

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

№ 4 (16)



Moscow 2018

Российский государственный гуманитарный университет  
Институт языкознания Российской Академии наук

# Вопросы языкового родства

Международный научный журнал

№ 4 (16)



Москва 2018

*Advisory Board:*

H. EICHNER (Vienna) / Chairman  
W. BAXTER (Ann Arbor, Michigan)  
V. BLAŽEK (Brno)  
M. GELL-MANN (Santa Fe, New Mexico)  
L. HYMAN (Berkeley)  
F. KORTLANDT (Leiden)  
A. LUBOTSKY (Leiden)  
J. P. MALLORY (Belfast)  
A. YU. MILITAREV (Moscow)  
V. F. VYDRIN (Paris)

*Editorial Staff:*

V. A. DYBO (Editor-in-Chief)  
G. S. STAROSTIN (Managing Editor)  
T. A. MIKHAILOVA (Editorial Secretary)  
A. V. DYBO  
S. V. KULLANDA  
M. A. MOLINA  
M. N. SAENKO  
I. S. YAKUBOVICH

Founded by Kirill BABAEV

*Редакционный совет:*

Х. АЙХНЕР (Вена) / председатель

В. БЛАЖЕК (Брно)

У. БЭКСТЕР (Анн Арбор)

В. Ф. ВЫДРИН (Париж)

М. ГЕЛЛ-МАНН (Санта-Фе)

Ф. КОРТЛАНДТ (Лейден)

А. ЛУБОЦКИЙ (Лейден)

Дж. МЭЛЛОРИ (Белфаст)

А. Ю. МИЛИГАРЕВ (Москва)

Л. ХАЙМАН (Беркли)

*Редакционная коллегия:*

В. А. ДЫБО (главный редактор)

Г. С. СТАРОСТИН (заместитель главного редактора)

Т. А. МИХАЙЛОВА (ответственный секретарь)

А. В. ДЫБО

С. В. КУЛАНДА

М. А. МОЛИНА

М. Н. САЕНКО

И. С. ЯКУБОВИЧ

Журнал основан К. В. БАБАЕВЫМ

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2018. — № 4 (16). — x + 75 с.

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by V. A. Dybo. — Moscow, 2018. — No. 4 (16). — x + 75 p.

ISSN 2219-3820

<http://www.jolr.ru/>  
[gstarst@rinet.ru](mailto:gstarst@rinet.ru)

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

Подписано в печать 10.12.2018. Формат 60×90/8.  
Бум. офсетная.  
Тираж 1050 экз. Заказ №374

Издательский центр  
Российского государственного гуманитарного университета  
125993, Москва, Миусская пл., 6  
[www.rggu.ru](http://www.rggu.ru)  
[www.knigirggu.ru](http://www.knigirggu.ru)

# Table of Contents / Содержание

Table of Contents / Содержание . . . . .	vii
Contributors / Сведения об авторах . . . . .	viii
Note for Contributors / Будущим авторам . . . . .	ix

## Articles / Статьи

<i>John Bengtson</i> . Some notes on Euskaro-Caucasian phonology . . . . .	247
<i>Джон Бенгтсон</i> . Заметки об эускара-кавказской фонологии	
<i>Anna Kuritsyna</i> . Suspended affixation with Tocharian adjectival suffix A - <i>ṣi</i> / B - <i>ṣṣe</i> and its possible parallel in Old Uighur. . . . .	265
[ <i>А. В. Курицына</i> . К вопросу о групповой аффиксации в тохарском (посредством адъективного суффикса А - <i>ṣi</i> / В - <i>ṣṣe</i> ) и ее возможных параллелях в древнеуйгурском]	
<i>Artem Trofimov</i> . On the place of Parachi and Ormuri among the Iranian languages according to the data of annotated Swadesh lists . . . . .	277
[ <i>А. А. Трофимов</i> . О классификации языков парачи и ормури внутри иранской группы по данным аннотированных сводешевских списков]	
<i>А. И. Коган</i> . О языке поэмы <i>Bāṇāsarakathā</i> (XV в.) и его отношении к языку кашмири . . . . .	293
[ <i>Anton Kogan</i> . On the language of the <i>Bāṇāsarakathā</i> poem and its relation to Kashmiri]	
<i>М. А. Молина</i> . Методы определения фокуса в мертвом языке: на пути к реконструкции праиндоевропейского синтаксиса . . . . .	304
[ <i>Maria Molina</i> . On the methodology of defining focus in an extinct language: towards a reconstruction of Proto-Indo-European syntax]	

## Contributors

*John Bengtson* — independent researcher, Association for the Study of Language in Prehistory, palaeojdb@hotmail.com

*Anton Kogan* — candidate of sciences (Philology), researcher, Department of Asian and African languages, Institute of Oriental Studies of the Russian Academy of Sciences (Moscow), kogan\_anton@yahoo.com

*Anna Kuritsyna* — junior researcher, Institute for Oriental and Classical Studies, Higher School of Economics (Moscow), linnansaari@gmail.com

*Maria Molina* — junior researcher, Institute of Linguistics, Russian Academy of Sciences (Moscow), maria.molina@me.com

*Artem Trofimov* — candidate of sciences (Philology), researcher, School of Advanced Studies for the Humanities, Russian Presidential Academy (Moscow), artemii.trofimov@gmail.com

## Сведения об авторах

*Бенгтсон, Джон* — независимый исследователь, Ассоциация изучения языка в доисторический период, jdbengt@softhome.net

*Коган, Антон Ильич* — канд. филол. наук, науч. сотрудник Отдела языков Азии и Африки Института востоковедения РАН (Москва), kogan\_anton@yahoo.com

*Курицына Анна Вячеславовна* — младший научный сотрудник Институт классического Востока и античности НИУ ВШЭ, linnansaari@gmail.com

*Молина, Мария Александровна* — младший научный сотрудник Института языкознания РАН (Москва), maria.molina@me.com

*Трофимов, Артем Александрович* — кандидат филол. наук, научный сотрудник Школы актуальных гуманитарных исследований РАНХиГС, artemii.trofimov@gmail.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. All such submissions should be sent to the managing editor:

G. Starostin  
Institute for Oriental and Classical Studies  
Russian State University for the Humanities  
125267 Moscow, Russia  
Miuskaya Square, 6  
E-mail: gstarst@rinet.ru

Articles are published preferably in English or Russian, although publication of texts in other major European languages (French, German, etc.) is possible. Each article should be accompanied with an abstract (not exceeding 300 words) and keywords.

For more detailed guidelines on article submission and editorial policies, please see our website at: <http://www.jolr.ru> or address the editorial staff directly at [gstarst@rinet.ru](mailto:gstarst@rinet.ru).

## 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. Рукописи можно высылать непосредственно заместителю главного редактора по адресу:

125267 Москва  
Миусская площадь, д. 6  
Российский государственный гуманитарный университет  
Институт восточных культур и античности  
Г. Старостину  
E-mail: gstarst@rinet.ru

Предпочтительные языки публикации — английский или русский, хотя возможна также публикация статей на других европейских языках (французский, немецкий и т. п.). К каждой статье обязательно прикладывается резюме (не более 300 слов) и список 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 же непосредственно, обратившись к редакции по электронной почте ([gstarst@rinet.ru](mailto:gstarst@rinet.ru)).



## Some notes on Euskaro-Caucasian phonology

This paper explores a few features of the proposed reconstruction of Euskaro-Caucasian, the putative ancestor of Basque and the North Caucasian languages, as put forth in a recent monograph. Here some features of the consonantal system are discussed, namely (I) the development of proto-Euskaro-Caucasian *\*m* in Basque, (II) the non-initial Basque reflexes of Euskaro-Caucasian laryngeals, and (III) the Basque noun stem allomorphs involving an alternation between /rc/ and /št/. It is shown how these details of Euskaro-Caucasian comparative phonology illuminate important unsolved problems of historical phonology on both the Basque and North Caucasian sides.

*Keywords:* Basque language, North Caucasian languages, comparative linguistics, Euskaro-Caucasian language family, Proto-Euskaro-Caucasian, labial nasals, laryngeals, ‘best explanation’ hypothesis

In a recent monograph (Bengtson 2017: henceforth BCR) I have undertaken to outline a comprehensive scheme of comparative and historical phonology of the Euskaro-Caucasian languages, based on the comparison of my version of Proto-Basque, which relies heavily on Michelena’s *Fonética histórica vasca* (1961, 1990, hereafter FHV), with Proto-North Caucasian and subsidiary proto-languages, according to the *North Caucasian Etymological Dictionary* by Nikolaev & Starostin (1994, hereafter NCED). Since the complete discussion of historical phonology comprises hundreds of pages of the monograph, in this paper only selected aspects are discussed.

It will be reasoned here that an examination of this selection of phonological features helps to establish that the Euskaro-Caucasian hypothesis represents the best explanation for external relations of the Basque language, and concomitantly of the North Caucasian language family, and the fruitfulness of this hypothesis is demonstrated by the cross-validating nature of the three selected features (Proto-Basque *\*m*; Proto-Basque *\*h* and *\*fi* in non-initial positions; Basque noun stem allomorphs with the alternation /rc/ ~ /št/), in the sense that a particular etymology may exhibit the functions of two or even three of the selected features (see **Conclusions**).<sup>1</sup>

### I. The supposed non-existent phoneme *\*m* in Proto-Basque

Michelena’s FHV is an exceptionally important pioneering work that remains the most valuable reference for anyone attempting a study of Basque historical phonology. In the nineteenth chapter Michelena tackled the ancient consonantal system of Basque and proposed the opposition of fortis (tense or voiceless) vs. lenis (lax or voiced) consonants:

FORTIS:	–	t	c	ć	k	N	L	R
LENIS:	b	d	s	ś	g	n	l	r

---

<sup>1</sup> It is important to note that the putative “Euskaro-Caucasian” (EuC, or “Vasco-Caucasian”) hypothesis is in no wise a contradiction of S. A. Starostin’s “Sino-Caucasian” hypothesis; rather, EuC is considered to be a subgroup of Sino-Caucasian, as discussed by Bengtson & G. Starostin (2015).

Notably absent from Michelena's scheme are the fortis labial /p/, the nasal labial /m/, and the fortis and lenis shushing sibilants /č/ and /š/, all present in modern and historically attested Basque.

Here we are only concerned with the purported absence of /m/ in Proto-Basque. Note that Michelena's proposal is hedged, as usual, with his characteristic caveat that such a hypothetical scheme should be treated with reservations, *i.e.*, it was not a hard and fast doctrine.<sup>2</sup> Nevertheless, Trask (EDB 14) states as an absolute law "that no \***m** can be reconstructed for Pre-Basque. The modern **m** was introduced into the language by the phonological changes P\*\*{2}, P\*\*{3} and P\*\*{7}, described in section 6."

The phonological rules Trask referred to are:

**P2. Assimilation of /n/**

P1 [Intervocalic /n/ loss] was competing with another change, which sometimes won out. Intervocalic \*/n/ was sporadically assimilated to a preceding high vowel: to /ñ/ after /i/, to /m/ after /u/. The first was very frequent, the second uncommon. For example, Latin *pinum* 'pine' was borrowed as \**pinu*, which developed to *piñu*, and Latin *cunam* 'cradle' was borrowed as \**kuna*, which yields modern *kuma* in some varieties, alongside *kua*.<sup>3</sup> This development, where it occurred, protected the nasal from loss by P1.

**P3. /nb/ reduction**

The cluster \*/nb/ is usually reduced to /m/. For example, \**senbe* yields common *seme* 'son', and \**unbe* yields common *ume* 'child'.

**P7. /b/ nasalization**

Original \*/b/ has almost invariably changed to /m/ in the configuration \*/bVn/, by a process of nasal assimilation. This process clearly preceded P1. For example, original \**banats* 'grapes' has developed to \**manats* and then to *mahats*; original \**bini* 'tongue' has developed \**mini* > *mihi* (and other variants); and Latin *sabanum* 'covering', borrowed as \**sabanu*, appears today as *zamau* in some varieties. In rare cases, mainly in loan words, \*/b/ developed to /m/ even in the absence of a conditioning nasal, as in \**zubel* 'holm oak' > B *zumel*. (Trask 2008: 25–26)

S. A. Starostin (1996: 122) objected to the notion of the complete absence of /m/ on the grounds of typology: "In Eurasia, at least, I do not know of any languages without a labial nasal ..."<sup>4</sup> The prominent Vasconist Jacobsen (1995: 122) also demurred: "... I see no clear evidence that \**m*- was absent from early Basque. ... I am not convinced that this assumption suffices to exclude from being native Basque several words in our sets [under discussion]: *moko* 'beak, front face' ..., ... *muin* 'pith, marrow, inner part' [also 'brain'] ..., and *mutur* 'snout, muzzle' ..."<sup>5</sup>

<sup>2</sup> "En un trabajo aparte [Michelena 1957], basándonos principalmente en semejanzas en el comportamiento de los fonemas, sobre todo en la coincidencia de la neutralización de oposiciones en contextos análogos, llegamos, *con todas las reservas que impone una restitución tan hipotética como ésta*, a proponer la siguiente reconstrucción del sistema fonológico principal del vasco antiguo: ... [In a separate work (Michelena 1957), based mainly on similarities in the behavior of phonemes, especially in the coincidence of the neutralization of oppositions in analogous contexts, we arrive, *with all the reservations that such a hypothetical restitution imposes*, at proposing the following reconstruction of the main system of the phonology of ancient Basque: ...]" (Michelena 1990: 373–74: *emphasis added* by JDB).

<sup>3</sup> More precisely: Latin *cunam* 'cradle' > (AN, Azk, Sal, R) *kuna* > (B, G, EB) *kuma*, alongside (AN, Sal) *kua* (< \**kūa*) and other variants: (Z) *khüña*, (A) *guia*, (G) *kuia*, (AN) *kuba*, etc. (FHV 304; OEH).

<sup>4</sup> In a response to this statement Trask (1996: 128) lists a number of languages, mostly in the Americas, that allegedly lack the phoneme /m/; outside of the Americas the only languages cited are the Burmic Hakka, Kpelle (Mande, West Africa), and the Papuan Rotokas. The absence of /m/ in Hakka and Kpelle is contrary to the information found in Ruhlen (1975: 203, 225); and WALS (Feature 18A: Absence of Common Consonants) cites only one language (Eyak, now extinct, in North America) with no bilabials or nasals, and none on other continents. Trask's claim that Navajo has no /m/ would be a surprise to the thousands of Navajos who use words such as '*amá* 'mother, maternal aunt' and *mā'ii* 'coyote' (though the existence of /m/ in Proto-Athabaskan is clearly questionable).

<sup>5</sup> In BCR these words are connected with North Caucasian and other words; see the BCR sets **Z.3**, **A.1**, and **A.18**, respectively, and set I.19 in this paper.

There is no dispute here with Trask's P2 rule, but some of the examples given for P3 and P7 are highly questionable. For example, *\*senbe* and *\*unbe* are based on the Aquitanian inscriptional forms SEMBE, OMBE- / VMME, which may or may not be identified with Basque *seme* 'son' and *ume* 'young, offspring, child' (better *\*hume*, as in BN *hume*, Z *hüme*), which seems to be wobbly evidence for this grand generalization. There is no direct evidence for *\*banats* 'grapes',<sup>6</sup> where the *\*b* is simply generalized from other cases of *\*bVn* > *\*mVn*. Evidence is also dubious for the idea that *zumel* 'holm oak' comes from *zur* 'wood' + *\*bel-* 'black' (EDB 134, 381), a somewhat folk-etymological guess unsupported by external evidence.<sup>7</sup> As affirmed by Jacobsen (1995: 122) the whole problem with this claim of the non-existence of /m/ is overgeneralization; just because *some* instances of Basque /m/ can be ascribed to Trask's P2, P3, and P7 rules does not necessarily mean that *all* /m/s are of secondary origin. In Azkue's (1905) standard dictionary of Basque words with initial /m/ span pages 733 through 790;<sup>8</sup> it is hard to believe that *all* of these /m/s are of secondary origin or borrowed. As will become clear in the etymologies cited below, there is no support in the Euskaro-Caucasian hypothesis for an absence of the phoneme /m/ in Proto-Basque, since a significant number of words in both Basque and North Caucasian which are considered putative cognates have *\*m* in both proto-languages.

Based on comparison of Basque with PNC there is substantial evidence that PEuC *\*m* is retained as Basque /m/, initially and medially, so the following etymologies require us to re-examine the proposal by Michelena (and enthusiastically seconded by Trask) that Pre-Basque had no phoneme *\*m*.

On the other hand, we can find a certain grain of truth in Michelena's idea, in that no Basque dialects permit /m/ in final position, and all final nasals become /n/ or /ñ/ (see *\*gain*, *\*sain*, and other examples, I.36–I.42). Furthermore PEuC clusters of the type *\*-mgw-*, *\*-mχw-*, *\*-mxxw-* become Basque *\*b*; and in most other medial PEuC clusters original *\*m* becomes /n/ or disappears in Basque. In the examples the letter-number combinations in bold, like **C.28**, **P.20**, refer to the etymological sets in BCR, which provide more detailed information: attested NC words and local Basque variants, and phonetic and semantic notes; only some of the representative forms are cited here.

### Initial Basque *\*m-*: < PEuC *\*m-*, *\*mH-*, *\*Hm-*, *\*n-* ... *-w-*:

- [I.1] Basque *\*maguli* / *\*malugi* 'strawberry'; (AN) *maguri*, (Bzt, Sal) *malubi* id. ~ PEC *\*niw<sub>G</sub>V* 'a kind of berry'; Akhwakh *muq:ali* 'blackberry', Tabasaran *niq:<sup>w</sup> / nimq:* 'strawberry' (**P.20**)
- [I.2] Basque *\*mahać* 'grape(s)'; (BN, L) *mahats*, (B-Bergara, Leintza) *magats* id. ~ (with metathesis) PNC *\*ġämćō* 'apple; medlar'; Chechen *hamc* 'medlar', Archi *anš* 'apple' (**P.17**)<sup>9</sup>
- [I.3] Basque *\*mati-n*, *\*mal-* 'weak, soft, flexible'; (L) *mali-n* 'weak' ~ PEC *\*mhālV-* / *\*mhänV-* 'warm, weak'; Chechen *mal-dan-* 'to weaken, loosen' (**R.63**)
- [I.4] Basque *\*malka-ř* 'straw'; Basque (AN-Arakil, R, Sal) *malkar* 'straw; stems'<sup>10</sup> ~ PEC *\*mhělq<sub>w</sub>V* 'root; shoot, offshoot'; Ingush *merba* 'shoot, offspring; a single straw' (**C.28**)

<sup>6</sup> BCR prefers the etymology positing Basque *\*mahać* 'grapes': see etymology [I.2].

<sup>7</sup> Cf. instead North Caucasian words such as Budukh *žumal*, Lezgi *ć:umal*, Agul *žimil* 'cornel (tree)' (BCR C.2).

<sup>8</sup> Likewise in Aulestia & White's (1992) dictionary, where words with initial /m/ are found on pages 272–295.

<sup>9</sup> Chechen *hamc* 'medlar', Khinalug *mić* 'apple', etc. For the semantic change of Basque 'grape' ~ NC 'medlar, apple' cf. Rumanian *poamă* 'fruit, apple', Moldovan *poamă* 'grape', French *pomme* 'apple', *pomme de terre* 'potato', Venetian *pomela* 'laurel', etc. (REW 6645; Buck 5.71).

<sup>10</sup> More precisely, 'straw (of vetch); stems of bean plants after threshing' (OEH).

- [I.5] Basque \**manhai* ‘table’; (Z) *māhāñ* id. ~ PEC \**mfiǎʔnē* ‘door; part of door’; Bezhta *maha* ‘door-frame’ (Q.9)<sup>11</sup>
- [I.6] Basque \**mar[c]-fiarde* ‘fork, pruning hook’; (BN) *martxite* ‘pruning hook’<sup>12</sup> ~ PNC \**mǎrc(w)Ē* ‘knife, sickle’; Hunzib *mišu* ‘sickle’; Chamali *mirča* ‘dagger’ (Q.17)
- [I.7] Basque \**mardo* ‘robust; luxuriant’, \**ardu-l* ‘healthy, strong’; (G, B) *ardul* ‘substantial, robust, sturdy’ ~ PEC \**mōrLV* ‘male; man, hero’; Archi *mett-le* ‘male’, Chechen *mayra* ‘husband; brave man’ (R.58)
- [I.8] Basque \**maruhi* / \**mahuři* ‘strawberry’, \**mar-* / \**mar* ‘blackberry’; (BN-Aldude) *mahurri* ‘strawberry’, (Z) *marhüga* ‘mulberry, blackberry’ ~ PEC \**mer(?)V* ‘a kind of berry’; Lak *mamari* ‘blackberry’ (P.18; see I.56)
- [I.9] Basque \**meda-ı̄* ‘narrow, tight’; (G, AN) *medar* ‘narrow, tight’<sup>13</sup> ~ PEC \**fm̄V̄ǎV̄* ‘thin, narrow’; Chechen *nılχa* ‘thin, sparse’, (with metathesis) Bezhta *tłämmö* ‘thin, narrow’, etc. (R.39)
- [I.10] Basque \**mihu-ri* ‘seed, kernel’; (BN) *mihuri*, (Z) *mühüi* id. ~ PEC \**mǎhwV* ‘grain, seed; a kind of cereal’; Akhwakh *mühı* ‘seed, grain’ (C.33)
- [I.11] Basque \**miko* ‘a little, little bit’; (BN, L) *miko* ‘a little, a little bit, a pinch’ ~ PEC \**mikwV* ‘small; young one’; Budukh *mik’e* ‘few; a little, small’ (R.32)
- [I.12] Basque \**minhi* ‘tongue’, \**minco* ‘speech’, \**mis-to* ‘sting (of bee), bite (of snake)’; (Z) *mı̄hı̄*, (B-arc) *mı̄ı̄*, *mı̄* ‘tongue’; northern Basque *mintzo* ‘speech, voice’<sup>14</sup> ~ PNC \**mēlçı̄* ‘tongue’; Andi *mic’i*, Tabasaran *melz*, Dargi (dial.) *limc:i* id. (A.9)
- [I.13] Basque \**moc* ‘sex organs’; (c) *motz* ‘cunnus, female genitals’, *motz-ak* ‘male genitals’ ~ PEC \**mǎçu* ‘hair; fibre, strip’; Avar *móc’o* ‘strip, piece’ (A.61)<sup>15</sup>
- [I.14] Basque \**moku-ı̄* ‘tree trunk, tree base’; (R-Uztárróz) *mokor* ‘stunted, rounded tree’ ~ PEC \**mǎhōqwe* ‘oak tree’; Tsakhur *moq’* ‘oak tree, acorn’ (C.42)
- [I.15] Basque \**mardo* ‘bunch, group’; (B, G, AN) *mordo* ‘bunch (of grapes); group, bunch, abundance’<sup>16</sup> ~ PEC \**mǎr[χ]o* ‘handful, armful’; Hinukh *mettu* ‘handful’, Chechen *mor* ‘armful’ (L.9)
- [I.16] Basque \**moś-ko-ı̄* ‘tree trunk’; B, G, AN) *moskor* ‘trunk (of a tree), burl’<sup>17</sup> ~ PNC \**muçŬ* / \**çumŬ* ‘stick, wood’; Dargi *murç’* ‘rod, stick’ (C.41)

<sup>11</sup> For NC ‘door’ ~ Basque ‘table’, probably from ‘board, plank’, from which either tables or doors may be made; for typology cf. Indo-Aryan: Old Indic *bhitta-* ‘split timber’; ‘fragment; wall’ > Dameli *bit* ‘rafter, roof-plank, table’; Pashai (Areti) *bidók* ‘table’; Panjabi *bhitt* ‘half of double door, window’, Panjabi (Kangra) *bhittu* ‘board, door, half of double door’, etc. (CDIAL 9493).

<sup>12</sup> For the element \**-fiarde* cf. Bsq \**šafiarde* ‘fork’. (G, AN, Z) *matxite* ‘machete’ is probably the source of Spanish *machete*, unlikely to be derived from Latin.

<sup>13</sup> Northern Basque *mehar* is blended with Basque \**menhe* ‘thin’ (BCR Z.12), of distinct origin.

<sup>14</sup> A long discussion is devoted to “The case of Basque \**minhi* ‘tongue’ ~ PNC \**mēlçı̄* ‘tongue’” in BCR, pp. 219–24.

<sup>15</sup> Compare the Basque and NC words with Burushaski \**mús-ki* ‘pubic hair’ (SCG 143), an outgroup comparison that supports the original meaning ‘pubic hair’. For semantic typology, cf. Lithuanian *kūšys* ‘female pubic hair; vulva’, *kuškis* ‘wisp, tuft, bunch, shag (of hair)’, Latvian *kūsis* ‘pubic hair’, *kušķis* ‘wisp, tuft, bunch, shag (of hair)’, (dial.) *kūse* ‘vulva’; Persian *kus* ‘pubic hair’ (Steinbergs 1996, §3.2.2).

<sup>16</sup> BCR has Basque \**mordo*, but *mardo* is attested in older Basque authors (e.g., Larramendi, 18<sup>th</sup> c.), and the metathesis *morda* in parts of Bizkaia; Proto-Basque \**mardo* (> (c) *mordo* by anticipatory assimilation) is a perfect phonetic match to PEC \**mǎr[χ]o*. The original lateral (cf. PEC \**χ*) leaves a trace in Basque (BN, L) *molkho*, (L) *mulko* ‘cluster’ (with diminutive suffix *-ko*).

<sup>17</sup> Analyzed as \**moś-* (root) + \**-ko-* (diminutive suffix) + \**-ı̄* (fossilized plural morpheme); see “General notes to etymologies,” following Conclusions.

- [I.17] Basque \**mošu* ‘nose, face, kiss’; (G) *musu* ‘nose, snout, face, lip, kiss, point, tip’ ~ PEC \**mħərçū* ‘protruding part’; Chechen-Ingush *muc’ar* ‘snout, muzzle, trunk’; Avar *móc’:u* ‘teat, nipple; tip’ (A.17)
- [I.18] Basque \**muga* ‘boundary, limit’; (c) *muga* [muʎa] id. ~ PEC \**mörqwǂ* ‘line, stripe’; Chechen *мова* ‘line, row’ (L.2)
- [I.19] Basque \**mune* ‘brain, marrow, pith’; (Bzt) *muiñ* ‘marrow, pith, brains’ ~ PEC \**mǎhnū* ‘brain, head’; Dargi *mehe* ‘brain’, Akhwakh *mina* ‘head’ (A.1)
- [I.20] Basque \**mucu-ř* (< \**murcu-ř*) / \**muštu-ř* ‘snout, muzzle’; (BN, R-Uztárroz) *mutzur-keta* ‘rooting made by snout of a pig’, (B) *mustur* ‘snout, muzzle’ ~ PEC \**mHǎrčwV* ‘snot, mucus, pus’; Andi *maču* ‘snot’, Chechen *marš* id. (A.19)
- [I.21] Basque \**muti-l* ‘boy; apprentice; worker’; (c) *mutil* id. ~ PNC \**mǎłǂǂ* ‘husband, male, worker’; Dargi *muru-l*, (dial.) *murʎu-l*, *murgu-l* ‘man, husband’ (J.4)<sup>18</sup>

### Basque medial \*-m-: < PEuC \*-m-, \*-mH-:

- [I.22] Basque \**ama* ‘mother’, \**eme* / \**ema-* ‘female’ ~ (with gender reversal) PEC \**ǂemV* ‘father; paternal uncle’; Avar *emé-n* ‘father’, Akwakh *ima* ‘father’, Tsakhur *emi* ‘paternal uncle’, etc. (J.10)<sup>19</sup>
- [I.23] Basque \**čamo* ‘weevil, caterpillar’; (B) *txamo* ‘caterpillar’, (G) *txamo* ‘weevil’, ~ PEC \**čǎmhV* / \**čwǎnhV* ‘flea; nit’; Hinukh *č’ima* ‘nit’, Tsakhur *č’in* ‘flea’ (B.45)
- [I.24] Basque \**čehume* ‘half-span’; (BN) *xehume* id.<sup>20</sup> ~ PNC \**čǂǂmhV* ‘span’; Tsakhur *č’im* / *č’um*, Agul *č’ib* / *č’eb* id.<sup>21</sup> (L.7)
- [I.25] Basque \**čimiča* ‘bedbug’, \**čimi-ri-* ‘butterfly’; (B, G, Sal, R) *tximitxa* ‘bedbug’, (G) *tximirrika*, *tximirrita* ‘butterfly’ ~ PEC \**čimVčV* ‘fly, butterfly’; Tabasaran *č’amč* ‘fly’, Lak *č’imuč’a-li* ‘butterfly’ (B.39)<sup>22</sup>
- [I.26] Basque \**e=meni* ‘to put’; (BN-arc) *imeni*, *imini*, (B, BN, Sal) *imiñi* id., alongside dissimilated forms (B, G, AN) *ibini*, *ipini*, *ipiñi*, etc. ~ PEC \**ǂima-* ‘to stay, to be’ < \**ǂiman-*; Archi *imm(a)-eχ:as* ‘to stay’, Khinalug =*mä* ‘to be’ (V.27)<sup>23</sup>
- [I.27] Basque \*(eci-)[*lamu*] ‘two days after tomorrow’; (B-arc) *etzi-luma*, (B) *etzi-limo*, *etzi-lume*, (A) {*ešçirimu*} / *ecirimu* / id. ~ PEC \**ǂǂmV* ‘one day (two days) after tomorrow’; Batsbi *lamö* ‘day after tomorrow’, Karata *la-di* ‘two days after tomorrow, three days hence’ (H.9)
- [I.28] Basque \**hama-(ř)* ‘ten’; (c) *hamar*, *amar* id. ~ PEC \**χǂmHǎ* / \**mHǂǂ* ‘handful’; Agul *χab* ‘armful’, Udi *maχ’a* ‘palm, handful’ (M.7)<sup>24</sup>
- [I.29] Basque \**lami-* / \**limi* (~ \**mili-*) ‘to lick, taste’; (AN, Sal, R) *lami-ka-tu* ‘to lick’ ~ PEC \**ǂamV* ‘licking, to lick’; Andi, Karata *lam-* ‘to lick’ (V.5)

<sup>18</sup> Note the similar suffix *-l* in Basque and Dargi.

