexicostatistic trees."** The prototype, developed since 2025, is designed for constructing and dating phylogenetic classifications on the basis of lexical data. It accepts both standard lexicostatistical databases, including those created in Starling or Excel, and precalculated cognacy percentages. Among its features are an enhanced neighbor-joining procedure capable of producing non-binary nodes, the option of manually defining tree structures, and statistical tools for assessing the reliability of the source data and the resulting classification. Confidence intervals for individual nodes and graphical visualization of the results are also provided; further functions are planned for subsequent versions.

As usual, the 21st Sergei Starostin Memorial Conference brought together a particularly broad range of approaches to the reconstruction and classification of languages. Traditional comparative-historical analysis was represented alongside lexicostatistical and phylogenetic methods, while questions of semantic change, grammaticalization, language contact, and the organization of etymological data received considerable attention. The diversity of the material — from Iranian, Uralic, Nakh-Daghestanian, and Indo-European languages to Polynesian, Kartvelian, Turkic, Tibetan, Aramaic, Mataguayan, and Gyalrongic — once again reflected the wide comparative perspective that has long characterized the Starostin Memorial Conference.

The complete program of the Conference, along with Russian- and English-language abstracts and presentations from most of its participants, can be found online at the «Tower of Babel» web site (URL: starlingdb.org/meetings.php?lan=en).