<sup>19</sup> In their global survey Bancel & Matthey (2013; Bancel, et al. 2015: see Table 1) found that (P)APA words, which usually denote ‘father’ or ‘dad’, denote ‘mother’ (or ‘mother’s sister’) in 6% of languages surveyed, while (M)AMA words denote ‘father’ (or ‘father’s brother’) in 26.1% of languages surveyed. In this case only two Dagestania families (Avar-Andian and Lezgian) have words for ‘father’ (‘paternal uncle’ in Lezgian) derived from PEC \**ǂemV*, while other North Caucasian words for ‘father’ derive from the expected (P)APA or (T)ATA types.

<sup>20</sup> A measure = distance between thumb and forefinger.

<sup>21</sup> A measure = distance between thumb and little finger

<sup>22</sup> This seems to be an expressive variant of a similar root: Basque \**simi(n)c(a)* ~ PNC \**ǂimiǂǂ* (see I.32).

<sup>23</sup> Older \**ǂiman-* is based on Hurrian *mann-*, Urartian *man-* ‘to be’; Burushaski \**man-* ‘to be, become’; Proto-Yeniseian \**pan-* ‘to sit’ (SCG 140).

<sup>24</sup> The difference of Basque /a/ ~ NC/o/ could reflect ablaut allomorphs: see “General notes to etymologies,” following Conclusions.

- [I.30] Basque \**limu-ri* ‘slippery, moist, smooth’; *limuri* (Z) ‘moist, humid’, (G, AN) ‘smooth, soft’, (Z, Sal) ‘slippery’, etc. ~ PEC \**ḤHwemV* ‘liquid’ (adj.); Avar *t̤:amí-ya-*, Andi *t̤:emi*, Lezgi *žimi* id. (R.53)
- [I.31] Basque \**sama-ř* (expressive diminutive \**čama-ř*); (AN, BN, Sal, Bzt) *zamar* ‘fleece, mass of wool shorn from a single sheep’ ~ PEC \**čhwěme* ‘eyebrow’; Kryz *fül-c’äm*, Lak *it:a-c’ani* id.<sup>25</sup> (A.23)
- [I.32] Basque \**simi(n)c(a)* ‘bedbug’; (L-Gethari) *zimintza* ~ PNC \**zimižǺ* ‘a kind of stinging insect’; Dargi (Chiragh) *zimiza-l* ‘ant’, Tabasaran *c’imc* ‘ant’ (B.38)<sup>26</sup>
- [I.33] Basque \**sumhe* ‘a kind of tree (willow, oak)’; (Z) *zúnhar* ‘a kind of elm tree’, (Sal) *zugar* id. ~ PEC \**žfūmV* ‘bush, herb; a kind of fruit’; Chechen *žolam* ‘small bushes, shrubs’ (< \**žǫlVm* < \**žǫm-Vl*) (C.2)<sup>27</sup>
- [I.34] Basque \**śama* ‘neck’; (B) *sama* ‘neck, throat’ ~ PEC \**šVmV* / \**mVšV* ‘spine; edge’; Avar *š:omé-li* ‘edge, end’ (A.29)
- [I.35] Basque \**śami-n*; (B, G, BN, L, Z) *samin* ‘bitter, sour’, also (BN, L, Z, Bzt) ‘pungent, piquant, spicy’ ~ PEC \**čěmhV* / \**čwěnhV* ‘salt, salty’; Avar *c’:am* ‘salt’ (R.15)

### Basque stem-final /n/ < PEuC \*-m-, \*-Hm-, \*-mH-:

- [I.36] Basque \*(beHa-)sun ‘bile, gall’; (BN) *behazun* ‘bile, gall; hatred, bitterness’, (L-Sara) *behazuñ*, *beazuma* id.<sup>28</sup> ~ PNC \**čwǻymě* ‘gall, anger’; Archi *s:am* ‘gall’, Dargi (Chiragh) *sume* ‘gall, anger’, (A.88)
- [I.37] Basque \**gain* ‘top, surface; above’; (Z) *gañ* ~ PEC \**hqwěmV* ‘head, horn’; Tsakhur *q’om* ‘summit, top’, Andi *ɠʷon* ‘horn’ (I.13)
- [I.38] Basque \**hagin* ‘tooth’ < PEuC \**HaḡimV*; (BN, L) *hagin* ‘molar tooth’ (B) *agin* ‘tooth, incisor tooth’ ~ (with transposition) PNC \**HěmḡǺ* ‘to bite’; Khinalug *q’ini* ‘to eat’, Karata *q’:am-* id. (A.11)<sup>29</sup>
- [I.39] Basque \**kain*; (B) *kain* ‘fog, vapor, mist, storm clouds’ ~ PNC \**k̲wǻmhV* ‘cloud, fog’; Khwarshi *qema* ‘clouds’ (G.6)
- [I.40] Basque \**sain* ‘vein, root, nerve, root (of hair)’; (Z) *zań* ‘nerve, vein, artery; root’ ~ PEC \**šěhmV* / \**hěmsV* ‘muscle, vein; intestine’; Chamali *s:ē*, *s:ey* ‘sinew, muscle’, Dargi *šems* ‘muscle’ (A.58)
- [I.41] Basque \**sen*-[-bera] ‘soft cheese, cream cheese’; (Z) *zen-béa*, (R) *zén-bra* id. ~ PEC \**Hžěmī* ‘curds, butter’; Hinukh *cen* ‘curds’ Udi *č:äin* ‘butter’ (P.6)
- [I.42] Basque \**soin* ‘shoulder, (upper) back’; (AN, BN, L) *soin* ‘shoulder (general); midsection (of pork)’, ~ PEC \**Hčwěymǻ* / \**Hčwěynǻ* ‘leg bone’; Tsakhur *c’om* ‘shin-bone’ (A.38)

<sup>25</sup> The NC words are compounds with a word meaning ‘eye’, implying that PEC \**čhwěme* itself originally had a more general meaning (‘hair, fur’), like its Sino-Caucasian cognates: Burushaski \**še[m]* ‘wool’; Proto-Yeniseian \**cəje* ‘hair’; Proto-Sino-Tibetan \**chām*: Lushai *sam* ‘hair (of head)’, Kinnauri *tsam* ‘wool’, Garo *mik-sam* ‘eyebrow’, Limbu *mik-sāŋ* ‘eyelashes’ (SCG 27).

<sup>26</sup> See I.25, which seems to be an expressive variant of this root.

<sup>27</sup> Both Basque and EC have derivatives with the suffix \*-atV: Basque: (B) *zumel* ‘holm oak’, (B, G) *zumal-ika-r*, *zumal-aka-r* ‘a kind of willow or osier’, (G) *zumal-di* ‘grove’: EC: Budukh *žumal* ‘cornel’, Kryz *žimel*, Agul *žimil* id.; Hunzib *šumal* ‘bushes, shrubs’; (with metathesis) Chechen *žolam* ‘small bushes’; Avar *žulám* / *žulán* ‘cornel’, etc. (NCED 1107).

<sup>28</sup> The Bsq word is a compound of \**beha-* + \**-sun*, the latter of which is compared with the EC words.

<sup>29</sup> The transposition \**HaḡimV* ~ \**HěmḡǺ* is typical of Sino-Caucasian and Euskaro-Caucasian (see BCR 90-91). “There are many cases where we have to reconstruct verbal root structures like \*CVRV or \*RVCV, based on [Sino-Tibetan] and Yeniseian evidence, while in NC their reflexes look like \*VRCV (with a reduction of the second vowel, and a regular development -CR- > -RC-)” (SCP, page 1). There seems to be a transposition similar to that of Basque \**hagin* in Andian languages: Andi *q’:am-*, Chamali *q’:an-* ‘to bite’, etc. (NCED 559).



**PEuC \*-mC- clusters:** PEuC \**m* in these clusters is highly unstable in its Basque reflexes. In some cases PEuC \**-m-*, in clusters with sibilants, seems to disappear without a trace. Some convergent North Caucasian developments are cited as well:

- [I.43] Basque \**e=sari* ‘to put, sit’ ~ PEC \*=*VmċVr* ‘to stand, stay’; Tsakhur ul’o=*zar-* ‘to stand’ (V.31)
- [I.44] Basque \**hace, \*hac*; (B, G, AN) *atze* ‘back part (of house, boat, etc.)’, (BN, L) *hatz, atz* ‘track, trace, vestige, footprint’<sup>30</sup> ~ PNC \**hāmċĔ* ‘half, middle’; Tabasaran *hac* ‘half; body from head to waist’, Ubykh *ža* ‘back part’ (L.8)
- [I.45] Basque \**hic* ‘word’; (BN, L, Z) *hitz* ‘word’ ~ PNC \*=*[i]mċŪ* ‘to speak, tell, talk’; Avar =*ic-* ‘to say, tell’ (L.10; V.52)
- [I.46] Basque \**isu* ‘fright, terror’; (B-Ispaster, Markina, G) *izu*, (AN, L, Z, R, Sal) *izi* id. ~ PNC \**hVmċV* ‘fear, fright’; Hunzib *hīč-* ‘to fear, afraid’ (K.6)
- [I.47] Basque \**listo-* ‘hornet, wasp’, \**e=lco* ‘gnat, midge, mosquito’<sup>31</sup> ~ PEC \**λāmċV* ‘a biting insect’; Akhwakh *ʔac’:u* ‘ant; bug, bedbug’ (B.31)
- [I.48] Basque \**sisali, \*sisa-ni*; (Bzt) *zizari* ‘worm’, *zizain* ‘leech’ ~ PEC \**sēmšā* ‘tadpole, water flea, bug, ant’; Chechen *sēsa* ‘tadpole’, Tsakhur *semsä-lä* ‘ant’ (B.47)

In others PEuC \**-m-* is associated with the development of the Basque labial vowels /u/ or /o/:<sup>32</sup>

- [I.49] Basque \**honda-ř* ‘sand, residue, bottom’; (BN, L) *hondar* ‘sand’ ~ PEC \**ʔamtV* ‘earth, ground’; Khinalug *ant* id. (D.18)<sup>33</sup>
- [I.50] Basque \**hu[m]ki* ‘to touch’; (BN, L) *hunki-tu*, (B) *uku-tu, iku-tu* id.<sup>34</sup> ~ PEC \**fimķwV* ‘fist’; Bagwali *hunk’a*, Archi *χ:ik* id. (V.13)
- [I.51] Basque \**[H]uśu* ‘pus, serum’; (BN) *usu* ~ PNC \**h[a]mzū* ‘saliva, mucus; herpes’; Lak *unts’u* ‘snot’ (A.93)
- [I.52] Basque \**i=hunc* ‘dew’; (A) *iunz* id. ~ PEC \**xwĕmċwĕ* ‘bog, dirt’; Lak *xunc’a* ‘bog’ (E.11)
- [I.53] Basque \**i=soc* ‘frost, ice’; (c) *izotz* < PEuC \**zVmċ(w)V* ? ~ PEC \**zwnċ(w)V* ‘snowdrift, snowstorm’; Lak *zanč* ‘snow drift (in the mountains)’ (G.16)<sup>35</sup>
- [I.54] Basque \**konko-ř / \*kunku-ř* ‘hump, joint, bone’; (R) *kunkur* ‘joint (of bones)’ ~ PNC \**qāmċ(w)ă* ‘knee, leg-bone’; Dargi *q’u(n)q’a* ‘knee’ (A.76)
- [I.55] Basque \**ondo*<sup>1</sup> ‘joint’; (c) *gar-ondo* ‘nape, back of neck’ ~ PEC \**ʔamdV* ‘joint, bone’; Avar *hod* ‘spine, backbone’ (A.77)

<sup>30</sup> “Es razonablemente seguro que *atze* [‘back part, past’, etc.] no es sino *hatz* ‘huella’ [‘track, trail’, etc.] que, en su empleo local, adquirió no sólo un nuevo valor, sino además, una *-e* adventicia” (OEH).

<sup>31</sup> EB *liztor* ‘wasp’, *eltxo* ‘mosquito’ (Aulestia & White 1992); these seem to be permutations of the same root. Basque \**listo-* is an example of Basque “st-clusters” that correspond, under certain conditions, to PNC sibilants (usually tense) \**š*, \**ç*, \**ç*, \**ž*, \**š* (see BCR 138).

<sup>32</sup> To judge from North Caucasian evidence, Basque \**hog<sup>[w]</sup>i* and \**ortú-* originally had labial vowels: Basque \**hog<sup>[w]</sup>i* ‘gums (mouth flesh)’, (BN, L, Z) *hobi* ~ PEC \**hōmċwĕ* ‘mouth, throat’ (A.13); Basque \**ortú-ši* ‘take off (shoes, stockings)’; \**ortú-č* ‘barefoot’ ~ PEC \*=*ōmċV* ‘to put on (trousers, shoes)’ (R.9; V.34); the Basque suffixes \**-ši* and \**-č* apparently reverse (‘take off footwear’) the sense in EC (‘put on footwear’) (BCR 443).

<sup>33</sup> Because this PEC reconstruction was based on only three Dagestanian languages, Tsez, Hinukh, and Khinalug, the nature of the internal nasal (whether \**m* or \**n*) was not clear. Based on the Basque evidence with the labial vowel /o/, \**m* may be preferable for the PEuC stage.

<sup>34</sup> In this word we have an unusual contrast between NE Basque /un/ ~ /on/ (*hunki-*, *onki-*) vs. SW Basque /u/ (*uki-*), thus a tentative reconstruction \**hu[m]ki-*.

<sup>35</sup> The lack of /n/ in the Basque word could suggest that it came from PEuC \**zVmċ(w)V*, related by labial metathesis to PEC \**zwnċ(w)V*. An \**m*-cluster would more readily explain the loss of nasal and development of the labial vowel /o/ than an \**n*-cluster.

- [I.56] Basque \**susa* (in \**mar-susa* ‘blackberry, mulberry’); (R) *mar-zuza* ~ PEC \**zim̥V* ‘a kind of berry (or flower, blossom)’; Agul *ziwz-ar* ‘strawberry’ (P.19; see I.8 for the morph \**mar-* / \**maɾ-*)

From PEuC clusters of the type \**-m̥gw-*, \**-m̥χw-*, \**-mxw-* nasality is lost and they become Basque \**-b-*, in some cases only in certain dialects:

- [I.57] Basque \**hibai* ‘river’; (BN, L) *hibai* id. ~ PEC \**yim̥χwÄ* ‘river’; Godoberi *inχ:i* ‘river, brook’ (E.7)
- [I.58] Basque \**hag<sup>[w]</sup>une* / \**gahune* ‘foam’; (BN, L) *hagun*, (L) *haboin*, *habuiñ*, (Z) *gahūñ*, etc. ~ PEC \**xwamV* / \**xwamxwV* ‘foam, liquid’; Dargi (Sirgokala) *x:umx:a* ‘foam’ (E.10)
- [I.59] Basque \**hog<sup>[w]</sup>i* ‘gums (mouth flesh)’ > (BN, L, Z) *hobi*; but (Bzt) *ogi*, (B, G) *oi*, (B) *ui* ~ PEC \**hōm̥gwī* ‘mouth, throat’; Akhwakh *oq’o* ‘throat’ (A.13)

This appears to be a counter-example, not yet explained:

- [I.60: see I.62] Basque \**eiho* ‘to grind’, \**eihera* ‘mill’ ~ PNC \**Hēm̥χwV* ‘to grind’, PEC \**Hēm̥χwV-rV* ‘mill, mill-stone’ (Q.20; V.68)

In several cases the \**-m-* (whether it disappears or remains as /n/) is associated with the development of a Basque diphthong, /ai/, /ei/, /oi/, or /au/:

- [I.61] Basque \**e=aici* ‘to milk (a cow)’; (Z) *jaitzi* ~ PNC \**ām̥zŪ* ‘to milk’; Archi =*ac:a* id. (V.26)
- [I.62 = I.60] Basque \**eiho* ‘to grind’, \**eihera* ‘mill’; (BN, L) *eho* ‘to grind’, (L) *eihera* ‘mill’ ~ PNC \**Hēm̥χwV* ‘to grind’, PEC \**Hēm̥χwV-rV* ‘mill, mill-stone’; > Ingush *ah-* ‘to grind’ *ħayra* ‘mill’, etc. (Q.20; V.68)
- [I.63] Basque \**e=roic* ‘cliff, fissure’; (c) *erroitz* ~ PEC \**hrōm̥çwe* ‘stone’; Khinalug *ric’in* ‘stone’ (D.13)
- [I.64 = I.37] Basque \**gain* ‘top, surface; above’ ~ PEC \**h̥q̥wēm̥V̄* ‘head, horn’ (I.13)
- [I.65] Basque \**gau* / \**gab-* ‘night’; (c) *gau*, (AN-Ulzama) *gabe* id. ~ Proto-Lezgian \**χ<sup>f</sup>:am:* / \**χ<sup>f</sup>:an:* ‘evening, night’; Tsakhur *χ<sup>f</sup>am* ‘night’ (H.5)<sup>36</sup>
- [I.66] Basque \**haic* ‘stone, rock, crag’; (BN) *haitz* ‘rock, stone’ ~ PNC \**hām̥Vçö* ‘stone’; Avar *ħuc*: id. (D.15)
- [I.67] Basque \**Hainc-eihar* ‘frost’;<sup>37</sup> (B-Oñate) *aintzigar* ‘faint frost on leaves’, (G-Tolosa) *antzigar* ‘fine-grained snow’ ~ PNC \**yām̥zÄ* ‘snow’; Tindi *anzi* ‘snow’ (G.15)
- [I.68] Basque \**haise* ‘wind’; (L) *haize* ‘wind, air’ ~ PNC \**zām̥sə* ‘breath, soul; cloud, sky; god’; Akhwakh *as:i* ‘breath’, Udi *haso* ‘cloud’ (G.3)
- [I.69] Basque \**hauć* ‘ashes, powder’; (L) *hauts* id. ~ PEC \**yōm̥çV* ‘earth’; Batsbi *yobst*: id. (F.5)
- [I.70] Basque \**hauśin* / \**haśuin* ‘nettle’; (BN, L) *hausin* id. ~ PEC \**h̥wām̥šwi* ‘a kind of herb or weed’; Agul *haś<sup>w</sup>* ‘caraway’ (C.23)
- [I.71 = I.39] Basque \**kain* ‘fog, clouds’ ~ PNC \**kw̥m̥hV* ‘cloud, fog’ (G.6)

<sup>36</sup> S. A. Starostin cited these forms as Proto-Lezgian \**χI:am:* / \**χI:an:* ‘evening, night’, with a pharyngealized voiceless uvular fricative, the symbol /l/ (палочка) signifying pharyngealization, and reconstructed Proto-Sino-Caucasian \**χV[n]HV* ‘dark’, based on Basque \**gau* ‘night’, Proto-Lezgian \**χI:an:* ‘evening’, PY \**χoʔn-* ‘dark’, and PST \**γVm(H)* ‘dark, shade’ (SCG 245–6; also TOB: Sino-Caucasian Database). With the symbol \*[n:] in the reconstruction Starostin left open the possibility of a different nasal, namely \**m:*, which is supported by the PST form, the Proto-Lezgian alternate form \**χI:am:* ‘evening’ (Tsakhur *χ<sup>f</sup>am* ‘night’), and the Basque labials in \**gau* ‘night’, \**gab-* (in compounds).

<sup>37</sup> The Basque word seems to be a compound of \**Hainc-* + \**eihar* ‘dry’ (BCR R.22), i.e. ‘dry snow’.

- [I.72] Basque *\*koipe* ‘rendered fat, grease’; (AN, BN, L) *goipe* ‘butter, animal fat, melted fat, grease’ ~ PEC *\*kāmpā* ‘butter; sour cream’; Tindi *k’wabi* ‘sour cream’ (P.9)
- [I.73] Basque *\*lainho* ‘fog, mist, vapor’; (BN, L) *lanho* ‘vapor’, (Z) *lānhú* ‘mist, fog’ < PEuC *\*λamHV* ~ (with metathesis) PEC *\*hemλĀ* ‘dream’; Godoberi *hanlu* ‘fog, cloud’, Karata *hantt’u* id. (G.7)
- [I.74] Basque *\*paita* ‘duck’; (B) *paita* id. < PEuC *\*pHām[ʔ]V̄* (SCP 30) ~ PEC *\*bHāmʔV* ‘duck’; Dargi *baʔ* id. (N.24)
- [I.75] Basque *\*saie* ‘vulture’, (arc) ‘eagle’; (B, G, BN, Z) *sai* ‘vulture’ ~ PEC *\*çwāmʔV̄* ‘eagle, vulture’; Karata *c’:ūyi* ‘eagle’, Andi *c’:un* ‘eagle, vulture’ (B.13)
- [I.76 = I.40] Basque *\*sain* ‘vein, root, nerve, root (of hair)’ ~ PEC *\*sēhmV / \*hēm̄sV* ‘muscle, vein; intestine’ (A.58)
- [I.77 = I.42] Basque *\*soin* ‘shoulder, (upper) back’, etc. ~ PEC *\*Hçwēymǎ / \*Hçwēynǎ* ‘leg bone’ (A.38)

In the following cases a trace of the original nasal is left as Basque /n/:

- [I.78 = I.49] Basque *\*fionda-ř* ‘sand, residue, bottom’ ~ PEC *\*ʔamtV* ‘earth, ground’ (D.18)<sup>38</sup>
- [I.79 = I.50] Basque *\*hu[m]ki* ‘to touch’<sup>39</sup> ~ PEC *\*fimkwV* ‘fist’ (V.13)
- [I.80 = I.52] Basque *\*i=hunc* ‘dew’ ~ PEC *\*xwimçwī* ‘bog, dirt’ (E.11)
- [I.81 = I.54] Basque *\*konko-ř / \*kunku-ř* ‘hump, joint, bone’ ~ PNC *\*qāmǎ(w)ǎ* ‘knee, leg-bone’ (A.76)
- [I.82 = I.73] Basque *\*lainho* ‘fog, mist, vapor’ ~ (with metathesis) PEC *\*hemλĀ* ‘dream’; Godoberi *hanlu* ‘fog, cloud’ (G.7)
- [I.83] Basque *\*e=ncu-n* ‘to hear, listen’; (c) *entzun* id., (dial.) ‘to smell (odors)’ ~ PNC *\*=āmçĔ* ‘to know, see’; Hunzib *=āc’-* ‘to see’ (V.48)
- [I.84 = I.67] Basque *\*Hainc-eihar* ‘frost’ ~ PNC *\*yāmzĀ* ‘snow’ (G.15)
- [I.85] Basque *\*lance-ř* ‘drizzle, fog’; (Z) *lántzer* ‘drizzle, fine rain’, ~ PEC *\*lǎ[m]çV / \*lwǎçV* ‘to filter, sift’; Bezhta *tatsa-ri* ‘sieve’ (G.9)<sup>40</sup>
- [I.86 = I.55] Basque *\*ondo*<sup>1</sup> ‘joint’ ~ PEC *\*ʔamdV* ‘joint, bone’ (A.77)
- [I.87] Basque *\*tanta*, *\*tanto* ‘drop, dung pellet’; (c) *tanta* ‘drop (of liquid)’ ~ PEC *\*tēmtV* ‘dirt, dust’; Andi *toto* ‘saliva’ (E.13)
- [I.88] Basque *\*tintV* ‘gland, uvula’; (BN) *tinto-la* ‘uvula’, ~ PEC *\*tēmʔV* ‘freckle, spot’; Bezhta *t’ot’i-lo* ‘pimple’ (A.34)

## II. Archaic laryngeals in Basque

The Basque aspirate /h/, which is still a contrastive phoneme in northern dialects (L, BN, Z), and is attested in older texts of the southern dialects as well, is treated separately in FHV (523-26) by Michelena, who “clearly assumed that a phoneme *\*h* was found in Proto-Basque” (Jacobsen 1995: 123), while Trask (1995: 9) thought that the “pre-Basque aspiration was not etymological, but was merely a suprasegmental feature” (BCR 37).

<sup>38</sup> Because this PEC reconstruction was based on only three Dagestania languages, Tsezi, Hinukh, and Khinalug, the nature of the internal nasal (whether *\*m* or *\*n*) was not clear. Based on the Basque evidence with the labial vowel /o/, *\*m* may be preferable for the PEuC stage.

<sup>39</sup> In this word we have an unusual contrast between NE Basque /un/ ~ /on/ (*hunki-*, *onki-*) vs. SW Basque /u/ (*uki-*), thus a tentative reconstruction *\*hu[m]ki-*.

<sup>40</sup> For semantic typology, cf. Swedish *sil* ‘strainer’, *sila* ‘to strain; trickle, pour (of rain)’, etc.

In North Caucasian “laryngeals are the most unstable class of consonants ... Their exact reflexes are often hard to establish (especially in [Proto-Nakh] and [Proto-Andian], where in some cases we use the symbol H, denoting an arbitrary laryngeal)” (NCED 62). Even more so, in the reconstruction of Proto-Sino-Caucasian “laryngeals in general are very unstable phonemes and tend to disappear easily. They are most numerous and best reconstructed in PNC, with ... correspondences in other families” (SCP 22).

In Basque we can probably reconstruct, at the most, two laryngeals: *\*h*, which remains as /h/ in all three of the aspirating dialects (BN, L, Z = Basque dialects in France), and *\*h̄* which has the reflex /h/ in BN and L but zero in Z. Whether this *\*h* / *\*h̄* opposition should really be traced back to Proto-Basque is, so far, uncertain; possibly it is the result of dialect borrowing or some other force. It is of interest, however, that the PNC uvular fricative *\*χ* always corresponds to Basque *\*h*, never to Basque *\*h̄* (BCR 167-69).

Both *\*h* and *\*h̄* are lost in the non-aspirating dialects (B, G, AN, R, Bzt, Azk, Sal = “Spanish” Basque dialects), except in some of the earliest records, and in some cases where old *\*h* between vowels has become the voiced velar continuant [v], phonemically /g/ and traditionally written as *g*, e.g. (B, G, AN) *igar* [iʋar] ‘dry’ < *\*eihar*; cf. PEC *\*hwēri* ‘morning, dawn’ (BCR R.22), and (R, Bzt) *agur* [aʋur] ‘palm, hollow of hand’ < *\*a=huř*; cf. PEC *\*kHwōrV* ‘hole, pit’ (BCR A.67).

Basque also preserves some laryngeals in combination with resonants, often with the laryngeals no longer existing as such in modern Caucasian languages. For example, the EC etymon represented by Avar *qʷ:warí-da-b* ‘narrow, cramped’ was reconstructed by Nikolayev & Starostin as PEC *\*q̄warHV* ‘thin, narrow’ (NCED 933). The cognate proposed for this word is Basque *\*garhi*, attested with /rh/ in northern Basque *garhi* ‘thin, slender, slim, slim-waisted’; in some Lezgian languages, for example, the reflex of *\*H* may consist of a pharyngealized initial consonant, e.g. Agul *qʷ:ure-* ‘thin, emaciated’.<sup>41</sup>

**Basque transposed laryngeals** (in relation to PNC). In the following cases Basque has a laryngeal *\*h* between vowels or in clusters, while the hypothesized PNC or PEC cognate has a laryngeal in a different position or different vowel structure. Note the contemporary NC cognates with a laryngeal or pharyngeal (/h/, /h̄/, /ʎ/, /ʎ̄/, /ʎ̄/), or a pharyngealized vowel (/ə̄/, /ē/, /ī/, /ō/, /ū/);<sup>42</sup> laryngeals and pharyngeals can also influence accent patterns in Avar,<sup>43</sup> and various other changes to vowels and consonants.<sup>44</sup> Selected Basque and NC cognates are cited:

[II.1] Basque *\*barhe* ‘slug, snail’; (BN, L, Z) *barhe* ‘slug’ ~ PEC *\*bHōr-* (in *\*bHōr-χwVtV* ‘snake, snail, worm’); Lak *bařčalu* ‘snail’ (with pharyngeal /ə̄/) (B.28)<sup>45</sup>

<sup>41</sup> “Pharyngealization (and the pharyngealization of uvulars in particular) [in Tsezian, Lak, Dargwa, Lezghian, Ubykh] is apparently the result of the fall of certain laryngeals ..., therefore we do not reconstruct a special pharyngealized uvular series for PNC” (NCED 59).

<sup>42</sup> “Pharyngealized vowels in Lak, [Proto-Dargwa] and [Proto-Lezghian] usually correspond to each other rather well and can be traced to PEC fallen laryngeals (not only in initial combinations, but in medial clusters as well, ... In addition to Lak, Dargwa and Lezghian languages pharyngealization is also present in [Proto-Tsezian] and [Proto-West Caucasian]” (NCED 83).

<sup>43</sup> “A typical feature of the emphatic laryngeals is that nouns that contain them reveal in Avar an immobile accent paradigm (the so-called ‘Paradigm A’), while words with plain laryngeals (or without laryngeals) have in Avar either an oxyton or a mobile accent paradigm (paradigms B and C)” (NCED 62).

<sup>44</sup> “One can also pay attention to the unmotivated variation of voice/voicelessness (and also glottalization) in some PWC reflexes. The reasons for a secondary voicing in PWC are not clear yet (it could be possibly connected with the fall of laryngeals in PWC)” (NCED 48); “The development of vowels is also influenced by pharyngealization (that appears as a result of the fall of laryngeals)” (NCED 72).

<sup>45</sup> Alternatively the PEC proto-form could have been *\*wHōr-*, since all attested forms belong to languages in which PNC *\*w* and *\*b* have merged as /b/. The same situation holds for the next set.

- [II.2] Basque *\*berho* ‘bramble, thicket’; (L) *berho* id. ~ PEC *\*bHörV* ‘a kind of thorny plant’; Lak *burū* ‘thistle’ (with pharyngeal /ʊ/) (C.14)<sup>46</sup>
- [II.3] Basque *\*bitho* ‘hair, mane’; (BN) *bilho* id. ~ PNC *\*pVhVtV* ‘feather, mane’; Dargi *pəhala* ‘feather’ (with pharyngeal /a/ and voiceless pharyngeal fricative /ħ/) (A.20)
- [II.4] Basque *\*čahal* ‘calf, heifer’; (Z) *xáhal* id. ~ PEC *\*HčwīlV̄* / *\*HlīčwV̄* ‘heifer, calf’; Ingush *šasa* ‘calf’, Avar *šáč’ár* ‘heifer’ (N.11)
- [II.5 = I.24] Basque *\*čehume* ‘half-span’; (BN) *xehume* id. ~ PNC *\*čV̄mhV* ‘span’; Tsakhur *č’im* / *č’um*, Agul *č’ib* / *č’eb* id. (with pharyngeal /i/, /ʊ/ or /e/) (L.7)
- [II.6] Basque *\*ethe* ‘word, speech’; (BN, L) *elhe* id. < *\*e=tehe* ~ PEC *\*lēHwV* / *\*tēHwV* ‘word’; Dargi *luši* ‘number’ (L.11)
- [II.7] Basque *\*geHet̪i* ‘ox; beef, meat’ / *\*i=lo=ket̪V*;<sup>47</sup> (L-arc) *geheli* ‘(fresh) beef’ (17th c.); (BN) *ikhel* ‘fattened ox’, (AN) *ikela* id.; (L) *okhela*, *okheli* ‘meat; piece (of meat, cheese), morsel’ ~ PEC *\*q̄hwēt̪V̄* / *\*q̄wēt̪hV̄* ‘large female domestic animal (cow, mare)’; Dargi *q<sup>ʳ(w)</sup>al* ‘cow’, Lak *ul* ‘cow’, Avar *šalá* ‘mare’<sup>48</sup> (P.12)
- [II.8] Basque *\*fierce* / *\*fiešte* ‘intestine’; (BN, L) *hertze* id., (BN-arc) *halsarr*-ak ‘entrañas’ < *\*fierce-ř*; ~ PNC *\*čĀrHV* / *\*HĀřčV* / *\*rHĀčV* ‘stomach, abomasum’; Avar *šorčo* ‘intestine’, Andi *ručo*-l id. (A.54; SCG 180)
- [II.9] Basque *\*fodoe* ‘cloud’; (BN) *hodei* ‘cloud, thunder’ ~ (with metathesis) PEC *\*dwiHV* ‘wind’; Dargi *dəʃ* id. (with pharyngeal /a/ and voiced pharyngeal fricative /ʕ/) (G.4)<sup>49</sup>
- [II.10] Basque *\*itha-ř* ‘pea(s), bean(s)’; (Z) *ilhar* ‘beans’ < *\*i=hVta-ř* ~ PEC *\*hōwt[ā]* ‘bean(s), lentil’; Tsezi *hil* ‘pea(s)’, Avar *holó* ‘bean(s)’ (O.8)
- [II.11] Basque *\*lahar* ‘thorn, bramble’; (BN, L) *lahar* ‘bramble, creeping thorn’, (Bzt) *lagar* id. ~ PEC *\*λwřřV* ‘leaf’; Karata (Tokita dial.) *alzi*, Lezgi *riq<sup>ř</sup>-äy* id. (C.17)<sup>50</sup>
- [II.12] Basque *\*leher* ‘pine (tree)’; (Z) *leher* id. ~ PEC *\*λfwētū* ‘a kind of tree’; Lak *halu* ‘grove’, Lak (Arakul dial.) *x<sup>w</sup>alu* id., Avar *ł:alú* ‘yew-tree’ (C.7)
- [II.13] Basque *\*luHur* ‘earth, ground, land’; (BN-Baigorri) *luhur*, (B) *luur* id. ~ PNC *\*lhēmŁwř* ‘earth’; Avar *rat<sup>ł</sup>*: ‘earth, ground’, Tsakhur (Gelmets dial.) *nok* ‘field’ (with pharyngeal /o/) (D.1)<sup>51</sup>
- [II.14 = I.2] Basque *\*mahać* ‘grape(s)’ ~ PNC *\*Āmćō* ‘apple; medlar’ (P.17)
- [II.15 = I.8] Basque *\*mařuhi* / *\*mahuri* ‘strawberry’ ~ PEC *\*mer(?)V* ‘a kind of berry’; Lezgi *mere* ‘blackberry’ (P.18)<sup>52</sup>
- [II.16] Basque *\*sahar* ‘old’ (human), ‘old person’; (BN, L) *zahar* ‘old (human, animal, plant), old person’, (Bzt) *zagar* id. ~ PNC *\*swērho* / *\*r(h)ěswō* ‘old; year’; Avar *asrá-ya-* ‘old, worn out’ (R.40)

<sup>46</sup> *\*bHörV* or *\*wHörV* (see the preceding note).

<sup>47</sup> Here we have an unusual case of the extended, unprefixated form *\*geHet̪i* coexisting with forms with fossilized class prefixes (articles), *ik(h)el(a)*, *ok(h)eli/a*; in the latter the syllabic structure is syncopated from three to two syllables and the initial plosive is devoiced, which also occurs in a few other etymologies (BCR 128–29).

<sup>48</sup> Note Dargi pharyngealized /q<sup>ʳ(w)</sup>/, Lak pharyngeal /ʊ/, Avar pharyngeal fricative /ʕ/.

<sup>49</sup> For semantic connections of ‘wind’ ~ ‘cloud’, cf. Basque *\*haise* ‘wind’ ~ Akhwakh *as:i* ‘breath’ ~ Udi *haso* ‘cloud’, etc. (BCR G.3).

<sup>50</sup> The Lezgi form represents metathesis (< Proto-Lezgian *\*riλ<sup>w</sup>*). For semantic typology of ‘leaf’ ~ ‘thorn’ cf. Old Indic *třna-* ‘grass, blade of grass, straw’, Slavic *trůnũ* ‘thorn’, Gothic *þaurmus*, Eng. *thorn*, etc. (Buck 8.51).

<sup>51</sup> Initial Avar /t/ and Tsakhur /n/ are the regular reflexes of PNC initial *\*l-*, when preceding *\*-m-* in the same root; Avar /t<sup>ł</sup>:/ and Tsakhur /k/ are the regular reflexes of the PNC tense lateral affricate *\*L* (NCED 53, 55, 131).

<sup>52</sup> Here NCED suggests a laryngeal with the notation (?). The Basque evidence for /h/ is quite clear.

- [II.17] Basque \**solho* ‘meadow, field’; (L) *sorho* ‘meadow’<sup>53</sup> ~ PEC \**čHätu* ‘earth, ground, sand’; Lak *š:alu* ‘earth, ground’ (with pharyngeal /a/) (D.3)
- [II.18 = I.33] Basque \**sumhe* ‘a kind of tree’; (Z) *zúnhar* ‘a kind of elm tree’, (Sal) *zugar* id. ~ PEC \**šfūmV* ‘bush, grass; a kind of fruit’; Chechen *žolam* ‘small bushes, shrubs’ (< \**žǝlVm* < \**žǝm-VI*) (C.2)

**PEuC medial laryngeal clusters:** In each of these cases northern Basque retains /h/ in a resonant + /h/ cluster /Rh/, and the PNC/PEC cognate also has a /Rh/ cluster, usually in the same order as in Basque; there are two possible cases of metathesis [II.27, II.28]. Selected Basque and NC cognates are cited:

- [II. 19] Basque \**arhe* ‘harrow, rake’; (BN, L) *arhe* id. < \**harhe* ~ PEC \**karhV* ‘harrow’; Avar *vár-ize* ‘to harrow’, *vári* ‘wooden ladder’ (Q.14)
- [II.20] Basque \**betħa-ř* ‘forehead, side, edge’; (BN, L) *belhar* ‘forehead’ ~ PEC \**bġāthǝ* ‘edge, end’; Lezgi *p:el*, Kryz *bel* ‘forehead’ (A.5)
- [II.21] Basque \**butha-ř* ‘breast, chest’; (BN, L) *bulhar*, (Z) *búlhar* id.<sup>54</sup> ~ PEC \**gwālhē* ‘udder, breast’; Avar *vwári* ‘udder’ (A.40)
- [II.22] Basque \**bur(a)ho* ‘curse, blasphemy, hatred’; (BN, L) *burho*, (B, G) *birago* ~ PEC \**gwātho* ‘gossip, rumor; offence, anger’; Avar *v<sup>w</sup>el* ‘gossip, rumor; abuse’, Tsakhur *qal* ‘offence, anger’ (K.1)
- [II.23] Basque \**elhi* ‘finger’ < \**helhi*; (BN, L) *erhi* ‘finger’, (B) *el-eztun* ‘finger-ring’ ~ PEC \**χġtHe* ‘sleeve’; Rutul *χili-di* ‘mitten’ (A.69)
- [II.24] Basque \**ethi* ‘cattle, herd’; (L) *elhi* ‘cattle, livestock, herd, flock’ < \**hethi* ~ PEC \**χwōlhV* / \**χHwōlV* ‘cattle, cows’; Chechen *hēlī* ‘cows’ (N.5)
- [II.25] Basque \**garhi* ‘thin, slender’; (BN, L) *garhi* ‘thin, slender, slim, slim-waisted’ ~ PEC \**ǵwarHV* ‘narrow, thin’; Agul *q:ure-* ‘thin, emaciated’ (R.59)
- [II.26] Basque \**gurhi* ‘butter, grease’; (Z) *gurhi* ‘fat, grease, butter’ ~ PEC \**χarHV* / \**χHərV* ‘butter, cheese’; Lezgi *vəri*, Rutul *χ<sup>h</sup>ara-d* ‘butter’ (P.7)
- [II.27] Basque \**ithu-n* ‘dark, night’; (BN) *ilhun* ‘night’ < PEuC \**ħilHu-* ~ (with metathesis) PNC \**ʔlġħ* ‘night’; Chamali *yeħa*, Tindi *reħa* id. (R.18)
- [II.28 = I.73, I.82] Basque \**lainho* ‘fog, mist, vapor’ ~ (with metathesis) PEC \**ħemħĀ* ‘dream’; Godoberi *hanlu* ‘fog, cloud’ (G.7)
- [II.29 = I.5] Basque \**manhai* ‘table’ ~ PEC \**mfiǝnē* ‘door; part of door’ (Q.9)

**Archaic Basque laryngeals, apparently lost in PNC:** There are four cases, all involving PNC/PEC tense sibilants (\**š*, \**š*, \**š*), in which the Basque cognate contains an internal laryngeal, but the PNC/PEC cognate does not. In NCED there is a hint that may explain this: “The [PEC] root [\**šwāyV*] ‘last year’ may have also contained a laryngeal,” based on the pharyngeal vowel /a/ in Archi *š<sup>w</sup>āš* ‘last year’ (NCED 975). The implication may be that Basque preserves an archaic laryngeal that was lost in PNC/PEC or absorbed into the tense sibilant, in the case of the ‘last year’ set and the other three sets.

- [II.30] Basque \**ihās* ‘last year’; (Z) *ihaz*, (B, G, AN) *igaz* [ivas] id. < PEuC \**iHašV* ~ (with metathesis) PEC \**šwāyV* ‘last year’ (< \**šwāyHV* or \**šHwāyV* ?); Chamali *sas*; Archi *š<sup>w</sup>āš* id. (H.3)
- [II.31] Basque \**sathui* ‘quick, nimble, flexible’, etc.; (L) *zalhui* id. ~ PNC \**šitV* ‘light (of weight)’ (< \**šitHV* or \**sHitV*?); Archi *s:ala* id. (R.42)

<sup>53</sup> The supposed derivation of this word from Latin *solum* (or a Romance descendant) is rejected here, for phonetic and semantic reasons; the corresponding loanword in Basque is (G) *zoru* ‘ground, floor’, (B, G) ‘sole (of shoe)’ (see BCR 238).

<sup>54</sup> One of the cases of PEuC \**gw* > Basque \**b*: see also the next set, and BCR 93–94, 167.

- [II.32] Basque *\*sethai* ‘plain, prairie’; (BN, L) *zelhai* ‘plain, flat land’, ‘flat’ (adj.) ~ PEC *\*ʒəlV* ‘plain, plateau’ (< *\*ʒəlHV* or *\*ʒHəlV* ?); Avar *c:or* ‘plain’ (D.5)
- [II.33] Basque *\*sutfio* ‘hole’; (BN, L) *zulho*, *zilho* ‘hole, burrow’, (B) *sama-zulo* ‘gullet’ (‘neck-tube’) ~ PEC *\*śwōtV* ‘hollow tube’ (< *\*śwōtHV* or *\*śHwōtV* ?); Avar *šulu* ‘pipe’, Lezgi *sulu-r* ‘throat’ (I.12)

### III. Basque /rc/ ~ /št/ stem alternation

As noted by Michelena (FHV 363–65), in a very few cases Basque exhibits a stem alternation between medial/final /rc/ and /št/. It so happens that PNC provides a key, since these Basque words correspond to NC words containing PNC clusters of *\*r* + sibilant affricates (*\*c*, *\*ć*, *\*č*):

- [III.1] Basque *\*borc* / *\*bošt* ‘five’, *\*bošte-ko* ‘hand’ ~ PEC *\*b[ə][r]cV* ‘paw; part of arm or leg’; Tsezi *besi* ‘fist’, Tabasaran *bac* ‘paw’ (M.6)<sup>55</sup>
- [III.2 = II.8] Basque *\*herce* / *\*hešte* ‘intestine’; (BN, L) *hertze*, (B, G, AN) *este* id. ~ PNC *\*HĀřčV* / *\*rHĀřčV* / *\*čĀřHV* ‘stomach, abomasum’; Avar *řorčo* ‘intestine’, Andi *ručo-l*, Chechen *čŭ-r-aš* ‘intestines’ (A.54; SCG 180)
- [III.3] Basque *\*fiorci* / *\*fiošti* ‘sky, storm’; (L) *hortz-adar*, *holtz-adar* ‘rainbow’,<sup>56</sup> (B) *osti* ‘storm, thunder’ ~ PEC *\*HĀřčwVn* ‘to become cloudy, gloomy’; Karata (Anchikh, Archo dial.) *orč-* ‘to become cloudy’, Avar *roč:n-* id. (G.12)
- [III.4 = I.20] Basque *\*mucu-ř* (< *\*murcu-ř*) / *\*muštu-ř* ‘snout; edge, extremity’ ~ PEC *\*mHĀřčwV* ‘pus; mucus, snot’ (A.19)
- [III.5] Basque *\*Perc* / *\*Pešt* ‘cauldron, pail’;<sup>57</sup> (G, AN) *pertz*, (AN, BN, Bzt, Z, R) *bertz*, (BN-ms) *best* ~ PEC *\*parčē* ‘water jug’; Avar *parč:i* ‘earthenware water jug’. (Q.23)

The distribution of /rc/-forms ~ /št/-forms is not regionally determined in Basque: for example, *bost* ‘five’ prevails both in the far west (B, G) and far east (Z, R), with *bortz*, *borz* in most of the more central dialects; *este* ‘intestine’ is mostly southwestern (B, G, AN), with (*h*)*er(t)ze* elsewhere; the same dialect may have both variants of *\*fiorci* / *\*fiošti*: (BN-Garazi) *ortzi* ‘storm cloud’ / *ostegun* ‘Thursday’; (Z) *ortze* ‘storm’ / *ostégün* ‘Thursday’, etc. (See the long lists of related words in EDB 316–19; Trask 1997: 278.)

Michelena sought to explain this alternation as stemming from a sandhi feature in compounds, in which a multistep change of *\*rtz* > *\*rzt* > *\*rst* > *\*st* /št/ occurred. Whatever the mechanism, it is likely that *\*berc*, *\*borc*, *\*herce*, *\*fiorci*, *\*mu(r)cu-ř* were originally the free-standing nouns, and *\*best-*, *\*bošt-*, *\*hešt-*, *\*fiošt-*, *\*muštu-*, respectively, were the stem variants used in compounds or before suffixes. Subsequently the two allomorphs were confounded and merged, so that the /rc/-forms could then be used in compounds, and /št/-forms in free-standing forms. It is interesting that the Basque words with this alternation all have NC cognates with the rhotic + sibilant clusters *\*rc*, *\*rć*, *\*rč*.<sup>58</sup>

<sup>55</sup> Problematic: outside of Basque the cluster *\*rc* is supported only by Avar *púrc:i* ‘ham’ (which may be secondary, < *\*bVc:-ri*: see NCED).

<sup>56</sup> The L words are compounds with *\*a=dar* ‘horn’, i.e. ‘sky-horn’; the form *holtz* (for *hortz*) by dissimilation from the following rhotic.

<sup>57</sup> *\*P* signifies “some labial stop,” since attested forms have both /p/ and /b/.

<sup>58</sup> Except Basque *\*berce* / *\*bešte* ‘other’, for which a satisfactory NC cognate has not been found. Possibly it is related in some way to Basque *\*bi* ‘two’ (cf. Udi *pə*, Lak (Kumukh dial.) *pk’i* ‘two’ < PNC *\*qHwā* id.), as in the common semantic link of ‘second’ and ‘other’.

In the following cases Basque has *\*št* or *\*st* corresponding to a PNC *\*r* + sibilant cluster, but without any known stem alternations:

- [III.6] PEuC *\*r[...]ś*: Basque *\*aštu-n* ‘heavy’; (B, G) *astun* ‘heavy, weighty; important, serious’ ~ PNC *\*=HrVŷśĒ* ‘thick, dense, fat’; Chechen *=arst-* ‘to fatten, become fat’, Tsezic *ŷašyu* ‘thick (of a flat object)’ (R.29)
- [III.7] PEuC *\*rš*: Basque *\*b=ušte-l* ‘rotten’; (L) *bustel*, (Z) *üstel* id. ~ PNC *\*=VršĒ* ‘to rot, ferment’; Dargi (Chiragh dial.) *=arš- / =urš-* ‘to rot’, Udi *baš’a(y)* ‘rotten’ (*b=aš’a-* ?) (R.45)<sup>59</sup>
- [III.8] PEuC *\*rč*: Basque *\*ga[ś]te* ‘young (person), youth’<sup>60</sup> ~ PNC *\*kwirčĒ / \*kwirčĒ* ‘young (of animals, birds)’; Lak *karč* ‘puppy’, Khwarshi *k’uč’i* id. (R.66)

## Conclusions

It is affirmed here, based on ample evidence, that a phoneme *\*m* did exist in Proto-Basque, as well as in Proto-Euskaro-Caucasian, and that it remained quite stable in Basque in most initial and medial environments. But PEuC *\*m* did indeed disappear, or change to /n/, especially in reflexes of internal PEuC clusters. Thus Trask’s doctrine that “no *\*m* can be reconstructed for Pre-Basque” (based on a more cautious suggestion by Michelena) turns out to contain a small smidgen of truth.

There is also copious evidence that the Proto-Basque laryngeals *\*h* and *\*h̥* are preserved intact in northern Basque reflexes of the same words in which a laryngeal is postulated for Proto-North Caucasian or Proto-East Caucasian (according to NCED). In internal clusters northern Basque frequently preserves these laryngeals in cases where there is only circumstantial evidence of them in modern North Caucasian languages (e.g., as pharyngealized consonants and vowels, or as particular accentual patterns). This is one of several ways in which Basque is more conservative than its cousins, the North Caucasian languages.

At the same time, Basque naturally exhibits some important innovations. Phonetically, a much simpler vowel and consonant system developed in Basque than Proto-Euskaro-Caucasian is thought to have had (BCR chapter IV). Grammatically, the Euskaro-Caucasian noun class system loses its productivity, but leaves traces in fossilized prefixes such as *\*be/\*bi=*, *\*e/\*i=*, *\*o/\*u=*, etc. (BCR chapter V). Lexically there is, for example, the development of a new word for ‘eye’, *\*b=egi* (originally ‘seer’, cf. PEC *\*=agwV* ‘to see’), while North Caucasian (*\*ʔwīʔi* ‘eye’; Dargi *ule*, Abkhaz *á-la*), Burushaski (*\*-l-či / \*il-* ‘eye’) and Yeniseian (*\*de-s* ‘eye’)<sup>61</sup> retain the older form, which Basque preserves only in the truncated form *\*le-* in words for ‘eye-tooth’ (*\*le-t-agin*), and in the verb *\*uter-tu* ‘to understand, comprehend; foresee, suspect; feel, sense’ (see BCR, etymologies A.8, A.11, A.12, V.14).<sup>62</sup> Other Basque innovations are evident, for example, in kinship terminology and numeral words that are radically different from those of North Caucasian; here we see the results of innovations on both sides (see BCR 329–336; 344–48).

Greenberg (1995: 207) refers to an essential attribute of a “best explanation” hypothesis of linguistic relatedness. “Such hypotheses have a further important characteristic, which we may call fruitfulness, That is, they can be built upon and lead to further discoveries and explanations of hitherto puzzling phenomena.” Seven decades earlier Edward Sapir had written in

<sup>59</sup> Udi *possibly* has the same fossilized class prefix (*\*b=*) as postulated in the Basque form.

<sup>60</sup> The reconstruction *\*ga[ś]te* is provisional, since dialectal evidence /*gaste/* ~ /*gašte/* is ambiguous.

<sup>61</sup> Proto-Yeniseian lacks initial laterals, and the initial resonants *\*l-*, *\*l̥-* and *\*n-* regularly become *\*d-* (SCP 38, 68).

<sup>62</sup> George Starostin proposed the *\*b=egi* and *\*uter-tu* etymologies. The Basque verb is typologically similar to Greek ὑπόψομαι ‘I shall suspect’, with -οψ- < PIE *\*okw-s-* ‘eye’ (Buck 17.44).



a letter that the Sino-Dene hypothesis “is all so powerfully cumulative and integrated that when you tumble to one point a lot of others fall into line” (letter to Alfred Kroeber in 1921, quoted in Golla 1984: 374). So it is with the Euskaro-Caucasian hypothesis. Just within the small selection of etymologies discussed here, several of them figure in more than one of the three major topics discussed: [I] the labial nasal *\*m*; [II] archaic Basque laryngeals; [III] the /rc/ ~ /št/ alternation, for example:

- [I.2] = [II.14] Basque *\*mahać* ‘grape(s)’ ~ PNC *\*šämćō* ‘apple; medlar’;  
 [I.8] = [II.15] Basque *\*mařuhi* / *\*mahuri* ‘strawberry’ ~ PEC *\*mer(?)V* ‘a kind of berry’;  
 [I.5] = [II.29] Basque *\*manhai* ‘table’ ~ PEC *\*mfhǎžnē* ‘door; part of door’;  
 [I.20] = [III.4] Basque *\*mucu-ř* / *\*muřtu-ř* ‘snout; edge, extremity’ ~ PEC *\*mHǎřčwV* ‘pus; mucus, snot’;  
 [I.24] = [II.5] Basque *\*čehume* ‘half-span’ ~ PNC *\*čřmħV* ‘span’;  
 [I.33] = [II.18] Basque *\*sumhe* ‘a kind of tree’ ~ PEC *\*žřmV* ‘bush, grass; a kind of fruit’;  
 [I.73] = [I.82] = [II.28] Basque *\*lainho* ‘fog, mist, vapor’ ~ PEC *\*hemĶĶ* ‘dream’; Godoberi *hanlu* ‘fog, cloud’;  
 [II.8] = [III.2] Basque *\*fierce* / *\*hešte* ‘intestine’ ~ PNC *\*HǎřčV* / *\*rHǎřčV* / *\*čřrHV* ‘stomach, abomasum’; Avar *řorćo* ‘intestine’.

These and other cases function as a kind of cross-validation (the same etymology confirmed by more than one rule), and exemplify what Sapir meant by “when you tumble to one point a lot of others fall into line.”

### General notes to etymologies

1. The fossilized plural or collective suffix *\*-ř* occurs in many Basque nouns (see BCR 76-78). In this paper see etymologies [I.4] Basque *\*malka-ř* ‘straw’; [I.14] Basque *\*moku-ř* ‘tree trunk, tree base’; [I.16] Basque *\*moś-ko-ř* ‘tree trunk’; [I.20 = III.4] Basque *\*mucu-ř* / *\*muřtu-ř* ‘snout, muzzle’; [I.28] Basque *\*hama-ř* ‘ten’; [I.31] Basque *\*sama-ř* / *\*čama-ř* ‘fleece, shorn wool; mane’; [I.49 = I.78] Basque *\*fonda-ř* ‘sand, residue, bottom’; [I.54 = I.81] Basque *\*konko-ř* / *\*kunku-ř* ‘hump, joint, bone’; [I.85] Basque *\*lance-ř* ‘drizzle, fog’; [II.8 = III.2] Basque (BN-arc) *halsarr-ak* ‘entrañas’ < *\*fierce-ř*; [II.10] Basque *\*itha-ř* ‘pea(s), bean(s)’; [II.20] Basque *\*bethia-ř* ‘forehead, side, edge’; [II.21] Basque *\*butha-ř* ‘breast, chest’.

2. Basque correspondences to PNC lateral affricates *\*Ķ*, *\*Ķ* (= *\*Ķ*), *\*Ķ* in initial (*Anlaut*) position. Basque laterals are very simple: only the resonant /l/ appears in initial position: see [I.27] Basque *\*(eci-)[lamu]* ‘two days after tomorrow’ ~ PEC *\*ĶřmV* ‘one day (two days) after tomorrow’; [I.30] Basque *\*limu-ri* ‘slippery, moist, smooth’, etc. ~ PEC *\*ĶHwemV* ‘liquid’ (adj.); [I.47] Basque *\*listo-* ‘hornet, wasp’ ~ PEC *\*ĶšmčV* ‘a biting insect’; [I.73] Basque *\*lainho* ‘fog, mist, vapor’ ~ PEC *\*hemĶĶ* ‘dream’; [II.11] Basque *\*lahai* ‘thorn, bramble’ ~ PEC *\*ĶwřřV* ‘leaf’; [II.13] Basque *\*luHuř* ‘earth, ground, land’ ~ PNC *\*lhěmĶwř* ‘earth’. (In cases I.73, II.11 and II.13 there has been metathesis. See BCR 154-55 for further examples and explanations.)

3. Basque correspondences to PNC lateral affricates *\*Ķ*, *\*Ķ* (= *\*Ķ*), *\*Ķ* in medial (*Inlaut*) position. There are two main reflex types, (a) Basque /rd/, and (b) Basque /rt/. See BCR 156–58 for further examples and a possible explanation of reflex (a) *vs.* (b). In this paper, with reflex type (a), see etymologies [I.7] Basque *\*mardo* ‘robust; luxuriant’, *\*mardu-l* ‘healthy, strong’ ~ PEC *\*mōrEV* ‘male; man, hero’; [I.9] Basque *\*meda-ř* ‘narrow, tight’ (< *\*merda-ř*) ~ PEC *\*fmVřřV* ‘thin, narrow’; [I.15] Basque *\*mardo* ‘bunch, group’ ~ PEC *\*mǎř[Ķ]o* ‘handful, armful’; With reflex type (b) see [I.21] Basque *\*muti-l* ‘boy; apprentice; worker’ (< *\*murti-l*) ~ PNC *\*mĶřřV* ‘husband, male, worker’. The Basque reflexes are the same whether the proto-phoneme was simple (e.g., *\*-Ķ-*) or in a cluster with a resonant (e.g., *\*-rĶ-*, *\*-rĶ-*, *\*-lĶ-*).

4. Matches between Basque and NC vowels may reflect a hypothetical Euskaro-Caucasian apophony (*Ablaut*) system (see BCR 106–110). The clearest evidence for such a system are found in some lexicalized traces in Basque, for example the Basque adjective *\*ośo* ‘whole, complete’ (R.65) and the verb *\*aśe* ‘to be filled, satiated’ (V.66), parallel with PNC *\*=fiđčV* / *\*=fiǎčV* ‘full, to fill’ (NCED 525), as in Proto-Nakh *\*=uč-* ‘to be filled, satiated’ / *\*=ač-i(n)*

‘heavy’ (Chechen =üz-na ‘full’ / =eza ‘heavy’), corresponding to Bsq \*ošo / \*aše, respectively. As regards PNC nouns, this system was based on an opposition of mid vowels for the direct case *vs.* high vowels for oblique cases (see NCED 81-82), exemplified by the noun \*nëwqũ [direct] ‘tear(s), pus’ *vs.* \*nřwqř- [oblique] (NCED 848); in Basque this alternation was lexicalized as regional variants, southwestern Basque *negar* ‘weeping; tear(s)’ *vs.* northeastern Basque *nigar* id. (BCR 106–07; A.78). A hypothesis needing further testing suggests that some vowel matches, such as \*i ~ \*e, in [I.30] Basque \*límu-ri ‘slippery, moist, smooth’ ~ PEC \*λHwemV ‘liquid’, and [I.48] Basque \*sisa-li ‘worm’, \*sisa-ni ‘leech’ ~ PEC \*sēmsā ‘tadpole, water flea, bug, ant’, could have resulted from the lexicalization of different apophonic allomorphs.

5. It is postulated in BCR that certain initial syllables in Basque derive from fossilized remnants (“stage III articles”) corresponding to the class/gender markers in Proto-North Caucasian that persist in most of the modern NC languages (BCR 11–12; 58–71). It happens that within the collection of etymologies in the current paper there are a few examples of the fossilized articles \*e=, \*i= and \*o=: see [I.47] Basque \*e=lco ‘gnat, midge, mosquito’ ~ PEC \*λāmc̣V ‘a biting insect’; [I.52 = I.80] Basque \*i=hunc ‘dew’ ~ PEC \*xwiṃc̣wī ‘bog, dirt’; [I.53] Basque \*i=soc ‘frost, ice’ ~ PEC \*ẓwVṇc̣(w)V ‘snowdrift, snowstorm’; [I.63] Basque \*e=roic ‘cliff, fissure’ ~ PEC \*hṛṃc̣we ‘stone’; [II.6] Basque \*e=the ‘word, speech’ < \*e=tehe ~ PEC \*ḷēHwV / \*ṭēHwV ‘word’; [II.7] Basque \*i=ketV ‘fattened ox’ / \*o=ketV ‘meat; piece (of meat, cheese), morsel’ ~ PEC \*q̣hwēřV̄ / \*q̣wēřhV̄ ‘large female domestic animal (cow, mare)’; [II.10] Basque \*itha-ř ‘pea(s), bean(s)’; (Z) *ilhar* ‘beans’ < \*i=hVta-ř ~ PEC \*hōwt[ā] ‘bean(s), lentil’.

6. **Reconstructions:** Reconstructions based on attested forms are designated by [\*], in the usual manner. More speculative or less substantiated “quasi-reconstructions” are designated by [+]. For example, PEuC \*HaçimV ‘tooth’ is a possible, but not necessarily the *only* solution to reconciling Basque \*hagin ‘tooth’ and PNC \*HēmçĀ ‘to bite’, by a common type of transposition (see I.38); likewise PEuC \*e=tehe is a possible, but not necessarily the only conceivable intermediate form between Basque \*ethe ‘word, speech’ and PEC \*ḷēHwV / \*ṭēHwV ‘word’ (see II.6).

### Abbreviations of languages and dialects

A	Araban: old Basque dialect of Araba (Álava) province (= Michelena’s <i>meridional</i> )
AN	Alto Navarro = High Navarrese (Basque dialect)
arc	Archaic or obsolete form
Azk	Aezkoan (Basque dialect)
B	Bizkaian = Biscayan (Basque dialect)
Bzt	Baztanese (Basque dialect)
BN	Bas-navarrais = Low Navarrese (Basque dialect)
c	Common Basque: the form in all dialects / or the most widespread form
dial.	dialectal
EB	<i>euskara batua</i> (standard Basque)
G	Gipuzkoan (Basque dialect)
L	Lapurcian = Labourdin (Basque dialect)
PEC	Proto-East Caucasian
PEuC	Proto-Euskaro-Caucasian
PNC	Proto-North Caucasian
PWC	Proto-West Caucasian
R	Roncalese (Basque dialect)
Sal	Salazarese (Basque dialect)
Z	Zuberoan = Souletin (Basque dialect)

### Abbreviations of sources

BCR	Bengtson (2017)	FHV	Michelena (1990)
Buck	Buck (1949)	NCED	Nikolaev, Starostin (1994)
CDIAL	Turner (1962–66)	OEH	Michelena (1987–)
EDB	Trask (2008)	SCG	Starostin (2005b)

SCP Starostin (2005a)

TOB Starostin (1998–2005)

WALS Dryer, Haspelmath (2013)

## References

- Aulestia, Gorka, Linda White. 1992. *Basque-English English-Basque dictionary*. Reno, Nevada: University of Nevada Press.
- Azkue, Resurrección María de. 1905–1906. *Diccionario vasco-español-francés / Dictionnaire basque-espagnol-français*. Bilbao / Paris: The Author / Paul Geuthner.
- Bancel P.J., A. Matthey de l'Etang, J.D. Bengtson. 2015. Back to Proto-sapiens (Part 2): The global kinship terms PAPA, MAMA, and KAKA. In D. Jones & B. Milicic (eds.). *Kinship, language, and prehistory: Per Hage and the renaissance in kinship studies*: 38–45. Salt Lake City: University of Utah Press.
- Bancel P.J., A. Matthey de l'Etang. 2013. Brave new words. In: C. Lefebvre, B. Comrie, H. Cohen (eds.). *New perspectives on the origins of language*: 333–377. Amsterdam: John Benjamins.
- Bengtson, John D. 2017. *Basque and its closest relatives: a new paradigm*. Cambridge, Mass.: Mother Tongue Press / Association for the Study of Language in Prehistory.
- Bengtson, John D. & George Starostin. 2015. The Dene-Sino-Caucasian hypothesis: state of the art and perspectives. Discussion draft posted on Academia.edu, 2015.
- Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University of Chicago Press.
- Dryer, M. S., M. Haspelmath (eds.) 2013. The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info/> [accessed 08.15.2018].
- Golla, Victor, ed. 1984. *The Sapir-Kroeber correspondence: letters between Edward Sapir and A.L. Kroeber 1905–1925*. Berkeley: University of California Press.
- Greenberg, Joseph H. 1995. The concept of proof in genetic linguistics. *Mother Tongue* 1: 207–216.
- Jacobsen, William H. Jr. 1995. Comment on R.L. Trask's 'Basque and Dene-Caucasian: a critique from the Basque side'. *Mother Tongue* 1: 120–142.
- Meyer-Lübke, Wilhelm. 1911. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.
- Michelena, Luis. 1957. Las antiguas consonantes vascas. In: D. Catalán (ed.). *Miscelánea homenaje a André Martinet*, vol. I: 113–157. Tenerife: Universidad de La Laguna.
- Michelena, Luis. 1961. *Fonética histórica vasca*. San Sebastián: Diputación Provincial de Guipúzcoa.
- Michelena, Luis [Koldo Mitxelena]. 1987–. *Orotariko euskal hiztegia / Diccionario general vasco*. Bilbao: Euskaltzaindia. [Edited since Michelena's death by Ibon Sarasola.] Available: [http://www.euskaltzaindia.eus/index.php?option=com\\_content&view=article&id=276&Itemid=413&lang=eu/](http://www.euskaltzaindia.eus/index.php?option=com_content&view=article&id=276&Itemid=413&lang=eu/) [Accessed 08.15.2018].
- Michelena, Luis [Koldo Mitxelena]. 1990. *Fonética histórica vasca*. (Anejos del *Anuario del Seminario de Filología Vasca «Julio de Urquijo»*, IV Serie «Obras completas de Luis Michelena», 1.) San Sebastián: Diputación Foral de Guipúzcoa. [Reprint of 2<sup>nd</sup> edition.]
- Nikolaev, Sergei L., Sergei A. Starostin. 1994. *A North Caucasian etymological dictionary*. Moscow: Asterisk Press.
- Ruhlen, Merritt. 1975. *A guide to the languages of the world*. Stanford: Language Universals Project.
- Starostin, Sergei A. 1996. Response to Trask [1996]. *Mother Tongue* 2: 119–123.
- Starostin, Sergei A. (ed.). 1998–2005. *The Tower of Babel. An etymological database project*. Available at: <http://starling.rinet.ru/> [accessed 08.15.2018].
- Starostin, Sergei A. 2005a. *Sino-Caucasian [phonology]*. <http://starling.rinet.ru/Texts/scc.pdf>
- Starostin, Sergei A. 2005b. *Sino-Caucasian [comparative glossary]*. <http://starling.rinet.ru/Texts/glossary.pdf>
- Steinbergs, Aleksandra. 1996. Intimate body parts and related terms in proto-Baltic. In: H. I. Aronson (ed). *Linguistic studies in the non-Slavic languages of the commonwealth of independent states and the Baltic republics*: 337–357. Chicago: Chicago Linguistic Society.
- Trask, R. L. 1995. Basque and Dene-Caucasian: a critique from the Basque side. *Mother Tongue* 1: 3–82.
- Trask, R. L. 1996. Second response to Starostin [1996]. *Mother Tongue* 2: 125–129.
- Trask, R. L. 1997. *The history of Basque*. London/New York: Routledge.
- Trask, R. L. 2008. *Etymological dictionary of Basque*. Ed. by Max W. Wheeler. University of Sussex. Linguist List Publications: <http://www.sussex.ac.uk/linguistics/1-4-1-2.html>
- Turner, R. L. 1962–66. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

*Джон Бенгтсон. Заметки об эускара-кавказской фонологии*

Статья посвящена обоснованию ряда особенностей фонологической системы, реконструируемой для праэускара-кавказского языка (предполагаемого общего предка баскского и северокавказских языков) и опубликованной в недавней монографии автора. Детальному анализу подвергаются следующие моменты: (I) развитие праэускара-кавказского сонанта *\*m* в баскском; (II) отражения эускара-кавказских ларингалов в начальной позиции в баскском; (III) чередование /rc/ ~ /št/ в алломорфах некоторых баскских именных основ. В статье показано, как сценарий, предлагаемый для праэускаро-кавказского, позволяет объяснить ряд важных нерешенных проблем в исторической фонологии и морфонологии как баскского языка, так и северокавказской семьи.

*Ключевые слова:* баскский язык, северокавказские языки, сравнительное языкознание, эускара-кавказская языковая семья, губные носовые, ларингалы

## Suspended affixation with Tocharian adjectival suffix A *-ši* / B *-ṣṣe* and its possible parallel in Old Uighur

In this paper, I discuss a hitherto undescribed property of the Tocharian adjectival suffixes (most commonly A *-ši* / B *-ṣṣe*, more rarely also *-tstse* and *-ññe*) to be shared by conjuncts. I will attempt to prove that such constructions are based on coordinated nouns and not on *dvandva*-type compounds, and compare Tocharian data with the situation in the areally close Old Uighur. If my assumption is correct, suspended affixation is possible not only with inflectional morphemes of secondary cases, but also with derivational ones, although the latter phenomenon is less widely spread in Tocharian.

Keywords: group derivation, suspended affixation, Tocharian adjectival suffixes, Old Uighur language, areal contacts

### Introduction

The Tocharian adjectival suffix A *-ši* / B *-ṣṣe* is well known for its agglutinative properties. As pointed out in previous works, it is usually added not to the stem, but to the oblique (= accusative)<sup>1</sup> form of a word, which can be Obl.Sg. as well as Obl.Pl. or Obl.Du. (Sieg, Siegling, Schulze 1931: 23; Krause, Thomas 1960: 144; Pinault 1992: 101; Zimmer 1982/83: 278; Adams 2009: 302). Cf. B *kleśaṣṣe* and *kleśanmaṣṣe* from *kleś* Nom.-Acc.Sg., *kleśanma* Nom.-Acc.Pl. “affliction”; *asaṃkhyaiṣṣe* and *asaṃkhyaintaṣṣe* from Nom.-Acc.Sg. *asaṃkhyai*, Nom.-Acc.Pl. *asaṃkhyainta* “incalculably long period of time”; *ekaṣṣe* and *eśneṣṣe* from Nom.-Acc.Sg. *ek*, Nom.-Acc.Dual. *eśane* “eye”. In this regard A *-ši* / B *-ṣṣe* behaves similarly to secondary case suffixes which also require Obl.

Another remarkable property of A *-ši* / B *-ṣṣe* is the adjectivization of modified nouns (Adams 2009: 304–305), cf.:

*larenämpa*      *śinmalñe-ṣṣe*    *palskalñe* (B 511 a3)  
dear:Pl.m:Comit    coming-Adj    thought<sup>2</sup>  
“The idea of coming together with the dear ones”.

*wrocy*      *akālkāntse*    *knelñe-ṣṣe*      *tsārwo* (B 591 a6)  
big:Acc.Sg.m    wish:Gen.Sg    fulfillment-Adj    joy  
“The joy of the fulfillment of a big wish”.

Already in the edition of Tocharian A texts (Sieg, Siegling 1921) A *-ši* is consequently transliterated separately. According to Zimmer (1982/83: 277), the first editors of Tocharian manuscripts would treat this formant as an enclitical word expressing “grammatical relations of manyfold kind”.

<sup>1</sup> The Tocharian accusative is traditionally called oblique because it serves as the basis for secondary case forms (Pinault 2008: 462; Krause, Thomas 1960: 78).

<sup>2</sup> Here and below only glosses for marked grammatical meanings are indicated. The coinciding Nom-Acc. is unmarked and thus not glossed, e.g. Nom-Acc.Pl.f is glossed as Pl.f.

Indeed, this suffix does not provide any special meaning, and, as noticed in Hajnal 2004: 139, the use of A *-ši* / B *-šše* adjectives generally corresponds to the use of genitive in most other Indo-European languages.

Along with the peculiarities mentioned above, the suffix A *-ši*, B *-šše* has one more agglutinative feature which has not yet been specially described: namely, it can adjectivize coordinated nouns. In such constructions – rather marginal in Tocharian A and quite well presented in Tocharian B – the adjectival suffix behaves similarly to secondary case affixes in that it links only the last word in a chain of conjuncts. Cf. the following Tocharian B examples of adjectivization with *-šše* and examples of suspended affixation<sup>3</sup> with the perlative suffix *-sa* for one and the same pair of words:

- (1) *kest yokai-šše läkle* (B 284 a2)  
 hunger thirst:Acc.Sg-Adj suffering  
 “The suffering of hunger and thirst”  
 vs. *kest yokai-sa* (B 286 b3)  
 hunger thirst:Acc.Sg-Perl  
 “Because of hunger and thirst”
- (2) *sklok pr(o)sk(ai-šše or)k(a)mñe* (PK AS 17K a1)  
 doubt fear:Acc.Sg-Adj darkness  
 “The darkness of doubt and fear”.  
 vs. *sklok proskai-sa* (B 409 a1)  
 doubt fear:Acc.Sg-Perl  
 “Because of doubt and fear”

The problem with the treatment of such constructions is the ambiguity whether adjectivized words comprise a syntagm or a dvandva compound. Namely, hendiadyses are highly frequent in Tocharian literary texts, and if the Nom. and Acc. forms of non-final elements in such expressions coincide, i.e. coordination of non-final elements with the oblique of the final element is not explicitly marked, it is impossible to define their syntactic status – as it is for the words *kest* and *sklok* in the examples cited above.

However, along with the cases where dvandvas and conjuncts are indistinguishable, at least Tocharian B has examples of adjectivization with *-šše* which can be treated as cases of suspended affixation due to some morphosyntactic characteristics. These cases will be treated further in detail (Section 2.1), as well as the few found examples from Tocharian A (Section 2.2). I will also try to find out whether this derivational model is possible with two other very productive Tocharian B suffixes *-tstse* and *-ññe* (Section 3). After that I will attempt to find typological and areal parallels to such constructions in some other languages, first of all in Old Uighur (Section 4). All sources for used translations will be indicated unless they are specifically proposed by the author.

### 1. A *-ši* / B *-šše* from the diachronic point of view

The exact origin of A *-ši* / B *-šše* is unknown. Traditionally two reconstructions have been proposed: *\*-syo* ~ *\*siyo* (Ringe 1996: 117; Hajnal 2004: 140) or *\*-skyo-* (Van Windekens 1979: 133). However, both PIE morphemes would formally coincide in Tocharian after palatalization.

<sup>3</sup> The term 'suspended affixation' has been introduced in Lewis 1967: 35 for cases “when one grammatical ending serves two or more parallel words” and thus describes group inflection in coordination.

As already noticed in the introduction, synchronically this suffix seems to join the Obl. form of the word which replicates the use of Tocharian secondary case suffixes. However, the statement about A *-ši* / B *-ṣṣe* being added to the oblique (=Acc.) case is hardly applicable to Tocharian A where Nom.Sg. and Acc.Sg forms are usually indistinguishable, and not completely true for Tocharian B where, although rarely, *-ṣṣe* can be found added to the thematic vowel of the oblique stem, cf. *witsakaṣṣe* (in B 530 a2) < Nom.Sg. *witsako*, Acc.Sg. *witsakai*, Acc.Pl *witsakaṃ* “root” (i.e. not *\*witsakaiṣṣe* or *\*witsakaṃṣe*). Thus, it would be probably more appropriate to say that A *-ši* / B *-ṣṣe* is added to the oblique stem, which mostly coincides with the oblique case.

As shown in Hajnal 2004: 141–143, this must have been the exact situation in diachronic perspective: the suffix A *-ši* / B *-ṣṣe* was originally added to the stem and not to the oblique case form, but a formal identity of some stem types (PIE *u-*, *o-* and occlusive stems) and Obl.Sg. in the common Tocharian period resulted in a reinterpretation of the whole derivational model of A *-ši* / B *-ṣṣe*.

At any rate, cases such as *witsakaṣṣe* are relatively isolated, while the enumerated agglutinative features can be observed throughout the whole relative chronology<sup>4</sup> of Tocharian B manuscripts (there has been no special dating of linguistic stages of Tocharian A texts for now).

## 2. Adjectivization of conjuncts by means of A *-ši* / B *-ṣṣe*

### 2.1. Situation in Tocharian B

Word sequences that underlie adjectives ending in B *-ṣṣe* can be regarded as coordinated nouns due to the following reasons.

1. While the form to which the suffix *-ṣṣe* is directly added may be a stem without inflexion (cf. *witsakaṣṣe* mentioned above), the non-final forms which *-ṣṣe* also refers to represent a morphological word. This word is always an Obl. if distinguishable from Nom.:

- (3) *kektseñ reki pälsko-ṣṣe āstreṃ (warṣälñe)* PK AS 7L a2  
 body:Acc.Sg word mind-Adj pure:Acc.Sg.m exercise  
 “(Basis for) a pure exercise of body, word, [and] mind” (CEToM).
- (4) *atiyai pisäl melte ora-ṣṣe puwar* (B 194 b1)  
 grass:Acc.Sg chaff dung wood-Adj fire  
 “A fire of grass, chaff, dung and wood” (Adams 2013: 9).
- (5) *pätär mätär säsuwer-ṣṣe lareṣṣana ñemna* (B 266 a2–3)  
 father:Acc.Sg mother:Acc.Sg son\*-Adj dear:Pl.f name:Pl  
 “The dear names of father, mother, children”<sup>5</sup> (Adams 2013: 595).

(The adjective *säsuwerṣṣe* “pertaining to a son” is based on the independently unattested form *säsuwer*. The latter corresponds morphologically to the verbal noun in *-r* (whose Nom.Sg. and Acc.Sg coincide) from an unattested PpT *\*säsu* “born (?)”, cf. Winter 1985: 260–261).

<sup>4</sup> See Peyrot 2008.

<sup>5</sup> Remarkably, the suffix *-ṣṣe* in *säsuwerṣṣe* is not coordinated with Nom.Pl.f *ñemna* “names” (i.e. Nom.-Acc.Sg. *säsuwerṣṣe* and not Nom.-Acc.Pl.f *\*säsuwerṣṣana*, cf. the following adjective *lareṣṣana* with the same suffix). The probable explanation is that the three attributives *pätär(ṣṣe)*, *mätär(ṣṣe)* and *säsuwerṣṣe* refer logically not to one and the same multitude of “nouns” expressed with the plural *ñemna*, but each corresponds to just one of the components of this multitude (“father’s name, mother’s name and son’s name”).

2. Adjectivized chains of words can demonstrate either single use of the suffix *-ṣṣe* that is joined only to the final element, or its repetition after each word. Therein as well, the suffix *-ṣṣe* behaves similarly to the secondary case affixes, for which suspended affixation is optional. Cf. the example (3) mentioned earlier, B 7 a8 and B 386 b3:

- (6) *kektseñ reki-ṣṣana krenta : ///* (B 7 a8)  
body:Acc.Sg word-Adj:Pl.f good:Acc.Pl.f  
“The good (deeds) in (mind), body and words” (CEToM).
- (7) *kektseñä-ṣṣe (re)k(i)-ṣṣe se śīl westrä • ///* (B 386 b3)  
body:Acc.Sg-Adj word-Adj this moral behavior speak:Mid.3s.Prs  
“This moral behavior of body and word is spoken of” (Adams 2013: 634).

Another case of variative repetition of *-ṣṣe* is found in the following examples with the enumeration of the beings of different rebirths (Nom-Acc.Sg. *nrai* “hell (beings)”, Nom-Acc.Pl. *lwāsa* “animals” and Acc.Pl. *preteṃ* “preta ghosts”) quite frequent in Tocharian and other Buddhist texts from East Turkestan:

- (8) *ñākcīyana śāmñāna rūpanma sātkasamai*  
divine:Pl.f human:Pl.f form:Pl pervade:Mid.1s.Pt  
*taisa nrai lwāsa prete-ṣṣana •* (IOL 4 b5)  
thus hell animal:Pl ghost-Adj:Pl.f  
“I have pervaded divine and human forms ..., thus [also] hell, animal and preta [forms]” (CEToM).
- (9) */// po wešeññai ñākcyai śāmñai nrai lwasā-ṣṣai*  
all voice:Acc.Sg divine:Acc.Sg.f human:Acc.Sg.f hell animal:Pl-Adj:Acc.Sg.f  
*preteṃ-ṣṣai :* (PK AS 13C b7)  
ghost:Acc.Pl-Adj:Sg.f  
“Every voice, divine [or] human, [or of] animals, [or of] hell [or of] pretas” (CEToM).
- (10) *nrai(-ṣṣe lwā-ṣṣe) preten-ṣe tnek nai kṣanti kālale* (B 554 b5-b6)  
hell-Adj animal-Adj ghost:Acc.Pl-Adj here Emp forgiveness achieve:Ger  
“For hellish [being], animal or preta, only here [is] forgiveness achievable” (Adams 2013: 295). It should be however noticed here that the suffix is a conjecture for two first adjectives in this example.

A parallel to these adjectives is provided by the examples with variative repetition of the secondary case Loc. B *-ne*:

- (11) *nrai lwāsa śle prete(n-ne nekä)t lakle*  
hell animal:Pl and ghost:Acc.Pl-Loc destroy:2s.Sbj suffering  
*eṃṣketstse : 2 ||* (PK AS 17A b4–5)  
completely  
“You will completely destroy sorrow among the hell-[beings], the animals and the pretas” (CEToM).
- (12) *nrai-ne läklenta preten-ne śle lwāsā-ne :* (B 284 b3)  
hell-Loc suffering:Pl ghost:Acc.Pl-Loc and animal:Pl-Loc  
“The sufferings among hellish (beings), pretas and animals”.

3. The adjectivized chains of conjuncts can vary in length, e.g. the following cases with the enumeration of the concepts *kleś yāmor vastu* (Skt. *kleśa karma vastu*) “affliction, deed, basis”:



(13) *kleś yāmor vastu-ṣṣe* /// (B 174 a5)  
affliction deed basis-Adj

(14) *kleś yāmor-ṣ(ṣ)e ṅare* (B 286 a5-a6)  
affliction deed-Adj thread  
“Thread of affliction and deed”.

(15) *yāmor-ṣṣe vastu-ṣṣe menā(k-sa)* (B 174 b1)  
deed-Adj basis-Adj example-Perl  
“Through the example of deed and matter”.

Also, with the already mentioned expression *kektseñe reki pälsko*\*<sup>6</sup> “body, word (and) mind”:

*kektseñ reki pälskoṣṣe* (PK AS 7L a2) and *kektseñ rekiṣṣana* (B 7 a8)  
(from the examples 3 and 6)

(16) *kāyacittāvasthāviṣṣata • kektseñ pälsko-ṣṣai avastha-ṣṣai wāki-meṃ* (B 197 b3)  
body:Acc.Sg mind-Adj:Acc.Sg.f state-Adj:Acc.Sg.f difference:Abl  
“Skt. From the difference of the condition of body and mind”.

4. As it was already shown in some examples above, the suffix -*ṣṣe* can be shared by three (examples 3, 5, 8, 13) or more (four in example 4) nouns. In this regard, one of the most interesting examples is PK AS 2A a3–5 with a Sanskrit parallel in the preceding line of the same manuscript:

(17) ||(*e*)*raṅḍabilvā bṛhatidvayaṃṅica mātuluṅga : pāṣāṅabhi tṛkaṭu mūlakṛtakaṣāya :*  
*hirant • pīlamāti • wi praha(ti)nta • mā(tu)l(uṅk pāṣā)ṅa(bhit) • klyotaiṣṣana witsaka-ṣṣe*  
*kaṣāy* (PK AS 2A a3–5)

“A decoction consisting of ricinus, of the heart of the fruit of Aegle marmelos, of two Solanum indicum, of citron, of (Coleus) aromaticus [and] roots of Tribulus terrestris” (CEToM).

The enumeration of six names of medical ingredients gets one common suffix -*ṣṣe*, although the latter has no morphological correspondence in the Sanskrit version (functionally it renders the PPt *kṛta*). In this example, not only the length of the chain of conjuncts is remarkable, but also the fact that two of these conjuncts (*wi praha(ti)nta* “two Solanum indicum” and *klyotaiṣṣana witsaka-ṃ* “roots of Tribulus terrestris”) are modified noun phrases, with a cardinal number (*wi* “two”) and an adjective (also in -*ṣṣe*) as modifiers respectively.

This Tocharian B example is unique, but, in my opinion, it shows especially convincingly that constructions with the derivational suffix -*ṣṣe* should be treated as syntactic ones.

5. Although -*ṣṣe* is usually added to nouns, some rare examples can be found in which it is shared by two homogenous gerundives<sup>7</sup> (examples 18–19). Thus, the discussed affixation model seems to be quite productive in Tocharian and cannot be reduced to derivation from lexicalized expressions.

(18) *anāṣṣälle satāṣṣälle-ṣṣe ime*<sup>8</sup>  
inhale:Ger exhale:Ger-Adj idea  
“The idea of how to inhale and to exhale”.

<sup>6</sup> The expression *kektseñe reki pälsko*\* is attested only in oblique forms *kektseñ reki pälsko-*.

<sup>7</sup> Tocharian verbal adjectives in -*lle* with the meaning of necessity or possibility of some action.

<sup>8</sup> BHS *prāṅāpānasmṛti* (Thomas, 1972: 443, footnote 5).

- (19) /// *lle päkšalle-šše nräi-ššeṃts lakle* : /// (B 150 a6)  
 X:Ger cook:Ger-Adj hell-Adj:Gen.Pl. suffering  
 “The (roasting?) and cooking suffering of the hellish (beings)”.

### 2.1.1. Some ambiguous cases

Along with the already given examples 1–2, there are some more cases where non-final conjuncts demonstrate no coordination markers (Nom=Acc), making the interpretation of such constructions ambiguous. These are the following cases:

- (20) *kā(madhātu) rūpadhātu-ššana kleśanma* (PK AS 16.3 a1)  
 sphere\_of\_desire sphere\_of\_form-Adj:Pl.f affliction:Pl  
 “Afflictions of the realm of desire and of the realm of form”.
- (21) *aśiṣ maṅkāl-šana rekau(na)* (PK AS 17K a4)  
 greeting blessing-Adj:Pl.f word:Pl  
 “Words of greeting [and] blessing”<sup>9</sup>.
- (22) *kewye kuñcītā-šše pelaikne-šše śūke* (B 27 a8)  
 butter sesame-Adj law-Adj taste  
 “The butter (and) sesame taste of the law” (CEToM).
- (23) *tagaram palāśapatreṇa • tagar palāśā-šše pilta-sa •* (B 308 b5)  
 “Skt. With the leaf of *tagara* and *palāśa*”.
- (24) *asaṃkhyai-nta kalpa-nma-ššai lālyīs= aṣkār mā maukaṃ* (B 591b6)  
 asaṃkhyeya-Pl kalpa-Pl-Adj:Acc.Sg.f effort:Perl back Neg desist:3s.Sbj  
 “(Who) does not desist because of the effort of *asaṃkhyeyas* and *kalpas*”<sup>10</sup>.

Differently from the hendiadyses like *sklok prosko* “doubt and fear” (example 2) or *aśiṣ maṅkāl* “greeting and blessing” (example 21) where the words are semantically close and may easily build compounds, the pairs like *kewye kuñcīt* “butter and sesame” or *tagar palāś* (Sanskrit names of two different plants) (examples 22 and 23) have no other attestations and cannot be recognized as collocations.

Overall, around 20 examples of the suffix *-šše* adjectivizing conjuncts have been found in Tocharian B. All of them are quite equally distributed between archaic, classical and late manuscripts and are possible in prose as well as in poetry.

### 2.2. Situation in Tocharian A

As already mentioned in the introduction, the examples of this derivational model are marginal in Tocharian A. There are two occurrences of the adjective *pācar mācarši* “pertaining to father and mother”:

<sup>9</sup> Cf. further the Tocharian A example 28.

<sup>10</sup> This line is translated in Peyrot 2013: 675 as “... who does not turn back from the effort for countless ages”; however, the rendering of *asaṃkhyainta kalpanma* as “countless ages” is, in my opinion, not very precise, because literally these are two nouns in Nom-Acc.Pl (with the noun plural suffixes *-nta* and *-nma* respectively), so the Nom-Acc.Pl. form *asaṃkhyainta* could hardly serve as an attribute for Nom-Acc.Pl. *kalpanma*.

- (25) *pācar mācar-šiṃ kāpñiune pukaṃ tampewāts wrasaśsi* (YQ II.3 b5)  
 “The love for father and mother (is) the strongest (sentiment) among the beings”  
 (Ji, Winter, Pinault 1998: 79).
- (26) *sām pācar mācar-ši wci märkampal-ši* (A 291 a2)  
 “(Two bodies of the Buddha) The one (stems) from father and mother, the second from the Law” (CEToM).

However, Adams (2015: 183) treats Nom-Acc *pācar mācar* “father (and) mother” as a compound with the meaning “parents”. Indeed, this expression is highly frequent in the languages of the world in general (cf. Wälchli 2005: 3–5) and in the languages of East Turkestan in particular (cf. Rastorgueva 1981: 445), so that the syntactic status of this hendiadys is at least disputable, and the adjective can be translated simply as “parental”.

The next pair of words *ásiṣ mañkal* “greeting and blessing”, already seen in Tocharian B (example 21), can be used in Tocharian A with a common plural suffix *-ntu* (*ásiṣ mañgalntu* in A 63 a5) which also does not support the assumption about a syntactic relation between the two words:

- (27) *ásiṣ mañkalšinās rakentuyo* (A 16 a6)  
 “With words of blessing and good luck” (CEToM).

There is, however, one example which quite convincingly shows that the Tocharian A suffix *-ši* also refers to conjuncts: it is shared by two abstract nouns derived by means of the very productive suffix *-une* / *-one* from the adjectives *kāwālte* “beautiful” and *ciñcār* “tender”:

- (28) *kus ne ārkisōṣṣaṃ k(ā)wāltune ciñcronesi wākā(m) ṣ[e](ṣ)* (A 175+178+THT 2968 fgm. g a5)<sup>11</sup>  
 “Who was the paragon of beauty and tenderness in the world”.

Cf. the use of the same words *kāwāltune* “beauty” and *ciñcrone* “tenderness” in the following line: *ote tāpreṃ kāwāltun[e] ote tāpreṃ ciñcro[ne]* “Ah, such a beauty, ah, such a tenderness!” (A 186+1143+202 b4)<sup>12</sup>.

### 3. Tocharian B suffixes *-tstse* and *-ññe*

#### 3.1. *-tstse*

Along with *-šše*, there are two other particularly productive adjectival suffixes in Tocharian B, namely the suffixes *-tstse* and *-ññe*<sup>13</sup>. To our knowledge, the ability of these suffixes to adjectivize syntactic groups has not yet been specially described, although at least examples for *-tstse* referring to subordinated noun phrases are well attested, e.g. *kārtse yāmor-cepi* Gen.Sg. < *kārtse yāmor-tstse*\* “having deeds of welfare (=virtuous)” in B 121 a2 or *orotstse cimpamñe-cci* Nom.-Acc.Pl. < *orotstse cāmpamñe-tstse*\* “having great capabilities” in B 506 a3. Noticeably, the suffix *-tstse* does not require an Obl form from the adjective Nom.Sg. *orotstse* (*\*orocce cimpamñe-cci*), which would be the case with *-šše*, cf. *läksaññai klautsai-šše ṣpel* “fish ear (i.e. gill) pellet” (*läksaññai klautsai* Acc.Sg.f < *läksaññe klautso* “fish ear”) in PK AS 3b B2.

Moreover, *-tstse* adjectivizes prepositional phrases, such as *śle krake-tse* from *śle krake* “with dirt (=dirty)” (PK NS 54 b4) or *snai yparwe-ccē* Acc.Sg. < *snai yparwe-tstse* from *snai yparwe* “end-

<sup>11</sup> Join according to Itkin (in print).

<sup>12</sup> Join according to Ilya B. Itkin (personal communication).

<sup>13</sup> See more on their use and semantics in Krause, Thomas 1960: 146–147.

less (lit. “without beginning”)” (IOL Toch 5 b6). In Tocharian, prepositional phrases like *śle krake* and *snai yparwe* are usually lexicalized and function both as nouns and as adjectives. It is thus possible that their semantically excessing adjectivization with *-tstse* serves to distinguish between nominal and attributive functions of such constructions.

We have found only one example of *-tstse* possibly referring to a coordinated group; this example also belongs to the adjectives derived from prepositional phrases (preposition *snai* “without”):

- (29) • *asaṅgam anapragraham* • *snai trenkāl snai krämpālyñetse* (B 251 b2)  
 “Skt. Without attachment (to the world), without distraction”.

However, the interpretation “Without attachment (to the world), having no distraction”, i.e. with *-tstse* referring only to the second prepositional phrase, is plausible as well.

It should be noticed that the suffix *-tstse* can be added to verb stems, e.g. *prekṣātstse* < *pärk-* “ask” (B 331 b1), *kautātstse* < *kaut-* “destroy” (THT 1681 a5), *tsakātstse* < *tsāk-* “sting, bite” (B 88 a3–4), *rīnātstse* < *rīn-* “renounce” (B 245 a4), *anaiwatstse* < *aiw-* “turn” with the negative prefix *\*e(n)-* (B 5 a5–6)<sup>14</sup>. It is probably the ability of *-tstse* to link syntactically non-independent verbal stems which explains its low productivity in the discussed derivational model: perception of the suffix *-tstse* by Tocharian speakers to be more “bound” within a word could impose a restriction on wide-scope affixation with conjuncts. On the contrary, the suffix *-ṣṣe* never occurs with verbal stems.

### 3.2. *-ññe*

Cases of group derivation by means of the suffix *-ññe* are exceedingly rare. As for subordinated groups, it occurs in the adjective *śwālyai paiyye-ññe* from *śwālyai paiyye* “right foot” (the adjective *śwālyai* is indeclinable).

Another possible example is PK NS 32 a4-a5, although the word *aiṣṣeñcaññe* (from agent noun *aiṣṣeñca* “one who gives”) is treated in CEToM as an abstract noun “giving” derived with the identical nominalizing suffix *-ññe*:

- (30) *ambari lānte*                      *pontaṃts*      *aiṣṣeñca-ññe*      *kärtse-ṣṣe*  
 Ambara king:Gen.Sg      all:Gen.Pl      giving-Der      virtue-Adj  
*ñem-kālywe*      *klyauṣāte* (PK NS 32 a4–5)  
 fame                      hear:Mid.3s.Pt

“The king Ambara was heard of, [namely] the goodness based on the giving of everything” (CEToM).

In my opinion, *pontaṃts aiṣṣeñca-ññe* can be understood as an adjective derived from the phrase *pontaṃts aiṣṣeñca* “giving to all” (*pontaṃts* Gen.Pl. “all”, *aiṣṣeñca* “giving”). This adjective is an attribute for *kärtse* “virtue” which is in turn the basis for the adjective in *-ṣṣe* that modifies the dvandva *ñem-kālywe* “fame”. The structure of all these nested noun phrases can be schematized as follows: [[*ambari lānte* [*pontaṃts aiṣṣeñca*]-*ññe* *kärtse*]-*ṣṣe* *ñem-kālywe klyauṣāte*] “The fame of King Ambara’s virtue of giving to all was heard”.

Adjectivization of coordinated groups with the suffix *-ññe* can be found just in one example where it refers to three words *yasa ñkante wrāko\** “gold, silver, pearl”:

<sup>13</sup> See in Adams 2013: 14.

- (31) *şadvarginta yasa ñkante wrāka-ññeṃ wmera makci priyeṃ* (PK AS 18A a2)  
*şadvargi:Pl gold silver pearl-Adj:Acc.Pl jewel:Pl Refl wear:3pl.Ipf*  
 “The *şadvargi*-monks wore by themselves jewels of gold, silver and pearls”.

Again, it is remarkable that *-ññe* is added to the oblique stem *wraka-* (Nom. *wrako*, Obl. *wrakai* “shell, pearl”) while the non-final conjuncts Nom-Acc *yasa* “gold” and Nom-Acc *ñkante* “silver” are morphological words (cf. the same for example 5). An interesting parallel is found in B 109 a4 where each of the same material names is adjectivized with its own suffix: *ysāşşe* with *-şşe* from Nom.-Acc.Sg. *yasa*, *ñ(i)kañce* with stem-final palatalization and the suffix *-e* from Nom.-Acc.Sg. *ñkante*, *wrākaññe* with the suffix *-ññe* from Nom.Sg *wrāko\**, Acc.Sg. *wrakai*, cf.:

- (32) *ysāşşeṃ ñikañceṃ wmera wrākaññeṃ* (B 109 a4)  
*gold:Adj:Acc.Pl silver:Adj:Acc.Pl jewel:Pl pearl:Adj:Acc.Pl*  
 “Jewels of gold, silver and pearls”.

#### 4. Suspended affixation with adjectival suffixes in some other languages

The phenomenon of suspended affixation with derivational morphemes has so far received relatively little attention in the typological perspective. On the one hand, in Spencer, Luís 2012: 200 it is noticed that “wide scope affixation can be found with inflectional and derivational morphology in a number of languages, and it is often a symptom of recent and not quite complete morphologization”. On the other hand, this special type of suspended affixation is denied, for example, for the agglutinative modern Turkic, cf. “Suspended affixation is impossible with any of the derivational morphemes of the language” (Broadwell 2008: 7); some examples of a similar kind are traditionally treated as affixation of compounds, cf. Kabak 2007: 336.

Recently, discussion on the presence of this phenomenon in modern Turkic and, probably, in some other languages has been raised, for instance, by F. Akkuş (2016), who points out that cases of suspended affixation with derivational morphemes are “rather uncommon to find, quite many to ignore” (2016: 1). According to Akkuş, nouns sharing a common derivational morpheme can be treated as syntactic groups as they can change their positions like *tuz ve biber-lik* (“of salt and pepper”) vs. *biber ve tuz-luk* (“of pepper and salt”) or be complemented by further conjuncts as well as receive a reiterated suffix, cf. *bir ve yedi-nci bölüm-ler-i-ni...* “first and seventh chapters (of the book)...” vs. *bir, yedi ve yirmi bir-inci bölüm-ler-i-ni* “(first, seventh and twenty first chapters (of the book))” (Akkuş 2016: 8–10).

But the most interesting thing is to compare Tocharian data with the situation in other ancient languages of East Turkestan because they show mutual lexical borrowings as well as a tradition of literary translation (especially from Tocharian into Old Uighur).

Adjectivization of subordinated groups, as well as some further agglutinative features, are indeed known in some Eastern Iranian languages, e.g. in Sogdian (Gershevitch 1945: 13) or in Khwarazmian (Windfuhr 2009: 323)<sup>15</sup>. However, we have not found any examples or descriptions of derivation from coordinated groups in these languages.

As for Old Uighur, the observation made in Erdal 1991: 141–142 about the adjectival suffix *-lXg*<sup>16</sup> is noteworthy: “Exceedingly interesting is the formational, syntactic and semantic similarity between *-lXg* and Tokharian *-ši* / *-şşe* [...] The parallelism between the two conjunction-formatives which is maintained in details [...] has no likeness in other Indo-European lan-

<sup>15</sup> Remarkable typological parallels to Tocharian group inflection in general can be observed in Ossetic, cf. Belyaev 2014.

<sup>16</sup> We use the spelling with the invariant *-lXg* as in Erdal 1991 and Erdal 2004.

guages [...] As Uigur is also isolated among the Turkic languages in the extraordinary rich use which it makes of *-lXg*, all we can say is that it must be an areal phenomenon common to Uigur and Tokharian”. Erdal analyses contextual and semantic similarities between A *-ši*, B *-şşe* and Uighur *-lXg* and provides examples of derivation from subordinated noun phrases (1991: 139–150); such cases are not specifically considered in this paper.

However, there are also quite a few examples of adjectives built from hendyadises by means of *-lXg* in Old Uighur. According to Erdal, “when added to binomes, *-lXg* is usually repeated, [...] Occasionally, however, the two terms do receive it in common” (1991: 147). Along with some ‘binomes’ which are attested either only with repeated or only with single adjectival suffix, some constructions of this type demonstrate optional suffix iteration (33–36) as well as changing order of elements (35–36), cf.:

(33) *örtlüg yalınluğ* (Erdal 2004: 330)  
vs. *ört yalınluğ* (DTS: 389)  
“fiery and flaming”.

(34) *čogluğ yalınl(ı)ğ* (Mainz 751/09 = DKPAM 07740 (Wilkins 2016 II: 630))  
vs. *čog yalınluğ* (DTS: 152)  
“shiny and brilliant”.

(35) *asığ tusuluğ* (Maitr 5 v25, (Erdal 1991: 148))  
vs. *asığluğ tusuluğ* and *tusuluğ asığluğ* (DTS: 60)  
“advantageous and beneficial”.

(36) *ačuq adırtluğ*  
vs. *adırtluğ ačuğluğ* (DTS: 12–13)  
“clear and obvious”.

As for ‘binomes’, Erdal indicates that numerous lexemes in Old Uighur are used in “fixed two-word sequences to render a single notion” and can either show or lack morphological fusion (2004: 533), so it must be difficult to define whether such paired lexemes underlying the adjectives are compounds or nouns in coordination.

However, the suffix *-lXg* can refer to longer sequences of words as well. The most remarkable example of this kind known to us is found in the Uighur version of Maitrisimit (MaitrHami, XX, leaf 14, 10–11). The adjective in *-lXg* here is an attribute to *otın* “fire”, the suffix is shared by four conjuncts:

(37) *ačmak [suvsə]mak busuš kadguluğ otın*  
“(Suffering from) fire of hunger (and) thirst, sorrow (and) distress”.

We now know the original version in Tocharian A for this line of the Old Uighur manuscript; however, it differs from the translation in structure as well as partially in content. Namely, it has no adjectival correspondence to the Uighur construction in *-lXg*. The word *por* “fire” in the Tocharian version is not the syntactic head but one of three homogenous Instr forms in *-yo*; Uighur hendyadises *ačmak suvsamak* “hunger and thirst” and *busuš kadgu* “sorrow and distress” render single words *kašt* “hunger” and *šurām* “sorrow” of the Tocharian original, cf.:

(38) *[ka]štjo [šurā]m[yo po](ryo) ///* (THT 1308 b2)  
“(Tormented by) hunger (and) sorrow (and) fire”<sup>17</sup>.

It might be interesting to collect more examples like (37) (with *-lXg* linking longer chains of conjuncts) to evaluate how widespread this type of group affixation is in Old Turkic.

<sup>17</sup> Identification and restoration according to Itkin, Kuritsyna 2017: 67.

## 5. Conclusion

The suffix A *-ši* / B *-šše* has an almost exclusive ability among derivational morphemes of Tocharian to link conjuncts. This suffix can refer to significantly more than two nouns, its iteration or suspension is optional, and the coordinated nouns in such adjectives can represent extended noun phrases. Some rare examples of this kind can also be found with other productive suffixes – B *-tstse* and *-ññe*. This phenomenon replicates Tocharian suspended affixation, when a secondary case suffix is added to a chain of conjuncts. Such a model probably arose after A *-ši* / B *-šše* had acquired the ability to link morphological words (Nom-Acc. or explicitly Acc. forms). A typological and areal parallel may be represented in Old Uighur, at least with the adjectival suffix *-lXg*. In Old Turkic constructions, however, the boundary between morphological and syntactic units is hard to define, while in the Tocharian examples coordination markers can occur inside such constructions. The Tocharian data presented in this paper can contribute to the discussion on which types of affixes can be affected by suspended affixation in typological perspective.

## References

- Adams, Douglas Q. 2009. Genitive and adjective in Tocharian. *Journal of Indo-European Studies* 37 (3/4): 299–320.
- Adams, Douglas Q. 2013. *Dictionary of Tocharian B. Revised and Greatly Enlarged*. Amsterdam – New York: Brill / Rodopi.
- Adams, Douglas Q. 2015. *Tocharian B: A grammar of syntax and word-formation*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Akkuş, Faruk. 2016. Suspended affixation with derivational suffixes and lexical integrity. In: Nikos Koutsoukos, Francesca Masini, Jenny Audring (eds.). *Online Proceedings of 10th Mediterranean Morphological Meeting*: 1–15. Available online at: <http://mmm.lis.upatras.gr/index.php/mmm/article/view/2720> [accessed 10.11.2018].
- Belyaev, Oleg I. 2014. Osetinskij kak jazyk s dvuxpadežnoj sistemoj: gruppovaja fleksija i drugie paradoksy padežnogo markirovanija. *Voprosy Jazykoznanija* 6: 31–65.
- Broadwell, George A. 2008. Turkish suspended affixation is lexical sharing. In: Miriam Butt, Tracy Holloway King (eds.). *Proceedings of the LFG08 Conference*: 198–213. Stanford, CA: CSLI.
- Burlak, Svetlana A., Itkin Ilya B. (in progress). *Formal'naja Grammatika Toxarskogo A Jazyka: fonologija, morfonologija, morfologija*.
- CEToM = *A Comprehensive Edition of Tocharian Manuscripts*. Available online at: <http://www.univie.ac.at/tocharian/> [accessed 05.12.2018].
- DTS = V. M. Nadel'aeu, D. M. Nasilov, E. R. Tenišev, A. M. Ščerbak (eds.) 1969. *Drevnet'urkskij Slovar'*. Leningrad: Nauka.
- Geng, Shimin, Hans-Joachim Klimkeit, Jens Peter Laut. 1998. *Eine buddhistische Apokalypse. Die Höllenskapitel (20–25) und die Schlußkapitel (26–27) der Hami-Handschrift der alttürkischen Maitrisimit. Unter Einbeziehung von Manuskriptteilen des Textes aus Sänjim und Murtuk (Abhandlungen der Nordrhein-Westfälischen Akademie der Wissenschaften 103)*. Wiesbaden: Westdeutscher Verlag.
- Erdal, Marcel. 1991. *Old Turkic Word Formation. A functional approach to the lexicon. Vol. I*. Wiesbaden: Otto Harassowitz.
- Erdal, Marcel. 2004. *A Grammar of Old Turkic (Handbook of Oriental Studies, Section 8 Uralic & Central Asia. Vol. 3)*. Leiden / Boston: Brill.
- Gershevitch, Ilya. 1945. Sogdian compounds. *Philological Society* 44 (1): 137–149.
- Hajnal, Ivo. 2004. Zur Genese agglutinierender Flexionsmuster im Tocharischen: die Adjektive auf B *-šše* / A *-ši*. In: P. Anreiter, M. Haslinger, H. D. Pohl (eds.). *Artes et Scientiae. Festschrift für Ralf-Peter Ritter zum 65. Geburtstag*: 137–157. Wien: Praesens Verlag.
- Itkin, Ilya B. (in print). *Toxarskaja A rukopis' №№ 144–211 iz Šorčuka: novye dannye*.
- Itkin, Ilya B., Kuritsyna, Anna V. 2017. Chapter XX of the “Maitreyasamiti-Nāṭaka” and its hellish sufferings: the fragment THT 1308.a. *Tocharian and Indo-European Studies* 18: 63–71.

- Ji, Xianlin, Werner Winter, Georges-Jean Pinault. 1998. *Fragments of the Tocharian A Maitreyasamiti-Nāṭaka of the Xinjiang Museum, China. Transliterated, translated and annotated by Ji Xianlin in collaboration with Werner Winter, Georges-Jean Pinault (TSM 113)*. Berlin / New York: De Gruyter.
- Kabak, Barış. 2007. Turkish Suspended Affixation. *Linguistics* 45 (2): 311–347.
- Krause, Wolfgang, Werner Thomas. 1960. *Tocharisches Elementarbuch. Vol. 1*. Heidelberg: Carl Winter.
- Lewis, Geoffrey L. 1967. *Turkish Grammar*. Oxford University Press.
- Peyrot, Michaël. 2008. *Variation and change in Tocharian B (Leiden Studies in Indo-European 15)*. Amsterdam / New York: Rodopi.
- Pinault, G.-J. 1992. *Introduction au tokharien. LALIES 11. Actes des Sessions de Linguistique et de Littérature (Cortona, 20–31 août 1990)*: 3–163. Paris: École Normale Supérieure.
- Rastorgueva V.S. (ed.). 1981. *Osnovy Iranskogo Jazykoznanija. Sredneiranskie jazyki*. Moskva: Nauka.
- Ringe, Donald A. Jr. 1996. *On the Chronology of Sound Changes in Tocharian. Vol. 1: From Proto-Indo-European to Proto-Tocharian (American Oriental Series 80)*. New Haven, Conn.: American Oriental Society.
- Schmidt, K.H. 1969. Agglutination und Postposition im Tocharischen. *Münchener Studien zur Sprachwissenschaft* 25: 105–112.
- Sieg, Emil, Siegling, Wilhelm. 1921. *Tocharische Sprachreste. Bd. I. Die Texte. A: Transkription*. Berlin / Leipzig: Walter de Gruyter.
- Sieg, Emil, Wilhelm Siegling, Wilhelm Schulze. 1931. *Tocharische Grammatik*. Göttingen: Vandenhoeck & Ruprecht.
- Sieg, Emil, Wilhelm Siegling. 1949. *Tocharische Sprachreste. Sprache B. Heft 1. Die Udānālankāra-Fragmente. Text, Übersetzung und Glossar*. Göttingen: Vandenhoeck & Ruprecht.
- Spencer, A., A. R. Luís. 2012. *Clitics: An Introduction (Cambridge Textbooks in Linguistics)*. New York: Cambridge University Press.
- Thomas, Werner. 1972. Zweigliedrige Wortverbindungen im Tocharischen. *Orbis* 21: 429–470.
- Van Windekens, Albert Joris. 1979. *Le Tokharien Confronté avec les Autres Langues Indo-Européennes. Vol. II, 1: La morphologie nominale. (Travaux publiés par le Centre International de Dialectologie Générale de l'Université Catholique de Louvain 12)*. Louvain: Centre international de dialectologie générale.
- Wälchli, Bernhard. 2005. *Co-Compounds and Natural Coordination (Oxford Studies in Typology and Linguistics Theory)*. Oxford University Press.
- Wilkins, Jens. 2016. *Buddhistische Erzählungen aus dem alten Zentralasien. Edition der altuigurischen Daśakarmapathāvadānamālā. 3 vols*. Turnhout: Brepols.
- Windfuhr, G. (ed.). 2009. *The Iranian Languages*. London / New York: Routledge.
- Winter, Werner. 1985. Tocharian B soy, A se and related forms. *Journal of the American Oriental Society* 105: 259–264.
- Zimmer, Stefan. 1982/1983. Die Funktion der tocharischen *ṣi/ṣṣe*-Adjektive. *Zeitschrift für Vergleichende Sprachforschung* 96 (2): 277–289.

А. В. Курицына. К вопросу о групповой аффиксации в тохарском (посредством адъективного суффикса А -ṣi / В -ṣṣe) и ее возможных параллелях в древнеуйгурском

В статье рассматривается не описанная ранее способность тохарского суффикса А -ṣi / В -ṣṣe (в единичных случаях также -tstse и -ññe) образовывать прилагательные от однородных членов. Предпринимается попытка доказать, что в основе таких конструкций лежат сочиненные именные группы, а не композиты-двандвы. Полученные данные сравниваются с материалом ареально близкого древнеуйгурского языка. Если наше предположение верно, то групповая аффиксация сочиненных групп в тохарских языках возможна не только со словоизменительными морфемами вторичных падежей, но и со словообразовательными, хотя и менее распространена для последних.

Ключевые слова: групповая флексия, групповая деривация, тохарские адъективные суффиксы, древнеуйгурский язык, ареальные контакты



## On the place of Parachi andOrmuri among the Iranian languages according to the data of annotated Swadesh lists<sup>1</sup>

In this paper, I discuss the open issue of the phylogenetic position of Parachi andOrmuri among other Iranian languages. Based on data from freshly constructed annotated 110-item Swadesh lists of Parachi andOrmuri compared with the same list as tentatively reconstructed for the Proto-Iranian stage (including additional synonyms with Western or Eastern areal distribution), I conclude that Parachi andOrmuri more probably align with the Eastern Iranian branch of languages. This is primarily concluded from the fact that Parachi andOrmuri show correlations with secondary synonyms of Eastern origin. It is important to note that phonetic isoglosses are not nearly as indicative, since Parachi andOrmuri show both Western and Eastern peculiarities; however, the prosodic systems of these languages are strongly reminiscent of the one attested in Pashto and preserving the Proto-(Indo)-Iranian system better than any other Iranian language. These conclusions support Morgenstierne's original opinion on the Eastern character of Parachi andOrmuri; on the other hand, Morgenstierne's assumption of the existence of a separate Southeastern Iranian group requires additional confirmation.

Keywords: Parachi language, Ormuri language, 110-item Swadesh list, Proto-Iranian, Eastern Iranian languages, Western Iranian languages, phonetic isoglosses, prosodic systems, Pashto language.

Parachi and Ormuri are two minor Iranian languages which have always presented a significant problem for the internal phylogenetic classification of the Iranian group, with different scholars expressing different opinions regarding their position among other Iranian languages. Thus, G. Morgenstierne thought that Parachi and Ormuri belonged to the Southeastern Iranian group (Morgenstierne 1926: 27–36). On the one hand, he did concede that there were some sound changes common for Parachi, Ormuri and Western Iranian languages, emphasizing that “Parachi and W. Ir. have initial voiced stops (*b, d, g*) and the palatal affricate *ǰ*, unlike the eastern dialects which have fricatives ( $\beta$  (*v*),  $\delta$ ,  $\gamma$ , *ž*)” (Morgenstierne 1926: 28; Morgenstierne 1929: 8). On the other hand, he listed numerous examples of Parachi and Ormuri words having specific Eastern Iranian correspondences (Morgenstierne 1926: 31–33; Morgenstierne 1929: 8–12; 317–318).

The opposite opinion was first expressed by G. A. Grierson who wrote that Ormuri “agrees generally, and closely, with the Western Iranian dialects, including Kurdish” (Grierson 1918: 49–52). Similar views were presented in Tedesco 1921 and Oranskij 1979: 81–128. The most extensive argumentation is presented in the monograph Efimov 1986, recently translated into English (Efimov 2011)<sup>2</sup>. According to V. A. Efimov (Efimov 1986: 9; Efimov 2011: 5–6), the following reasons should lead us to consider that Parachi and Ormuri are Western, rather than Eastern, Iranian languages:

---

<sup>1</sup> I would like to thank Dr. Ilya Yakubovich for many valuable helpful comments and remarks. All remaining errors are my own.

<sup>2</sup> For the sake of convenience I cite the English translation throughout this article.

- 1) Ir. \*b- > Orm. b-;
- 2) Ir. \*d- > Orm. d-;
- 3) Ir. \*g- > Orm. g-;
- 4) Ir. \*-č > Orm. ž/z (Log. ž, Kan. z).

Furthermore, Efimov adduces certain reflexes that fit with the positioning of Ormuri in the Northwestern subgroup of Western Iranian: “the reflexes of Old Iranian \*z/\*d, \*s/\*θ and \*sp corresponding with Ormuri z, s, sp, respectively” and “the reflex of Ir. \*θr/\*θr as Log. š, Kan. ř” (Efimov 1986: 9–10; Efimov 2011: 6–7).

The view of Morgenstierne is accepted, for example, in Kieffer 1989 and Kieffer 2009. However, it is important to note that Ch. M. Kieffer is not that optimistic about Morgenstierne’s hypothesis in the first work: he mentions both Morgenstierne’s and Efimov’s views noting Efimov’s “*riche argumentation*” (Kieffer 1989: 452). Efimov’s view is accepted in EDIL where both Parachi and Ormuri forms are listed among the data of the Northwestern Iranian languages.

It is clear that Efimov takes into consideration exclusively those isoglosses that concern segmental phonetics. However, there is an additional important trait that is common for Parachi, Ormuri and Eastern Iranian, namely, traces of the Proto-Iranian accentuation, preserved in Pashto and some other Eastern Iranian languages (for example, in Wakhi and Yidgha-Munji languages; see Mayrhofer 1989: 13).

Parachi and Ormuri have free dynamic stress, mobile in the former language and non-mobile in the latter. Efimov characterizes the stress of pre-Ormuri in the following way: “However, as with the Old Iranian dialect to which Ormuri can be traced back, its position is not determined by counting back the long and short syllables from the end of the original word form, that is to say, it does not depend on quantitative factors. From analysis of the material available it can be postulated with sufficient confidence for the original, Old-Iranian dialect that there was a Vedic type of free stress, the reflexes of which may be observed in both the verb and noun systems in varying degrees, although in a more limited and sometimes altered form” (Efimov 2011: 90). This characteristic immediately reminds of Pashto stress, researched in detail by V. A. Dybo in two articles (Dybo 1974; Dybo 1989). Dybo has demonstrated that Pashto inherited free Indo-Iranian stress and preserved some traits of it even better than Vedic. Therefore, it would be very important to investigate the Parachi and Ormuri systems of accentuation and their connection to systems of Pashto and other Eastern Iranian languages that preserve to some extent traces of free stress. It seems that at least Ormuri barytone nouns generally correspond to barytone nouns in Pashto and the same rule is applicable to oxytone nouns, as suggested by the following matches:

1) barytona:

Pashto *áspa* f. ‘horse’ ~ Ormuri Kan. *yāsp* Log. *yāsp*, f. *yāspa* ‘horse’ ~ Vedic *ásvā-* ‘horse’ (Dybo 1974: 72; Efimov 2011: 90);

Pashto *wáwra* ‘snow’ ~ Ormuri Kan. *γoř* Log. *γoř* ‘snow’ ~ Vedic *vápra* ‘mound; earthwork, rampart, mud wall’ (Dybo 1974: 74; Efimov 2011: 90);

Pashto *wáč* f. *wáča* ‘dry’ ~ Ormuri Kan. *wyok* Log. *wuk* ‘dry’ ~ Vedic *súška-* ‘dry’ (Dybo 1974: 74; Efimov 2011: 90);

2) oxytona:

Pashto *muryá*, *mərgá* ‘bird (of prey)’ ~ Ormuri Kan. *mírgá* Log. *morgá* ‘bird / little bird / sparrow’ ~ Vedic *mrgá-* ‘a forest animal or wild beast, game of any kind, (esp.) a deer, fawn, gazelle, antelope, stag, musk-deer’ (Dybo 1974: 82; Efimov 2011: 90);

Pashto *plār* ‘father’ ~ Ormuri Kan. *pye* ~ *piyé* Log. *pe* ‘father’ ~ Vedic *pitár-* ‘father’ (Dybo 1974: 74; Efimov 2011: 90);

Pashto *nwí, niwí, nəwí* ‘90’ ~ Ormuri Kan. Log. *naŋwí* ‘90’ ~ OInd. *navatí* (Dybo 1974: 85; Efimov 2011: 292).

Such correspondences are hardly coincidental, which means that Parachi and Ormuri share at least one very important Eastern Iranian phonetic feature. In order to ascertain its importance for Iranian phylogeny, and also in light of the conflicting scholarly views on the role of phonetic isoglosses in the classification of Parachi and Ormuri on the whole, it seems reasonable to supplement phonetic data with a detailed analysis of the lexical material of the languages under investigation.

To do this, we will recur to lexicostatistical methodology and test the two conflicting hypotheses with the aid of 110-item Swadesh lists for Parachi and Ormuri, contrasting them with the same list as tentatively reconstructed for the Proto-Iranian stage. The methodology of the proto-list reconstruction is described in (Starostin 2016).

For the sake of additional accuracy, the Proto-Iranian list as used in this study includes not only the main (“optimal”) entry, but also other (usually less likely) candidates with indications of their areal distribution. After that, the data of Parachi and Ormuri (which, since the classification of these languages is a priori uncertain, was not used itself for the reconstruction of the Proto-Iranian wordlist in order to avoid circularity) is compared with the reconstructed Proto-Iranian list in order to discover whether these two languages have more specific lexical correlations with Western or Eastern Iranian.

I prefer such a methodology over the more standard lexicostatistical method (comparing data from modern and historically attested languages) for several reasons. First, at this moment we do not have a sufficient number of high-quality annotated Swadesh lists, agreeing with the specifications of the Global Lexicostatistical Database, for Iranian languages (see existing lists at <http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ier\irn&limit=-1>). Second, it is *per se* interesting to establish the number of specific Western and Eastern isoglosses contained in the list of Proto-Iranian and whether it is possible to make conclusions about the affiliation of Parachi and Ormuri using these data.

The complete data are presented below. The main sources for the Proto-Iranian reconstruction and the list of proto-forms are EDIL and Sergei Nikolaev’s Indo-European etymological database included into The Tower of Babel project (<http://starling.rinet.ru>). The main sources for the Parachi list are G. Morgenstierne’s texts and vocabulary (Morgenstierne 1929) and V. A. Efimov’s grammar with texts and vocabulary (Efimov 2009). The main sources for the Ormuri list are G. Morgenstierne’s Logar texts and vocabulary, including words from both the Logar and Kaniguram dialects (Morgenstierne 1929), V. A. Efimov’s Kaniguram and Logar texts and grammar with etymological vocabulary (Efimov 2011) and D. G. Hallberg’s 210-item wordlist for Kaniguram partly coinciding with Swadesh lists (Hallberg 2004). Based on these data, it is possible to make some observations on the Parachi and Ormuri lists and count areal isoglosses.

First, it is interesting that Kaniguram and Logar dialects show numerous differences in both inherited and borrowed lexicon. The following cases are attested:

10. ‘bone’: Kan. *yaḍ-* ~ *haḍ* ~ Log. *stoyán* {stoγ’ān}
30. ‘fly (v.)’: Kan. *bur-* / *bur-yék* ~ Log. *par-* / *par-ók*
44. ‘knee’: Kan. *zan-řak* / *zān-řak* ~ Log. *zānú* {zā’nū}
46. ‘leaf’: Kan. *puxay* ~ Log. *barg*
54. ‘moon’: Kan. *maryok* / Kan. *spožmay* ~ Log. *mátáw*
72. ‘see’: Kan. *dzun-* ~ Log. *juǰ-*
75. ‘skin’: Kan. *tsarmun* / *čāmṛa* ~ Log. *pōst*

78. 'smoke': Kan. *lugay* ~ Log. *dud* {dūd}  
 79. 'stand': Kan. *ašt-* / *ašt-yék* / Kan. *t-yek* ~ Log. *dar* / *dar-ók*  
 90. 'tree': Kan. *wína* ~ Log. *draxt*  
 101. 'far': Kan. *pets* ~ Log. *dur*  
 102. 'heavy': Kan. *grān* Log. *wazmīnd*  
 106. 'snake': Kan. *māngor* ~ Log. *mār*.

It must, however, be noted that some of these discrepancies may be fictitious due to lack of sufficient information in our sources – it is possible that some of the loanwords are actually used in both dialects, or that one dialect uses not one but many borrowings as synonyms.

Second, there are 3 exclusive Parachi-Ormuri isoglosses in the 110-item wordlist:

1. 'all': Parachi *hu* / *hos* ~ Ormuri (Kan., Log.) *ayéra* < PIr. \**harwa* (EDIL 3: 372–373).  
 19. 'drink': Parachi *t'er-* / *t'ar-* / *t'ōr-* ~ Ormuri (Kan., Log.) *tr-* / *ta-tak* / *tō-tk* < PIr. \**tṛ-ya-* / \**tṛšta(ka)-* (Morensterne 1929: 10, 295).  
 28. 'fire': Parachi *r'ínē* ~ Ormuri (Kan.) *ráwan*, (Log.) *rówon* < PIr. \**rauxšna-* (Morgensterne 1929: 284; Efimov 2011: 296).

Although this number seems to be rather small, it is important to remember that both languages have a significant number of borrowings and lexicographic lacunae; moreover, sometimes it is possible to find such exclusive non-lexicostatistical pairs with divergent meanings as Ormuri *gap* 'stone' ~ Parachi *gáp-âṛ* 'fireplace' (Morgensterne 1929: 10). The natural conclusion is that these two languages really form one subgroup, going back to an intermediate ancestor.

Finally, the number of areal isoglosses in the Proto-Iranian wordlist is significant. The following 15 isoglosses define the Western group:

4. 'belly': SW \**škamb-*, NW \**lab-*.  
 7. 'bite': SW, NW \**gaṣ-*.  
 11. 'breast': SW, NW \**cai-na-*.  
 26. 'fat': SW, NW \**čarp-īh-*.  
 36. 'hair': SW, NW \**maud-a-*.  
 43. 'kill': SW, NW \**kauš-*.  
 55. 'mountain': SW, NW \**kaufa-*.  
 56. 'mouth': SW, NW \**ṣaf-an-*.  
 61. 'nose': SW, NW \**wain-a-* ~ \**wain-i-*.  
 70. 'sand': SW, NW \**raika-*.  
 71. 'say': SW, NW \**gaub-*.  
 104. 'salt': NW \**hwād-*.  
 106. 'snake': SW, NW \**mār-a-*.  
 107. 'thin': SW, NW \**nāz-u-ka-*.

Two of these isoglosses are specifically Northwestern, one specifically Southwestern; all the others are common to both branches.

There are 10 Eastern isoglosses in the Proto-Iranian list:

5. 'big': \**stūra-*  
 12. 'burn': \**θaw-*  
 15. 'cold': \**stāba-*  
 18. 'dog': \**kuta-* / fem. \**kuti(ī)-*  
 26. 'fat': \**wāzdā-*  
 29. 'fish': \**kapa-*

35. 'green': \**axšaina-*  
 36. 'hair': \**gaun-a-*  
 93. 'warm': \**tap-*  
 106. 'snake': \**δανύγc-* (the Proto-Shughni form; Morgenstierne 1974: 30).

It seems clear that Parachi and Ormuri do not generally share Western isoglosses. There is only one clear case: Parachi *gas-* / *gast-* / *gōst-* 'to bite / to sting' is derived from \**ga3-* 'to bite / to sting' (EDIL 3: 263–265; a related word in Pashto is *ayzáy* 'thorn', see Cheung 2007: 118, but the verb itself is not attested in Eastern Iranian).

In contrast, Parachi and especially Ormuri do share several Eastern innovations (and this result fully agrees with Morgenstierne's lexical observations).

Parachi shows 3 matches with Eastern Iranian:

12. 'burn': Parachi *t'ēw-* / *thēyí* ~ EIr. \**θaw-*;  
 26. 'fat': Parachi *γazd* ~ EIr. \**wāzdā-*;  
 36. 'hair': Parachi *ginó* ~ EIr. \**gaun-a*.

Also worth mentioning is Parachi *ta'pō*, which at face value is recognizable as a Pashai loanword, but it is also possible that earlier \**tapaka-* was deformed under Pashai influence; compare Ormuri (Kan., Log.) *tok* 'warm / hot'.

Ormuri has 4 matches with Eastern Iranian:

5. 'big': Ormuri (Kan.) *stur*, (Log.) *stor* ~ EIr. \**stūra-*;  
 26. 'fat': Ormuri (Kan.) *γweztzi* ~ EIr. \**wāzdā-*;  
 35. 'green': Ormuri (Kan.) *šin*, (Log.) *šin* ~ EIr. \**axšaina-*;  
 93. 'warm': Ormuri (Kan., Log.) *tok* ~ EIr. \**tap-*.

My conclusions are based on the theoretical assumption that the Proto-Iranian 110-item Swadesh list is sufficient to demonstrate the affinity of two languages with a concrete subgroup of related idioms, while theories of classification that are based on phonetic isoglosses alone are generally weak and demand lexical support (cf. Dybo, Starostin 2008). If so, then evidence provided by the basic lexicon leads to the natural conclusion that Parachi and Ormuri are not Western (Northwestern) Iranian languages, and that it makes more sense to classify them as Eastern Iranian languages, following Morgenstierne (however, the notion of Southeastern Iranian as a genetic subgroup is controversial and requires independent confirmation).

That said, this result should not be considered as final, and further investigation is necessary to demonstrate the exact degree of relationship between Parachi–Ormuri and other Eastern Iranian languages. I think that there are three main directions in which this investigation should be taken. First of all, it is necessary to increase the number of high-quality annotated Swadesh lists for Iranian languages in order to obtain more precise lexicostatistical matrices and arrive at more refined phylogenetic conclusions. The second task is to revise the lexical isoglosses assembled by Morgenstierne and add new data gathered from text corpora and etymological dictionaries. It would also be useful to conduct a thorough analysis of the relatively scarce specific correspondences between Parachi–Ormuri and Western Iranian subgroups and see whether they can be explained as cases of accidental homoplasy or results of contact-induced change.

Finally, a third direction of further studies – arguably the most intriguing one for specialists not only in Iranian, but also in general Indo-European studies – should consist of research on the prosodic systems in Parachi and Ormuri and their relation to the corresponding systems in Pashto and other Eastern Iranian languages.

## Appendix 1. 110-item Swadesh lists for Proto-Iranian, Parachi and Ormuri.

*Preliminary note:* I mostly use Efimov's transcription for both Parachi and Ormuri. Present and past Parachi verbal stems are cited according to Efimov 2009 (see especially Efimov 2009: 65–71). Lexical synonyms are only given in those cases where available data does not allow to choose the most accurate semantic equivalent for the required Swadesh meaning. Borrowings are marked with negative indexes (–1), with the name of the donor language or language group indicated after the index.

All Ormuri forms are marked as Kan. (Kaniguram) and Log. (Logar), since there are significant differences between two dialects. Three verbal stems are cited consistently: the present stem, masc. sg. past participle, sg. fem. past participle<sup>3</sup>.

An extended annotated version of the Parachi list can be found at the Global Lexicostatistical Database (<http://starling.rinet.ru/new100>).

Word	Proto-Iranian	Parachi	Ormuri
1. all	*wic-w-a-	hu ~ hos	Kan. Log. ayéra
2. ashes	*ātr-ya-	ǰoŋd -1 Ind.	Kan. yānak ~ yāŋk
2. ashes <sub>2</sub>		b'āγ -1 Pers.	
3. bark	*pawasta-	pust -1 Pers.	Log. pāñī -1 Psht. ?
4. belly	*udar-a-	aštáw	Kan. ðim -1 Dard.
4. belly <sub>2</sub>	*škamb- SW		Kan. Log. nas -1 Psht.
4. belly <sub>3</sub>	*lab- NW		
5. big	*ma3- ~ *ma3-ant- ~ *ma3-a-na-	g'aŋd -1 Ind. / Dard.	Kan. stur Log. stor
5. big <sub>2</sub>	*stūra- EIr		
6. bird	wi-	not attested	Kan. mirgá Log. morgá
7. bite	*danc-	gas- / gast- / gōst-	Kan. Log. gran-
7. bite <sub>2</sub>	*ga3- SW NW		Kan. xr- / x'wá-lak / x'wā-lk Log. xr- / xó-lok
8. black	*cyāw-a-	pádō -1??	Kan. Log. γrās
9. blood	*wah-un-i- ~ *wah-un-a-	hin	Kan. sun Log. šun
9. blood <sub>2</sub>			Log. in -1 Psht. ?
10. bone	*ast- ~ *ast-i-	hađ -1 Ind. / Dard.	Kan. yađ- ~ hađ -1 Psht. < Ind. / Dard.
10. bone <sub>2</sub>			Log. stoyán {stoy'án}
11. breast	*war-ah-	siz	Log. sónok ~ sínak
11. breast <sub>2</sub>	*cai-na- SW NW	bar -1 Pers.	
12. burn	*dag-	t'ēw- / thēyí	Kan. braz- / braš-tak / bruš-k Log. bróš-tok
12. burn <sub>2</sub>	*θaw- EIr (Pamir)		
13. claw (nail)	*nāx-u-na- / *nāx-a-na-	nōrk	Kan. naxk Log. néxči (sg. and pl.)
14. cloud	*abr-a-	ayír	Kan. abar Log. yēwər
14. cloud <sub>2</sub>	*maig-a		
15. cold (adj.)	*car-ta-	eštáw	Kan. Log. tsak

<sup>3</sup> The third form is frequently not attested.

Word	Proto-Iranian	Parachi	Ormuri
15. cold <sub>2</sub> (adj.)	*stāba-		
16. come	*gam-	ži-	Kan. z- ~ z-ay- / z-ok / z-āk Log. z-ey- / z-ok / z-āk
16. come <sub>2</sub>		āγ-	
17. die	*mar-	mer- / moγ-	Kan. mr- / mú-lak / ma-lk Log. mr- / mó-lok
18. dog	*cwan-	espó	Kan. spak Log. spok
18. dog <sub>2</sub>	*kuta- / fem. *kuti(ī)-		
18. dog <sub>3</sub>	*gad-w-a-		
19. drink	*hwar-	t'er- / t'ar- / t'ōr-	Kan. Log. tr- / ta-tak / tō-tk
19. drink <sub>2</sub>	*pī- / *pā-		
20. dry	*huš-ka-	hóškō	Kan. wyok Log. wuk
21. ear	*gauš-a-	guš	Kan. Log. goy
22. earth	*zam-	d'arám -1 Ind. / Dard.	Log. búma ~ bóma Kan. búma ~ bumba
22. earth <sub>2</sub>			Log. jēr
23. eat	*hwar-	xar- / xuγ-	Kan. xr- / x <sup>w</sup> á-lak / x <sup>w</sup> ā-lk Log. xr- / xó-lok
24. egg	*āi-a- ~ *āi-a-ka-	ēx	Kan. ālk ~ hāwalk ~ wonk Log. wolk
25. eye	*čaš-man-	teč	Kan. tsom Log. tsími (sg. and pl.)
26. fat	*pī-wah-	čarbí -1 Pers.	Kan. γweztzi
26. fat <sub>2</sub>	*čarp-īh- SW NW	γazd	
26. fat <sub>3</sub>	*wāzdā- EIr		
27. feather	*par-na-	pōn	Kan. púxay (pl.)
27. feather <sub>2</sub>		p'ar -1 Pers.	
28. fire	*ā-tar-	ār	Kan. ráwan Log. rówon
28. fire <sub>2</sub>		r'ínē	
29. fish	*macy-a- (< *matsy-a-)	m'ásō	Kan. Log. māyí -1 Pers.
29. fish <sub>2</sub>	*kapa-		
30. fly v.	*pat-	rāz-	Kan. bur- / bur-yék
30. fly v.		p'arak- -1 Pers. ?	Log. par- / par-ók -1 Pers. ?
31. foot	*pad-	pā	Kan. pā-γi Log. pāy
32. full	*pγ-na-	t'ad	Kan. đāk Log. đāk -1 Psht.
33. give	*dā-	dah- / dā-	Kan. řáw- / řyuk / řuk Log. šáw- / šuk
34. good	*wah-u-	bakár -1 Pers.	Kan. sir ~ sirr Log. šer
35. green	*zar-i-	sawz -1 Pers.	Kan. šin Log. xín
35. green <sub>2</sub>	*axšaina- EIr		
36. hair	*warc-a-	ginó	Kan. dra / pl. dri Log. dri (sg. & pl.)
36. hair <sub>2</sub>	*maud-a-	dōš	
36. hair <sub>3</sub>	*gaun-a-		

Word	Proto-Iranian	Parachi	Ormuri
37. hand	* <i>zas-ta-</i>	<i>dōst</i>	Kan. <i>dist</i> Log. <i>dest</i>
38. head	* <i>car-ah-</i>	<i>sōr</i>	Kan. Log. <i>sar</i>
39. hear	* <i>craw-</i>	<i>harw- / hat- / hōt-</i>	Kan. <i>amár-</i> / <i>amar-yék</i> Log. <i>mar-</i> ~ <i>amar-</i> / <i>marók</i> ~ <i>amarok</i>
40. heart	* <i>zrd-</i>	<i>zoṛ</i>	Kan. <i>zli</i> Log. <i>zle</i> ~ <i>zli</i>
41. horn	* <i>cr-ū-</i>	<i>šax</i> -1 Pers.	Kan. <i>sukár</i>
41. horn <sub>2</sub>		<i>šī</i>	
42. I	* <i>aš-am</i>	<i>ān</i>	Kan. Log. <i>az</i>
42. I <sub>2</sub> (obl.)	<i>ma-</i>	<i>mon</i>	Kan. Log. <i>mun</i>
43. kill	* <i>gan-</i>	<i>mēr-</i> / <i>mānt-</i>	Kan. <i>wazn-</i> / <i>wáz-yok</i> Log. <i>wožn-áw-</i> / <i>wožn-aw-ók</i>
43. kill <sub>2</sub>	* <i>kauš-</i> SW NW		
44. knee	* <i>žān-u-</i>	<i>žānú</i> -1 Pers.	Kan. <i>zan-řak</i> ~ <i>zān-řak</i>
44. knee <sub>2</sub>			Log. <i>žānú</i> { <i>žā'nū</i> } -1 Pers.
45. know	<i>wid-</i>	<i>xabár bi-</i> -1 Pers.	Kan. <i>pazán-</i> / <i>pazan-yék</i> Log. <i>pazán-</i> / <i>pazan-ók</i>
45. know <sub>2</sub>			Kan. <i>poy awas-</i> Log. <i>poy b-</i> / <i>poy šáw-</i> -1 Psht.
46. leaf	* <i>war-ka-</i>	<i>pōn</i>	Kan. <i>puxay</i>
46. leaf			Log. <i>barg</i> -1 Pers.
47. lie	* <i>cai-</i>	<i>řiz-</i>	Kan. <i>nw-</i> / <i>nw-ástak</i>
48. liver	* <i>yak-ar</i> / * <i>yak-n-</i>	<i>žegár</i> -1 Pers.	Kan. <i>dzāř</i>
49. long	* <i>darg-a-</i>	<i>bišř</i>	Kan. <i>drāy</i> Log. <i>drāy</i>
50. louse	* <i>cwiš-</i>	<i>espó</i>	Kan. Log. <i>spuy</i>
51. man	* <i>nar-</i>	<i>mēr</i>	Kan. Log. <i>sařay</i> -1 Psht.
52. many	* <i>par-u-</i>	<i>γolú</i> -1 Pers. < Arab.	Kan. <i>dzut</i> ~ <i>zut</i> Log. <i>zot</i>
53. meat	* <i>gaw-</i>	<i>γuš</i>	Kan. <i>gáka</i> Log. <i>gáka</i>
54. moon	* <i>māh-</i>	<i>mahók</i>	Kan. <i>maryok</i>
54. moon <sub>2</sub>			Kan. <i>spožmay</i> -1 Psht.
54. moon <sub>3</sub>			Log. <i>mátaw</i> -1 Pers. ?
55. mountain	* <i>gar-i-</i>	<i>d'ár</i> -1 Dard. / Ind.	Kan. <i>gri</i> Log. <i>gri</i> ~ <i>giri</i>
55. mountain <sub>2</sub>	* <i>kaufa-</i>		
56. mouth	* <i>āh-</i>	<i>šoṇḍ</i> -1 Dard.	Kan. <i>pyoz</i> Log. <i>poz</i>
56. mouth <sub>2</sub>	* <i>žaf-an-</i> SW NW		
57. name	* <i>nām-an-</i>	<i>nām</i>	Kan. <i>nām</i> Log. <i>nām</i>
58. neck	* <i>grīw-a-</i>	<i>mandá</i> -1 Dard./Ind.	Kan. <i>gařdan</i> Log. <i>gardan</i> -1 Pers.
59. new	* <i>nav-a-</i>	<i>now</i>	Kan. <i>nyow</i> Log. <i>now</i>
60. night	* <i>xšap-</i>	<i>xawán</i>	Kan. <i>šyo</i> Log. <i>šo</i>
61. nose	* <i>nah-</i> / * <i>nāh-</i>	<i>nēšř</i>	Kan. Log. <i>ninní</i>
61. nose <sub>2</sub>	* <i>wain-a-</i> ~ * <i>wain-i-</i> SW NW		
62. not	* <i>na-</i>	<i>na</i>	Kan. Log. <i>nak</i>
63. one	* <i>ai-wa-</i>	<i>žu</i>	Kan. <i>sa</i> Log. <i>še</i>
64. person	* <i>mar-t-ya-</i> / * <i>mar-ta-</i>	<i>mānéš</i> -1 Ind.	Kan. Log. <i>sařay</i> -1 Psht.



Word	Proto-Iranian	Parachi	Ormuri
65. rain	*wār(a)-	γār	Kan. <i>bārān</i> Log. <i>bārān</i> -1 Pers.
66. red	*raud-i-ta-	sork ~ sórkō	Kan. <i>suř</i> Log. <i>šuš</i>
67. road	*pant-ā- ~ *paθ-	panán	Kan. <i>rāy</i> , Log. <i>rāy</i>
68. root	*waix-a-	γix	Kan. <i>walyi</i> {walye} -1 Psht.
68. root	*raiš-a-ka-		
69. round	*gart-a-na-	not attested	not attested
70. sand	*raika- SW NW	<i>seyá</i> -1 Ind. / Dard. / Ir. ?	Kan. <i>sigá</i> -1 Psht.
71. say	*mraw-	jař- / jaři:-	Kan. <i>γ<sup>w</sup>as-</i> Log. <i>γoš-</i>
71. say <sub>2</sub> (suppl.)	*wač-		
71. say <sub>3</sub>	*gaub- SW NW		
72. see	*wain-	b'óc- -1 Dard. / Ind.	Kan. <i>dzun-</i>
72. see <sub>2</sub>			Log. <i>juč-</i>
72. see <sub>3</sub> (suppl.)	*darc-	d'or-	Kan. <i>dyeK</i> Log. <i>dek</i>
73. seed	*taux-man-	p'ōř- -1 Dard. / Ind.	Log. <i>toxəm</i> ~ <i>toxom</i> -1 Pers.
74. sit	*āh-	n'in- / n'ašt- / n'ōšt	Kan. <i>āy-</i> ~ <i>ey-</i> / <i>nás-tak</i> Log. <i>nós-tok</i>
75. skin	*pawasta-	pust -1 Pers.	Kan. <i>tsarmun</i> / <i>čāmra</i> -1 Psht.
75. skin <sub>2</sub>			Log. <i>pōst</i> -1 Pers.
76. sleep	*hwap-	xōm kan-	Kan. <i>x<sup>w</sup>āw k-</i> / <i>dok</i> Log. <i>xaw k-</i> / <i>dāk</i> Kan. <i>x<sup>w</sup>al-gastak</i>
77. small	*kac-u-	činó -1 Dard.	Kan. Log. <i>zarí</i>
77. small <sub>2</sub>		ríza -1 Pers.	
78. smoke	*dū-ta- / *dū-ma-	d'i	Kan. <i>lugay</i> -1 Psht.
78. smoke			Log. <i>dud</i> {dūd} -1 Pers.
79. stand	*stā-	papá bi-	Kan. <i>ašt-</i> / <i>ašt-yék</i>
79. stand <sub>2</sub>			Kan. <i>t-yek</i>
79. stand <sub>3</sub>			Log. <i>dar</i> / <i>dar-ók</i>
80. star	*stār-	estéč	Kan. <i>stirrak</i> ~ <i>storak</i> Log. <i>stežáy</i> ~ <i>stéšak</i>
81. stone	*ac-man- / *ac-an-ga-	gīr	Kan. Log. <i>gap</i>
82. sun	*hwar- / *hwan-	ruč	Kan. <i>meř</i> Log. <i>meš</i>
82. sun <sub>2</sub>			Log. <i>tówa</i> Kan. <i>tōwō</i>
83. swim	*fraw-	aubāzī kan- -1 Pers.	Log. <i>aubāzī k-</i> -1 Pers.
84. tail	*dum-a- / *dum-b-a-	domb -1 Pers ?	Kan. <i>likye</i> -1 Psht.
84. tail		dayónd	Log. <i>dom</i>
85. that	*ana- 'that' (medial) / *haw- (masc., fem.) ~ *awa- (neut.) 'that' (distal)	ō	Kan. <i>afa</i> ~ <i>afo</i> Log. <i>afo</i>
86. this	*ai- (masc.) ~ *i- (fem., neut.) 'this'	ē	Kan. <i>o</i> Log. <i>a</i>
87. thou	*tū- ~ *tuw-am	tō	Kan. Log. <i>tu</i>

Word	Proto-Iranian	Parachi	Ormuri
88. tongue	*hiz-ū-	bān	Kan. zbān Log. zobān {zu'bān}
89. tooth	*dant- ~ *dant-an-	danān	Kan. gas Log. gīši (sg. and pl.)
90. tree	*dār-u-	b'in	Kan. wūna -1 Psht.
90. tree <sub>2</sub>			Log. draxt -1 Pers.
91. two	*dwa-	di ~ do	Kan. dyo Log. do
92. walk (go)	*ai-	par-	Kan. tsaw- / ts-yek Log. tsaw- / tsaw-ok
92. walk (go) <sub>2</sub> (suppl.)		č'i-	
93. warm (hot)	*gar-ma-	ta'pō -1 Dard.	Kan. Log. tok
93. warm (hot) <sub>2</sub>	*tap- EIr		
94. water	*āp- / *ap-	āw	Kan. wak Log. wok
95. we	*wai-am	mā	Kan. māx Log. māx
95. we <sub>2</sub> (suppl.)	*ah-ma-		
96. what	*ka-t	če	Kan. Log. tsa
97. white	*cwait-a- ~ *cwit-a-	č'áčō -1 Dard. / Ind.	Kan. spew ~ spiw Log. spew
98. who	*ka-	ka	Kan. kuk Log. kok
99. woman	*žan-i- ~ *žan-ī-	zaif -1 Pers. < Arab.	Kan. dzarká Log. zarká
100. yellow	*zar-i-	zítō	Kan. Log. zyer -1 Psht.
101. far	*dūr-ai	dorín	Kan. pets
101. far <sub>2</sub>			Log. dur -1 Pers. ?
102. heavy	*gur-u-	gerān -1 Pers.	Kan. grān
102. heavy <sub>2</sub>		b'eān -1 Ind.	Log. wazmīnd -1 Pers.
103. near	*nazd-	nazdík -1 Pers.	Kan. Log. boy
103. near <sub>2</sub>			Log. nezdék
104. salt	*namada-ka- ~ *namad-ka-	namá	Kan. mek Log. nemék
104. salt	*hwād- NW		
105. short	*mr̥z-u-	γánōk	Kan. Log. lanđ -1 Psht.
106. snake	*až-i-	kerm -1 Pers.	Kan. māngor -1 Psht.
106. snake <sub>2</sub>	*mār-a- SW NW		Log. mār -1 Pers.
106. snake <sub>3</sub>	*δəvũγc- (Shughni)		
107. thin	*tan-u-ka-	mahín -1 Pers.	Kan. narye Log. naráy -1 Psht.
107. thin <sub>2</sub>	*nāz-u-ka-		
108. wind	*wā-ta-	γā {γâ}	Kan. bād -1 Pers.
108. wind <sub>2</sub>		šamāl -1 Pers. < Arab.	
109. worm	*kr̥m-i-	kerm -1 Pers.	Log. kerm -1 Pers.
110. year	*yār-	sar	Kan. tsān Log. čān

### Notes on specific Swadesh items.

1. 'all': **Proto-Iranian:** \*wic-w-a- is the best candidate: its descendants mean 'all' in Avestan and in some Middle Iranian languages (like Sogdian). In many Iranian languages (both Western and Eastern) \*wic-w-a- was superseded by the stem \*ham-a- with the original meaning 'the same'; in Parachi and Ormuri it has been replaced by reflexes of \*harw-a- 'whole (totus)'.

3. 'bark': **Proto-Iranian:** *\*pawasta-* is likely to be analyzed as the past participle *\*pa=was-ta-* from the verbal root *\*was-* 'to wear clothes, be dressed' with the prefix *\*(a)pa-*. The first meaning of this form is 'skin'; polysemy 'skin / bark' is found in many Iranian languages and can be reconstructed for the proto-language.

4. 'belly': **Proto-Iranian:** *\*udar-a-* is retained in Avestan, Sogdian and in many Eastern Iranian languages. The choice is further supported by Indo-Aryan cognates. *\*škamb-* is a Southwestern isogloss; *\*lab-* is a Northwestern isogloss.

**Parachi:** Probably of Indo-Iranian origin.

5. 'big': **Proto-Iranian:** The root *\*ma3-* and its derivatives are retained in Avestan, Sogdian, Southwestern, Northwestern (Kurdish *mazin*, Balochi *mazan*) and in some Eastern Iranian languages (for example, Manichean Sogdian *mzyγ < \*ma3-ik-a*, a secondary formation of the comparative grade *\*ma3-iyah*). It has a reliable Indo-European etymology and exact matches in Old Indic. *\*stūra-* is attested in this meaning in Eastern Iranian languages.

6. 'bird': **Proto-Iranian:** The stem *\*wi-* is retained in Avestan and Middle Persian and is supported by Vedic evidence. In many Iranian languages, 'bird' is expressed with reflexes of *\*mṛga-*.

**Parachi:** It is possible to suggest that the main word for 'bird' is *morčē* 'sparrow' (Efimov 2009: 218; Morgenstierne 1929: 273) (like in Ormuri, see below), but this is uncertain.

**Ormuri:** Though this word is usually translated as 'sparrow / little bird', Hallberg's informants use it to translate phrases "the man killed (shot) the bird" and "the bird flew" (Hallberg 2004: 141). It means that Ormuri shows polysemy 'sparrow / little bird / bird in general'.

7. 'bite': **Proto-Iranian:** The stem *\*danc-* is retained in Avestan and Khotanese Saka and is supported by Indo-Aryan cognates. *\*ga3-* is attested in the meaning 'to bite / to sting' in Southwestern and Northwestern Iranian languages.

8. 'black': **Parachi:** Etymology is unknown. Looks like a borrowing. Phonetic variants: *pádu*, *p'adú* (Efimov 2009: 225).

9. 'blood': **Ormuri:** Morgenstierne thinks that Kan. *sun* Log. *šun* is borrowed from Indic languages; he compares Sanskrit *śóna-* 'red, crimson, purple / redness / blood'. This is rather uncertain, because this word has later Indic cognates only in Pali *sōṇita* and Prakrit *sōṇia* (Turner 1962–1985: 12625). The correspondence Kan *s* ~ Log. *š*, usual for genuine Ormuri words, is also problematic. I would rather hesitantly derive Ormuri *sun* from PI *\*xrū-* 'raw flesh' with the suffix *-na-*, compare Avestan *xrū-* 'raw flesh', *xrūniia-* 'Bluttat, blutige Mishandlung' (Bartholomae 1904: 539). Log. *in* is known only from one Logar informant. Morgenstierne considers this word to be a borrowing from Pashto *wín*; indeed, there are dialectal forms without *w-* (Hallberg 2004: 22).

11. 'breast': **Proto-Iranian:** *\*war-ah-*, found in Avestan and some other languages (for example, Balochi), has a parallel in Old Indian *úras-* 'breast'. *\*cai-na-* is attested in Southwestern and Northwestern languages.

**Ormuri:** Log. *sónok* ~ *sínak* shows polysemy 'breast / udder'. A dubious case: this word can designate only woman's breast. Distinct from Kan. Log. *čik* pl. *čiči* 'nipple / female breast (pl.)' (Efimov 2011: 270; Morgenstierne 1929: 390). Hallberg cites Kan. *sina* that seems to be borrowed from Persian (Hallberg 2004: 83).

12. 'burn': **Proto-Iranian:** *\*dag-* is retained in this or slightly altered meaning in Avestan, middle Iranian and some modern Iranian languages. The second candidate *\*sauk-* / *\*sauč-*, widespread in Iranian languages, has the original meaning 'to shine'.

*\*θaw-* is an Eastern isogloss (mostly Pamir languages).

14. 'cloud': **Proto-Iranian:** There are two Iranian stems, *\*abr-a-* and *\*maig-a-*, both attested in Avestan and all Iranian subgroups and having external cognates. It seems that the main word for 'cloud' in Young Avestan was *maēya-*, while *aβra-* had the meaning 'rain cloud'.

However, in other Iranian languages descendants of both these stems can mean ‘cloud’ in general. Therefore, it is reasonable to think that *\*abra-* and *\*maiga-* were semantically very close to each other in Proto-Iranian, and we have to treat them as synonyms.

15. ‘cold’ (adj.): **Proto-Iranian:** *\*stāba-* is an Eastern Iranian isogloss.

18. ‘dog’: **Proto-Iranian:** *\*cwan-* has Indo-Aryan cognates, a perfect Indo-European etymology and wide distribution. *\*kuta-* / fem. *\*kuti(ī)-* is widespread mostly in eastern Iranian languages and has a weak external etymology. Avestan *gaδuuu-* designates either a dog of a special breed or serves as a stylistic synonym; descendants of this root in other Iranian languages often mean ‘puppy’ (Ossetic, Yaghnobi) or ‘a multicolored dog; a dog with a white mark’ (Wakhi), ‘a dog with cut ears’ (Shughni).

19. ‘drink’: **Proto-Iranian:** *\*hwar-* features polysemy ‘to eat / to drink’ in Avestan, Southwestern languages (Middle Persian, Modern Persian etc.), Northwestern languages (Balochi), Eastern languages (Sogdian, Yaghnobi), Ormuri. Such distribution suggests that *\*hwar-* could be polysemous already in Proto-Iranian.

The second candidate is the root *\*pī-* / *\*pā-* whose antiquity is proved by Indo-Aryan cognates. This root forms the verbs ‘to drink’ in some languages of the Pamir group (Wakhi *puv-*, *pəv-* / *pit-*, Ishkashimi *pəv-* / *pəvd-*, Sanglechi *pöv-* / *pöv-* ‘to drink’). At the current stage of research it is hard to determine whether the Pamir languages possess a retention or represent a backward development.

22. ‘earth’: **Ormuri:** Log. *búma* ~ *bóma*, Kan. *búma* ~ *bumba* mostly mean ‘earth as opposed to heaven / ground’, but can also designate ‘earth (soil)’, compare “after this the seeds are sown (lit. ‘scattered on the earth’)” (Efimov 2011: 234).

G. Morgenstierne translates Log. *jer* as ‘clay’; moreover, it can be tentatively traced back to *\*grai-* / *\*gri-* ‘clay’ (EDIL 3: 83). However, it is used in the following contexts: “then for one month (they are occupied) with ditches and channels (to retain water), they take earth (soil, clay), they make vegetable patches (a kord is a plot of ground banked up) – they put earth on the plot”; “the month of Jawza has come – they take earth, put it on the plots, they water them in the spring” (Efimov 2011: 250).

26. ‘fat’: **Proto-Iranian:** *\*pī-wah-* is retained in Avestan and many middle and modern Iranian languages from all Iranian subgroups, and has reliable Indo-Aryan and Indo-European cognates. *\*čarp-īh-* is a Western Iranian isogloss. *\*wāzdā-* is an Eastern Iranian isogloss.

29. ‘fish’: **Proto-Iranian:** the word *\*macy-a-* (< *\*matsy-a-*) is retained in Avestan, Southwestern (middle and modern Persian), Northwestern Iranian languages (Parthian, Kurdish etc.) and in dialects of Pashto; moreover, it has clear Indo-Aryan cognates.

Eastern Iranian *\*kapa-* ‘fish’ does not occur in other subgroups and has no reliable etymology.

30. ‘fly’ (v.): **Proto-Iranian:** *\*pat-* still preserves its original meaning in several Avestan passages, the direct Indo-Aryan cognates speak in favor of its antiquity. In the majority of Iranian languages it was superseded by verbs with different semantics (often prefixed) or denominatives from *\*par-na-* ‘feather’.

31. ‘foot’: **Parachi:** Polysemy: ‘foot / paw’. It is possible to suspect polysemy ‘foot / leg’, compare ‘this woman is pregnant’, literally, ‘this woman has heavy legs’ (Efimov 2009: 160).

**Ormuri:** Polysemy: ‘foot / leg’. The Kaniguram form has an Indian suffix. Hallberg records *lanḡay* {lanḡa<sup>i</sup>} (Hallberg 2004: 85).

34. ‘good’: **Proto-Iranian:** *\*wah-u-* is retained in Avestan and some other Iranian languages (among them, Khotanese); it has direct Indo-Aryan and Indo-European cognates.

35. ‘green’: **Proto-Iranian:** *\*zar-i-* shows polysemy ‘green / yellow’; these meanings are found in Avestan and coincide with the Old Indian situation. In the majority of Iranian lan-

guages, this adjective with the meaning ‘green’ was superseded by *\*s abz-a-* ‘green’ of unclear origin. *\*axšaina-* is an Eastern isogloss.

36. ‘hair’: **Proto-Iranian:** *\*warc-a-* is retained in this meaning in Avestan, in Northwestern Iranian Languages (Pahlavi) and in several Eastern Iranian idioms of different periods (for example, Khotanese, Sogdian and Pashto). In Western Iranian it has altered meanings, e.g., Persian *gors* ‘curl’. *\*maud-a-* ‘hair’ is found only in Southwestern and Northwestern Iranian languages of Middle and Modern periods. The stem *\*gaun-a-* designating ‘hair’ in Eastern Iranian languages originally meant ‘body hair (of animals) / color (of horse)’.

39. ‘hear’: **Proto-Iranian:** *\*craw-* is retained in its basic meaning ‘to hear’ in Avestan, Balochi and Eastern Iranian languages (the Shughni group and Yazghulami); other Iranian languages of all subgroups preserve only derived meanings, like ‘to sing’, ‘to be heard’. Verbs with the root *\*gawš-* (mostly prefixed) are also attested as ‘to hear’ in many Iranian languages belonging to all subgroups. However, Avestan *gaoš-* means ‘to listen / to perceive’ and Old Indian *ghoṣ-* means ‘to sound’. Therefore, it is unnecessary to treat *\*craw-* and *\*gauš-* as Proto-Iranian synonyms.

43. ‘kill’: **Proto-Iranian:** the root *\*gan-* shows polysemy ‘to hit / to strike / to slay / to kill’. Retained in Avestan and all Iranian subgroups; has direct Indo-Aryan and IE cognates. In many Iranian languages, it preserves only the meanings ‘to hit / to strike’, but it is the main word for ‘to kill’ in Avestan, Parachi, Ormuri and many Eastern Iranian languages (Khotanese, Shughni, Ishkashimi etc.). In Northwestern and Southwestern Iranian languages the main verb with this meaning is *\*kauš-*.

45. ‘know’: **Proto-Iranian:** The opposition *\*wid-* ‘to know (a situation), know that...’ / *\*zan-* ‘to know how to, be acquainted with an object/person’ is found in Avestan and can be reconstructed for Proto-Iranian (the same is attested in Old Indian). In the majority of Iranian languages, *\*wid-* ‘to know (a situation), know that...’ in its primary meaning was superseded by *\*zan-*. In some languages, the meaning ‘to know’ is expressed with such verbs as *\*grab-* ‘to grab / seize / take’ etc.

46. leaf: **Parachi:** *pōn* has the polysemy: ‘feather / leaf’.

**Ormuri:** Kan. *puxay* shows polysemy ‘leaf / feather’. Also Log. *paṭ* (Morgenstierne 1929: 389), probably borrowed from Dardic languages; compare Pashai *paṭā* ‘leaf’.

47. ‘lie’: **Proto-Iranian:** *\*cai-* is retained in Avestan and has external cognates. Descendants of this root in later Iranian languages have different meaning: ‘to sleep’ (Wakhi), ‘to be seek, ill’ (Ossetic). In the majority of Iranian languages the term ‘to lie’ is derived from verbs with original meanings ‘to sleep’, ‘to fall’ etc.

54. ‘moon’: **Ormuri:** Kan. *spožmay* (Efimov 2011: 272; Hallberg 2004: 92). It is clear from texts in Efimov 2011 and from Hallberg’s list that now this borrowing from Pashto has superseded the original *maryok* in Kaniguram dialect.

55. ‘mountain’: **Proto-Iranian:** *\*gar-i-* is retained in this meaning in Avestan, Southwestern, Eastern languages and Ormuri. In Northwestern languages this root is preserved only in altered meanings (compare Kurdish *gir, girik* ‘hill / height’).

*\*kaufa-* has reflexes in Persian and Kurdish and is a Western isogloss.

56. ‘mouth’: **Proto-Iranian:** *\*āh-* is retained only in Avestan and Khotanese, but its antiquity is proven by direct Indo-Aryan and IE cognates. *\*zaf-an-* is used with this meaning in Southwestern and Northwestern Iranian groups. The original meaning may have been ‘muzzle’ (in Avestan – ‘mouth of daevic creatures’).

58. ‘neck’: **Proto-Iranian:** *\*grīw-a-* is retained in Avestan, Southwestern and Eastern Iranian languages, its antiquity is supported by direct Indo-Aryan cognates. The widespread stem *\*gard-a-na-* ~ *\*gart-a-na-* ‘round’ with the semantic development ‘round’ > ‘neck’ represents a later derivative from *\*gart-* ~ *\*gard-* ‘to turn’.

61. 'nose': **Proto-Iranian:** *\*nah-* / *\*nāh-* is retained in Avestan, Old Persian and all subgroups and has direct Indo-Aryan cognates. The stem *\*wain-a-* ~ *\*wain-i-*, derived from *\*wain-* 'to see', means 'nose' in some Iranian languages mostly of Southwestern and Northwestern subgroups. In Avestan it means 'nose of dragon'. We can treat it as a Western isogloss.

**Ormuri:** Kan. Log. *ninní* can represent a contamination of *\*nah-* / *\*na:h-* and *\*wain-i-*.

66. 'red': **Proto-Iranian:** *\*raud-i-ta-* is retained in Avestan and has direct Indo-Aryan and IE cognates. The stem *\*suk-ra-* ~ *\*sux-ra-*, derived from the root *\*sauč-* 'to shine', is widespread in Iranian languages with the meaning 'red', but its original semantics should be 'shining, bright; red (of fire)'.

67. 'road': **Proto-Iranian:** *\*pant-ā-* ~ *\*paθ-* is a basic term in Avestan and mostly in Eastern languages (in Middle and modern Southwestern and Northwestern languages it means specifically 'path / track / direction / advice, counsel'). Cf. the stem *\*raθ-ya*, which acquires the semantics 'road' in Western Iranian languages and Pashto, being derived from the word *\*raθa-* 'chariot'.

68. 'root': **Proto-Iranian:** both *\*waix-a-* and *\*raiš-a-ka-* are retained in all subgroups (sometimes they represent synonyms within one language). These stems do not have certain external cognates. We treat them as synonyms.

70. 'sand': **Proto-Iranian:** a very unstable term. One of the possible candidates is *\*rayka-*, attested in Southwestern and Northwestern Iranian languages.

71. 'say': **Proto-Iranian:** suppletive roots *\*mraw-* and *\*wač-* are the main terms with the meaning 'to say' in Avestan; the first one is used in present, the second one mostly in the aorist and perfect. Old Indian data confirm that Avestan retains the original suppletive paradigm. *\*gaub-* in this meaning represents a Western isogloss (EDIL 3: 232).

72. 'see': **Proto-Iranian:** *\*wain-* forms the present stem. *\*darc-* is used as perfect and aorist stem in Avestan and as past stem in Parachi and some Eastern Iranian languages. The third stem *\*dāi-* usually forms the past stem in all subgroups, but Indo-Aryan cognates demonstrate another meaning.

Both Parachi and Ormuri have suppletive paradigms with preservation of *\*darc-*. The stem *\*wain-* yields Kan. *dzun-* (Efimov 2011: 283).

74. 'sit': **Proto-Iranian:** according to Avestan and external Indo-Aryan data, Proto-Iranian *\*āh-* meant 'to sit' and *\*had-* meant 'to sit down'. *\*āh-* is retained for the stative meaning 'to sit' in Avestan and Middle Iranian languages. In the majority of Iranian languages, belonging to all subgroups, *\*had-* acquired the meaning 'to sit', having superseded *\*āh-*.

**Ormuri:** Compare the phrase {a-dúwa-wa mux-ki āyi aw řawa bu} 'her daughter sits in front of her and cries'. The present stem *āy-* ~ *ey-* is distinct from related Kan. *n-* (2<sup>nd</sup> conj.) ~ *nay-* Log. *ney-* 'to sit down' with the same participle Kan. *nás-tak* Log. *nós-tok* (Efimov 2011: 291–293; Morgenstierne 1929: 401).

83. 'swim': **Proto-Iranian:** the verb *\*fraw-* is retained in the meaning 'to swim' only in some Avestan texts, but direct Indo-Aryan and IE cognates prove its antiquity. In later Iranian languages, it was superseded by various other verbs or analytic expressions.

85. 'that', 86. 'this': **Proto-Iranian:** On the basis of Avestan data, the Proto-Iranian system can be tentatively reconstructed as *\*ai-* (masc.) ~ *\*i-* (fem., neut.) 'this' / *\*ana-* 'that' (medial) / *\*haw-* (masc., fem.) ~ *\*awa-* (neut.) 'that' (distal).

90. 'tree': **Proto-Iranian:** *\*dār-u-* 'tree / wood / stick' is retained in all subgroups. It or its derivatives mean 'tree' in Southwestern and Northwestern Iranian languages. In light of external comparanda, *\*dār-u-* can be safely posited as a Proto-Iranian term for 'tree'. Thus it is likely that *\*dār-u-* in the meaning 'tree' was superseded with *\*wan-a-* 'forest / (forest) tree' already in Avestan and in the Eastern Iranian group.

92. 'walk (go)': **Proto-Iranian:** the root *\*ai-* is retained in Avestan and all subgroups. It forms both present and aorist stems in Avestan. In many languages of all subgroups descendants of *\*čyaw-* 'to set in motion; to move' became the main verbs with the meaning 'to go'.

93. 'warm (hot)': **Proto-Iranian:** *\*gar-ma-* 'hot / warm' is retained in Avestan and all subgroups; it has direct Indo-Aryan and IE cognates. *\*tap-* forms words with this meaning in Eastern Iranian languages.

**Ormuri:** Kan. Log. *tok* shows polysemy 'hot / warm'.

104. 'salt': **Proto-Iranian:** Kurdish *xwē* 'salt' and Balochi *vād* are derived from *\*hwād-* 'delicious; to be delicious'.

105. 'short': **Proto-Iranian:** *\*mr̥z-u-* is retained in Avestan (in the temporal sense), Khotanese and Sogdian. External comparanda prove its antiquity.

106. 'snake': **Proto-Iranian:** *\*aš-i-* is retained in Avestan, in Middle Persian in the meaning 'dragon', in Khwarezmian and Munji; this stem has Indo-Aryan and IE cognates and this situation proves that it should be considered as the main Iranian word for 'snake'. *\*mār-a-*, probably derived from *\*mar-* 'to kill', serves as a designation of 'snake' in Southwestern and Northwestern Iranian languages. In many Iranian languages, especially Eastern, descendants of *\*k̥m-i-* 'worm' altered the meaning to 'snake' (or the polysemy 'worm / snake' can be found). Eastern languages also have one more stem whose reconstruction is uncertain, yielding Shughni *divūsk*, Sarikoli *tv̄f̄v̄isk*, Wakhi *fuks* 'snake' etc.

107. 'thin': **Proto-Iranian:** *\*nāz-u-ka-*, originally meaning 'tender', replaced this meaning in some Southwestern and Northwestern languages.

110. 'year': **Proto-Iranian:** according to Avestan data and IE cognates, *\*yār-* meant 'year' in Proto-Iranian. *\*car-da-* with the original Proto-Indo-Iranian meaning 'autumn' alters it to 'year (of somebody's age)' in Avestan and to 'year (in general)' in later Iranian languages.

#### Abbreviations of language names

Arab. – Arabic; Dard. – Dardic; Elr – Eastern Iranian; Ind. – Indian; Ir. – Iranian; Kan. – Kaniguram Ormuri; Log. – Logar Ormuri; OInd. – Old Indian; Pers. – Persian (Dari); Psht. – Pashto; SW – Southwestern Iranian; NW – Northwestern Iranian; PIr – Proto-Iranian.

#### References

- Bartholomae, Christian. 1904. *Altiranisches Wörterbuch*. Strassburg: Verlag von Karl. J. Trubner.
- Baranov, Kh. K. 1977 *Arabsko-russkij slovar'*. 5th edition. Moskva: Russkij jazyk.
- Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb*. Leiden-Boston: Brill.
- Grierson, George A. 1918. The Ōrmuṛī or Bargistā language. An account of a little-known Eranian dialect. *Memoirs of the Asiatic Society of Bengal* 7/1: I-XIV, 1–101.
- Dybo, Anna V., George S. Starostin. In Defense of the Comparative Method, or the End of the Vovin Controversy. In: I. Smirnov (ed.). *Aspekty komparativistiki III. Orientalia et Classica: Trudy Instituta vostochnykh kul'tur i antichnosti XIX*: 119–258. Moskva: RGGU.
- Dybo, V. A. 1974. Afganskoe udarenie i ego znachenie dl'a indoevropskoj i balto-slav'anskoj akcentologii. I. Imen'naja akcentuacija. In: T. M. Sudnik (ed.). *Balto-slav'anskije issledovanija*: 67–105. Moskva: Nauka.
- Dybo, V. A. 1989. Afganskoe udarenie i ego znachenie dl'a indoevropskoj i balto-slav'anskoj akcentologii. II. Glagol'naja akcentuacija. In: N. I. Tolstoj (ed.). *Slav'anskij I balkanskij fol'klor. Prosodija: Sbornik statej*: 106–147. Moskva: Nauka.
- EDIL = V. S. Rastorgueva, D. I. Edelman. 2000-. *Etimologičeskij slovar' iranskich jazykov* (Etymological dictionary of the Iranian languages). Moskva: Vostochnaja literatura RAN.

- Efimov, V. A. 1986. *Jazyk ormuri v sinkhronnom i istoricheskom osveschenii*. Moskva: Nauka.
- Efimov, V. A. 2009. *Jazyk parachi*. Moskva: Vostochnaja literatura.
- Efimov, V. A. 2011. *The Ormuri language in Past and Present. English translation edited by Joan L. G. Baart*. FLI language and culture series (Volume 6). Forum for Language Initiatives.
- Hallberg, Daniel G. 2004. *Sociolinguistic survey of Northern Pakistan. Volume 4. Pashto. Waneci. Ormuri. 2<sup>nd</sup> edition*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.
- Kieffer, Charles M. 1989. Le parāči, l'ormuři. In: R. Schmitt (Hrsg.) *Compendium linguarum Iranicarum*: 445–455. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Kieffer, Charles M. 2009. Parachi. In: G. Windfuhr (ed.). *The Iranian languages*: 693–720. London / New York: Routledge.
- Mayrhofer, Manfred. 1989. Vorgeschichte der iranischen Sprachen; Uriranisch. In: R. Schmitt (Hrsg.) *Compendium linguarum Iranicarum*: 445–455. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Morgenstierne, Georg. 1926. *Report on a linguistic mission to Afghanistan*. Oslo: H Aschehoug & Co.
- Morgenstierne, Georg. 1929. *Indo-Iranian frontier languages. Volume I. Parachi and Ormuri*. Oslo: H Aschehoug & Co.
- Morgenstierne, Georg. 1974. *Etymological Vocabulary of the Shughni group*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Morgenstierne, Georg. 2003. *A new etymological vocabulary of Pashto. By Georg Morgenstierne; compiled and edited by J. Elfenbein, D.N. MacKenzie and Nicholas Sims-Williams*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Oranskij, I. M. 1979. Vvedenie. Voprosy istoriko-dialektologicheskoi klassifikacii iranskikh jazykov. In: V. S. Rastorgueva (ed.). *Osnovy iranskogo jazykoznanija. Drevneiranskije jazyki*: 81–128. Moskva: Nauka.
- Rees, Daniel A. 2008. *Towards Proto-Persian: An Optimality Theoretic Historical Reconstruction* (Ph.D.). Washington, DC.
- Starostin, George S. 2016. From wordlists to proto-wordlists: reconstruction as 'optimal selection'. *Faits de langues* 47: 177–200.
- Starostin, Sergei (ed.). 1998–2005. *The Tower of Babel. An etymological database project*. Available at: <http://starling.rinet.ru/> [accessed 26.12.2018].
- Tedesco, Paul. 1921. Dialektologie der west-iranischen Turfantexte. *Le Monde oriental* 15: 184–258.
- Turner, Ralph L. 1962–1985. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press. Available at: <http://dsal.uchicago.edu/dictionaries/soas/> [accessed 26.12.2018].

A. A. Трофимов. О классификации языков парачи и ормури внутри иранской группы по данным аннотированных сводешевских списков

В данной статье рассматриваются языки парачи и ормури и их место в иранской классификации. На основании данных сопоставления двух аннотированных 110-словных списков Сводеша, составленных для этих языков, с реконструированным праиранским списком и западноиранскими и восточноиранскими регионализмами, делается вывод о более вероятном восточноиранском характере парачи и ормури – эти языки имеют совпадения с восточноиранскими изоглоссами и практически не имеют совпадений с многочисленными западноиранскими. При этом показательно, что фонетические изоглоссы не дают ясной картины: у парачи и ормури есть как западные, так и восточные фонетические черты, причём интересно, что акцентуационные системы этих языков напоминают систему, засвидетельствованную в пушту и в наибольшей степени отражающую древнее (индо)иранское состояние. Таким образом, более обоснованно мнение Г. Morgenstierne о принадлежности парачи и ормури к восточноиранским языкам. При этом выделение норвежским исследователем отдельной юго-восточной иранской подгруппы, единственными представителями которой и являются рассматриваемые языки, спорно и нуждается в дальнейшем подтверждении.

*Ключевые слова:* парачи, ормури, 110-словный список Сводеша, праиранский язык, восточноиранские языки, западноиранские языки, фонетические изоглоссы, акцентуационные системы, пушту.



## О языке поэмы *Bāṇāsarakathā* (XV в.) и его отношении к языку кашмири

Язык поэмы *Bāṇāsarakathā* («Рассказ о демоне Бане»), написанной в Кашмире в середине XV в., традиционно считается ранней формой кашмири. Ряд фактов исторической фонетики и лексики, однако, заставляет считать эту точку зрения сомнительной. В статье делается попытка внести в данный вопрос максимальную ясность. Автор приходит к выводу, что язык основной части поэмы должен классифицироваться как индоарийский и, при этом, несомненно, носит искусственный и стилизованный характер. Вместе с тем в тексте содержатся многочисленные фрагменты на раннем кашмири, которые следует считать первой дошедшей до нас датированной фиксацией дардского языкового материала. Ряд засвидетельствованных в памятнике индийских элементов, вероятнее всего, был заимствован из новоиндийского языка, распространенного в Кашмирской долине до XIII–XIV вв.

*Ключевые слова:* язык кашмири, дардские языки, индоарийские языки, пракриты, апабхрэнша, искусственные языки, языковой сдвиг.

Хотя кашмири до недавнего времени являлся единственным дардским языком, имеющим письменность, диахроническое исследование кашмирского материала с помощью филологического метода как правило оказывается малорезультативным. Причина тому – непродолжительность литературной традиции: самые ранние из дошедших до нас в оригинальном виде текстов отстоят от наших дней не более чем на три столетия. В подобной ситуации представляются вполне объяснимыми попытки некоторых ученых обнаружить более старые аутентичные литературные памятники на кашмири.

Одним из возможных старокашмирских памятников в прошлом считались стихотворные вставки в религиозно-философском трактате *Mahānayaṅprakāśa* (Grierson 1929)<sup>1</sup>. Проведенное нами специальное исследование, однако, показало, что язык этих вставок никоим образом не может являться предком современного кашмири. Он представляет собой искусственный индоарийский литературный идиом, произведенный из санскрита при помощи правил фонетического пересчета и содержащий лишь некоторое (притом относительно небольшое) количество заимствованных кашмирских элементов (Коган 2010).

Недавно в широкий научный обиход был введен еще один текст, язык которого, по мнению ряда исследователей, является ранней формой кашмири. Это поэма на сюжет из Махабхараты о битве Кришны с демоном Баной (Банасурой), известная под санскритским названием *Bāṇāsarakathā* («Рассказ о демоне Бане»). Автор поэмы – Аватарабхатта или Бхаттаватара – был одним из придворных поэтов кашмирского султана Зайн-уль-Абидина (1420–1470). Текст содержит прямое указание на дату написания – 26-ой год правления султана, что соответствует 1446 г. Единственная рукопись памятника была обнаружена во 2-ой половине XIX в. в г. Биканер<sup>2</sup> известным немецким санскритологом Георгом Бюлером. Попытка прочесть ее с помощью информантов из числа ученых

<sup>1</sup> Сколь-нибудь надежная датировка памятника отсутствует, известно лишь имя автора – Шитикантха.

<sup>2</sup> Ныне в штате Раджастан на западе Индии.

кашмирцев-индусов («кашмирских пандитов») успеха не имела<sup>3</sup>. Тем не менее, Г. Бюлер классифицировал ее язык как кашмири (Bühler 1877: 90), и в дальнейшем данная точка зрения нашла отражение в целом ряде кашмироведческих публикаций, где поэма *Bāṇāsarakathā* называется наиболее ранним датированным кашмирским литературным произведением (Grierson 1919; Эдельман 1965; Захарьин, Эдельман 1971; Edelman 1983). Название памятника было ошибочно прочитано Г. Бюлером как *Bāṇāsuravadha* («Убийство демона Баны»), а имя поэта не было распознано, из-за чего рассматриваемый текст иногда объявляют поэмой неизвестного автора (см., например, Захарьин, Эдельман 1971; Edelman 1983).

В течение весьма длительного времени после обнаружения (в общей сложности около 100 лет) рукопись памятника находилась в книгохранилищах нескольких индийских библиотек<sup>4</sup>. В конце 60-х годов XX в. она попала в поле зрения молодого филолога Шаши Шекхара Тошкхани, посвятившего описанию ее языка диссертацию на соискание степени PhD в Кашмирском университете<sup>5</sup>. Диссертация была защищена в 1972 г. на факультете языка хинди и написана на этом языке. Она содержит общие сведения о поэме, эпохе её написания и авторе, стилистических особенностях, грамматический очерк языка памятника, включающий также краткий обзор исторической фонетики, оригинальный текст с переводом на хинди, словарный указатель. Долгие годы исследование Ш.Ш. Тошкхани было практически недоступным для широкой научной общественности, причиной чему был не только языковой барьер, но и политическая напряженность в столице штата Джамму и Кашмир, где работа хранилась в единственном экземпляре. В декабре 2015 г. электронный вариант текста диссертации был опубликован в Интернете<sup>6</sup>, что впервые сделало возможным знакомство с ней ученых за пределами Индии. Это в свою очередь открыло путь для более детального изучения языка поэмы *Bāṇāsarakathā* в различных аспектах, в том числе, в сравнительно-историческом.

\* \* \*

В завершающей части поэмы автор называет язык своего произведения санскритским словом *dēśya* («местный, провинциальный, относящийся к данной области или стране»). По мнению Ш. Ш. Тошкхани, данный термин следует считать обозначением народно-разговорного языка Кашмира XV в. (Toṣkhānī 1972: 235). Мы видим, однако, некоторые основания полагать, что подобная интерпретация не является единственно возможной. Известно, например, что искусственный язык упоминавшегося выше трактата *Mahānāyaprakāśa* назывался синонимичным термином *dēśabhāṣā* («язык страны»)<sup>7</sup>. Данный факт, как нам представляется, указывает на необходимость осторожной и небуквальной трактовки лингвонимов с начальными компонентами *dēśa-* и *dēśya-*. Скорее всего, их следует считать названиями идиомов, применявшимися в той или иной функции только в определенных географических ареалах (в противоположность санскриту и литературным пракритам, использовавшимся во всей Индии), имея при этом в виду, что сфера функционирования таких идиомов часто неясна. В некоторых случаях речь вполне мо-

<sup>3</sup> По словам самого Г. Бюлера, информанты заявили, что способны понять текст поэмы, однако не берутся его объяснить (Bühler 1877: 90).

<sup>4</sup> В настоящее время она хранится в библиотеке Института востоковедных исследований им. Р. Г. Бхандаркара (Bhandarkar Oriental Research Institute) в г. Пуна.

<sup>5</sup> В г. Сринагар, столице штата Джамму и Кашмир.

<sup>6</sup> Toṣkhānī 1972. Режим доступа: <http://shodhganga.inflibnet.ac.in/handle/10603/56812>.

<sup>7</sup> См. Grierson 1929.

жет идти о местных литературных языках, не отражавших живую разговорную речь и носивших, подобно пракритам и апабхрэнша, стилизованный и искусственный характер

Сходство языка поэмы *Bāṇāsarakathā* с языком трактата *Mahānayaṅprakāśa* отнюдь не ограничивается близостью названий. Весьма примечательной общей чертой обоих текстов является очевидное преобладание в них индоарийской лексики<sup>8</sup>. Часть этой лексики не обнаруживает фонетических отличий от древнеиндийских прототипов и, таким образом, может представлять собой прямые заимствования из санскрита. Другая часть, несомненно, отражает более позднюю, чем древнеиндийская, ступень историко-фонетического развития. Показательно, что в целом ряде слов, относящихся к этой группе, обнаруживаются фонологические инновации, характерные для индоарийской ветви. Среди таких инноваций следует выделить две: стяжение общеиндоиранского краткого дифтонга *\*ai > ē* (*ēk-* ‘1’ при др.-инд. *ēka-* (< *\*aika-*), ср. авест. *aēva-*; *ēśu* ‘этот’ при др.-инд. *ēśa-* (< *\*aiśa-*), ср. авест. *aēšo*) и переход общеиндоиранской звонкой придыхательной аффрикаты *\*jh<sup>9</sup> > h* (*hanēta* ‘убив’ при др.-инд. *hanti* ‘убивает’ (< *\*jhanti*), ср. авест. *jaiṅtī*; *nēh-* ‘любовь, привязанность’ при др.-инд. *snēha-* ‘любовь, привязанность; липкость, клейкость, жир, слизь, мокрота’ (< *\*snaijha-*), ср. авест. *snaēžana-* ‘слюнявый (о ребенке, собаке, волке)’, *snaēžint* ‘снежащий’<sup>10</sup>; *duhit* ‘дочь’ при др.-инд. *duhitar-* (< *\*dujhitar-* < и.-е. *\*dhughater-*)).

Обращает на себя внимание значительное число примеров рефлексии ряда индоиранских консонантных сочетаний, широко распространенной в средне- и новоиндийских языках, но чуждой кашмири. К подобным примерам относятся, в частности, случаи утраты второго компонента в кластерах вида «согласный + r», т. е. переходы *\*tr > t* (*tās-* ‘страх’ при др.-инд. *trāsa-*, пали, пракр. *tāsa-* ‘дрожь, страх’, хинди *tāsnā* ‘ужасать, угрожать’; *ti-* ‘три’ при др.-инд. *tri-* *trayaḥ*, *trīṇi*, пали *tayo*, *tīṇi*, пракр. *taō*, *tiṅṅi*, хинди, мар. *tīn*, пандж. *tinn*, неп., бенг. *tin*, кашм. *trē*), *\*dr > d, dd* (*canda*<sup>11</sup> ‘луна’ при др.-инд. *candra-*, пали *canda-*, пракр. *caṅda-*, хинди, гудж., мар. *cād*, пандж. *cand*; *niddi* ‘спит’ при др.-инд. *nidrāyati*, пали *niddāyati*, пракр. *ṅiddāadi*, ср. также пали *niddā*, пракр. *ṅiddā*, хинди, неп. *nīd*, бенг. *nīd*, кашм. *nēndīr* ‘сон’), *\*pr > p* (*piyā* ‘любимый’ при др.-инд. *priya-*, пали, пракр. *piya-*, хинди *piyā* ‘любимый, возлюбленный’, лахнда *pīā*, пандж. *pīu*, гудж. *piyu*, мар. *piyo* ‘муж, возлюбленный’; *pathamo* ‘первый’ при др.-инд. *prathama-*, пали *paṭhama-*, пракр. *pathama-*, *paḍhama-*, ср. также хинди *pahlā*, пандж., бенг., мар. *pahlā*, лахнда *peihlā*, неп. *pailo*, гудж. *pahelū* ‘первый’ < *\*prathil(l)a-* (Turner 1966: 488); *pabha* ‘свет’ при др.-инд. *prabhā*, пали, пракр. *pabhā* ‘свет’, лахнда *pōh*, пандж. *pauih*, хинди *roh*, *rau* ‘рассвет’, асс. *piwō* ‘восход’, кашм. *prawi* ‘лучи, солнечный свет’; *pāvēha* ‘получает’ при др.-инд. *prāpayati*, пали *pāpēti*, пракр. *pāvāi*, ср. лахнда *pāvāṅ*, пандж. *pāiṅā*, хинди *pānā*, неп. *pāiṅi*, бенг. *pāoyā*, кашм. *prāwūn* ‘получать’), *\*bhr > bh* (*bham-* ‘заблуждение’ при др.-инд. *bhrama-* ‘ошибка, заблуждение; вращение; головокружение’, пракр. *bhama-* ‘блуждание; ошибка’, лахнда, пандж. *bhaū* ‘головокружение’, хинди *bhawō* ‘вращение, кружение’, асс. *bhuwō* ‘обман’, кашм. *bra-*

<sup>8</sup> Ср. указание Дж. Грирсона относительно языка *Mahānayaṅprakāśa*: “Except for two or three doubtful expressions, *Śitikaṅṭha*’s vocabulary is entirely Indo-Aryan” (Grierson 1929: 77).

<sup>9</sup> Рефлекс и.-е. *\*gh, \*g<sup>wh</sup>* в позиции палатализации.

<sup>10</sup> Древнеиндийская и авестийская формы в конечном итоге восходят к и.-е. *\*sneig<sup>wh</sup>-* ‘снежить’ (Mayrhofer 1996: 772).

<sup>11</sup> В примерах из языка *Bāṇāsarakathā* мы, следуя общепринятой индологической традиции, используем букву *c* для транслитерации знака индийских письменностей шарада и деванагари, передающего глухую непридыхательную палатальную аффрикату. При фиксации же кашмирского материала для данной фонемы нами используется символ *č*, в то время как при помощи символа *c* обозначается глухая непридыхательная зубная аффриката.

*miṇ* ‘блуждать’). Как можно видеть, кашмирские соответствия приведенных древнеиндийских слов сохраняют общеиндоиранские группы согласных без изменений и не могут являться отражениями форм с более инновационной фонетикой, засвидетельствованных в поэме *Bāṇāsarakathā*.

Следы средне- и новоиндийского развития обнаруживает также в отдельных случаях и соответствие древнеиндийской группы *kṣ* (*tikkho* ‘острый’ при др.-инд. *tikṣṇa-*, пракр. *tikkho*, хинди *tikhā*, пандж., лахнда *tikkhā*, неп. *tikho*)<sup>12</sup>. Как было показано нами в прошлых работах, в кашмири такое отражение характерно только для индоарийских заимствований, в то время как в исконной лексике др.-инд. *kṣ* соответствуют глухие придыхательные аффрикаты *čh* и *ch* (Коган 2011; 2015).

Все рассмотренные историко-фонетические изоглоссы однозначно говорят в пользу индоарийского характера языка *Bāṇāsarakathā*. В данной связи заслуживает упоминания и тот факт, что в поэме, как и в трактате *Mahānayaṅprakāśa*, имеются продолжения древнеиндийских композитов, характерных только для позднего санскрита и не обнаруживающих соответствий за пределами индоарийской группы: *paramātha-* ‘высшая душа’ < др.-инд. *paramātman-*; *tibhavana* ‘три мира’ < др.-инд. *tribhavana-*; *nirpēnda-* ‘Индра среди царей’ < др.-инд. *nṛpēndra-*; *svakamma-* ‘своя карма’; *candavadani* ‘лунолика’ < др.-инд. *candravadanā*.

Весьма примечательной чертой исторической фонетики языка *Bāṇāsarakathā* является ничем не мотивированный разницей в звукосоответствиях. Данное явление, характерное также для языка *Mahānayaṅprakāśa* и для среднеиндийских литературных языков, объясняется тем, что авторы текстов не строго следовали правилам пересчета с санскрита, описанным в грамматических трактатах, а в случае с *Mahānayaṅprakāśa* – еще и недостаточной разработанностью самих правил (Коган 2010). Для *Bāṇāsarakathā* наиболее частотны следующие типы неоднозначного историко-фонетического развития:

- 1) отражение зубных смычных в интервокальном положении: *t, d > y, v, 0* (*kēy-*, *kyo* ‘сделанный’ при др.-инд. *kṛta-*; *appū* ‘отданный, врученный’ при др.-инд. *arpita-*; *kīśū* ‘какой?’ при др.-инд. *kīdṛśa-*; *vāyān* ‘играющий (на музыкальном инструменте)’ при др.-инд. *vādayati* ‘играет’; *revān* ‘плачущий’ при др.-инд. *rudati* ‘плачет’); сохранение в *t, d* неизменном виде (*māt* ‘мать’ при др.-инд. *mātar-* (Nom Sg *mātā*); *duhit* ‘дочь’ при др.-инд. *duhitar-* (Nom Sg *duhitā*); *mito* ‘мертвый’ при др.-инд. *mṛta-*; *midu* ‘нежный’ при др.-инд. *mṛdu-*; *adiśya* ‘невидимый’ при др.-инд. *adṛśya-*; *vada-tā-s* ‘скажи ему!’ при др.-инд. *vadati* ‘говорит’; *pasādē* ‘подношением (Instr Sg)’ при др.-инд. *prasāda-* ‘подношение’);
- 2) отражение интервокального *p*: *p > v* (*pāvēha* ‘получает’ при др.-инд. *prāpayati*; *thāvēta* ‘установив’ при др.-инд. *sthāpayati* ‘устанавливает’; *nīvara* ‘ножной браслет’ при др.-инд. *nīpura-*); сохранение интервокального *p* в неизменном виде (*palāpē* ‘причитаниями (Instr Sg)’ при др.-инд. *pralāpa-* ‘причитания’; *kīpa* ‘милосердие’ при др.-инд. *kṛpā-*; *narapata* ‘царь’ при др.-инд. *narapati-*);
- 3) отражение начального и интервокального *ś*: *ś > s* (*siva* ‘Шива’ при др.-инд. *śiva-*; *sadiśo* ‘подобный’ при др.-инд. *sadṛśa-*; *asra* ‘слеза’ при др.-инд. *aśru-*); *ś > h* (*daha* ‘10’ при др.-инд. *daśa*; *kāha* ‘11’ при др.-инд. *ekādaśa*; *nāhen* ‘разрушают’ при др.-инд. *nāśayanti*); сохранение начального и интервокального *ś* в неизменном виде (*śira* ‘голова’ при др.-инд. *śiras-*; *śun-* ‘слышать’ при др.-инд. *śṛṇōti* ‘слышит’; *bhāśū* ‘12’ при др.-инд. *dvādaśa*; *kīśū* ‘какой?’ при др.-инд. *kīdṛśa-*);

<sup>12</sup> Отмечен также пример взаимозаменяемости *kṣ* и *kh*: *yakṣēta* и *yakhēta* ‘как’ (Toṣkhānī 1972: 247). В большинстве случаев, однако, др.-инд. *kṣ* в тексте поэмы сохраняется (по крайней мере, на письме) в неизменном виде (Toṣkhānī 1972: 258–259).

- 4) отражение интервокальной группы  $ṣṭ$ :  $ṣṭ > ṭṭh$  (*diṭṭho* ‘взгляд’ при др.-инд. *dṛṣṭi*;  
*biṭṭho* ‘сидящий’ при др.-инд. *upaviṣṭa*); сохранение группы  $ṣṭ$  в неизменном виде  
(*kudiṣṭa* ‘недобрый взгляд’ при др.-инд. *kudṛṣṭi*; *ukiṣṭa* ‘превосходный’ при др.-инд.  
*utkrṣṭa*).

В ряде случаев одно и то же древнеиндийское слово может иметь более одного соответствия<sup>13</sup>: *udayo* и *uyo* ‘поднявшийся’ при др.-инд. *udita*-; *sadiso* и *sadiśo* ‘подобный’ при др.-инд. *sadṛśa*-; *sahassa*, *sassa* и *sāss* ‘1000’ при др.-инд. *sahasra*-; *śatta* и *śatti* ‘сила’ при др.-инд. *śakti*-; *ratta* и *ratthā* ‘кровь’ при др.-инд. *rakta*-.

Весьма высокая степень историко-фонетической нерегулярности, демонстрируемая приведенным выше материалом, является, на наш взгляд, наглядным и неоспоримым свидетельством искусственного характера языка *Bāṇāsarakathā*. Представляя собой стилизованный идиом, подобный литературным пракритам и апабхрانشа, язык *Bāṇāsarakathā*, однако, не идентичен ни одному из них. Более того, он обнаруживает целый ряд черт, резко отличающих его от большинства средне- и новоиндийских языков. В исторической фонетике к этим чертам относится отсутствие перехода  $y > j$  в начальной позиции (ср. *yo* ‘который’ при др.-инд. *yaḥ*, пракр. *jō*, *je*, апабхрانشа, хинди, пандж., лахнда, синдхи, неп., мар. *jō*, бенг., гудж. *jē*, асс. *zi*; *yid* ‘если’ при др.-инд. *yadi*, пракр., апабхрانشа *jaī*, хинди, пандж., лахнда, синдхи *jē*, мар. *jai*; *yuvina* ‘молодость’ при др.-инд. *yauvana*-, пракр. *jōvaṇa*-, *jovvaṇa*-, *juvvaṇa*-, пандж. *jōvaṇ*, хинди, неп. *jōvan*), а также переход предвокального  $ṣ > ś$ ,  $śś$  (ср. *śiyi* ‘6’ при др.-инд. *ṣaṣ*-, ср. также пали *saṭṭhi*-, пракр. *saṭṭhī*-, хинди, мар. *sāṭh*, пандж., лахнда *saṭṭh*, синдхи *saṭhi*, неп., гудж. *sāṭhi* ‘60’ < др.-инд. *ṣaṣṭi*;- *ēśu* ‘этот’ при др.-инд. *ēṣa*-, ср. пали, пракр. *ēśo*, апабхрانشа *ēhō* ( $h < s$ ); *vaśśuna* ‘дождь’ при др.-инд. *varṣana*-, ср. также пали *vassati*, пракр. *vassadi* ‘дождит’, хинди *barasna*, пандж. *vassṇā*, лахнда *vassan*, неп. *barsani*, гудж. *varasvī*, мар. *varasṇē* ‘дождить’).

Обе рассмотренные изоглоссы характерны и для языка *Mahānayaṇaparakāśa*. Последняя из них представляется нам особенно интересной. Децеребрализация более раннего церебрального  $ṣ$  привела к совпадению его рефлекса с сохранившимся без изменений индоиранским  $*ś$  (= др.-инд.  $ś$ ) – отражением праиндоевропейского палатального  $*k$  (ср. приведенные выше примеры сохранения  $ś$ : *śira* ‘голова’ при др.-инд. *śiras*-; *śun* ‘слышать’ при др.-инд. *śṛṇōti* ‘слышит’; *śatti*, *śatta* ‘сила’ при др.-инд. *śakti*-; *bhāśū* ‘12’ при др.-инд. *dvādaśa*). Утрата оппозиции двух старых сибилантов и возникновение в её результате двухсибилянтовой системы ( $s$ ,  $ś$ ) отличает язык *Bāṇāsarakathā* от кашмири, где в исконной лексике сохраняется троичное противопоставление<sup>14</sup>, но при этом сближает его с некоторыми новыми индоарийскими языками, в частности, с цыганским и частью языков группы пахари. Аналогичный историко-фонетический процесс, по всей видимости, прошел также в языке – источнике индоарийских заимствований в кашмири (Коган 2011).

Примечательно, однако, что, обладая некоторыми редкими для индоарийского ареала чертами в исторической фонетике, рассматриваемый язык обнаруживает при этом изоглоссы, объединяющие его с новоиндийскими языками. Особенно заметны такие изоглоссы на лексическом уровне, где налицо целый ряд слов и корней, широко представленных главным образом в позднем индоарийском (ср. *b(h)uta* ‘много’ при хинди, пандж., гудж., неп. *bahut*, цыг. *but*; *caḍ(h)*- ‘взбираться, подниматься’ при хинди, пандж., лахнда, синдхи, кум. *caḥh*-, гудж., мар. *caḍh*-, бенг., неп. *caḥ*- ‘подниматься, взбираться, вскарабкиваться’; *ithaya* ‘сюда’ при пандж. *itthe*, лахнда *itth* ‘здесь’, кум. *ith*, *ithā*

<sup>13</sup> Об аналогичной ситуации в языке *Mahānayaṇaparakāśa* см. Коган 2010: 16.

<sup>14</sup> Для кашмири характерно сохранение в неизменном виде старого  $*s$  и передвижение двух остальных сибилантов:  $ś > h$ ,  $ṣ > ś$ . Подробнее см. Коган 2009.

‘сюда’; *jal-* ‘гореть’ при хинди, пандж., кум., неп. *jal-*, гудж., мар., зап. пахари (чамеали) *jal-*, бенг. *jol-*, асс. *zol-*; *paḍ-* ‘падать’ при хинди, кум., гудж. *paṛ-*, мар. *paḍ-*, бенг. *paṛ-*, цыг. *per-*; *rah-* ‘оставаться’ при хинди, пандж., лахнда, синдхи, неп., гудж., мар. *rah-*, бенг., асс. *rāh-*, кум. *rau-*, зап. пахари (бхадарвахи) *rēh-*).

\* \* \*

Рассмотренные выше факты, как представляется, однозначно говорят в пользу предложенной нами трактовки терминов *dēśya* и *dēśabhāṣā* как возможных обозначений местных литературных, притом, искусственно созданных идиомов. Стихотворные вставки в трактате *Mahānayaaprakāśa* и поэму *Bāṇāsarakathā* можно считать двумя дошедшими до нас текстами на одном из таких идиомов<sup>15</sup>. Его можно было бы назвать кашмирским *dēśya* или кашмирским *dēśabhāṣā*. Правила пересчета с санскрита на этот стилизованный язык, по всей видимости, ещё XV в. не прошли этап формирования и не были кодифицированы. На это указывает продемонстрированный нами низкий уровень регулярности в фонетических соответствиях.

Представляется несомненным, что данный язык, являющийся искусственным конструктором, никоим образом не мог быть народно-разговорным языком Кашмира. При этом, однако, в тексте *Bāṇāsarakathā* встречается немало лексических и морфологических элементов, обнаруживающих следы дардского историко-фонетического развития и/или имеющих наиболее близкие этимологические параллели в языках дардской группы, прежде всего, в кашмири. Данная особенность была отмечена нами и для трактата *Mahānayaaprakāśa* (Коган 2010: 17–20), но в *Bāṇāsarakathā* подобные вкрапления гораздо более многочисленны. Установить точное их число, впрочем, не представляется возможным, одна из причин чему – непоследовательность письменной фиксации некоторых фонемных оппозиций. Так, довольно часто встречаются случаи свободного варьирования на письме звонких придыхательных и непридыхательных согласных: *bhu-* и *bi-* ‘становиться’; *bhuta* и *buta* ‘много’; *bannēta* и *bhannēta* ‘связав’. Подобное варьирование, вероятно объясняется ошибками, допущенными переписчиками-кашмирцами – носителями языка, лишенного звонкой придыхательной серии. Поскольку совпадение двух серий вследствие утраты придыхания у звонких аспирирует представляет собой общедардскую классифицирующую историко-фонетическую инновацию, примеры, подобные приведенным здесь, следует признать неясными с точки зрения возможности их отнесения к дардскому этимологическому пласту.

В ряде случаев, однако, вариант с сохранившимся придыханием не засвидетельствован: *dāri* (при отсутствии *\*dhāri*) ‘держит’, ср. др.-инд. *dhārayati*; *gari* (при отсутствии *\*ghari*) ‘дом (Loc Sg)’, ср. пали, пракр. *ghara-*, хинди, неп., мар., гудж. *ghar*, синдхи *gharu* ‘дом’. Особенно показателен несомненный пример дезаспирации общеиндоиранской аффрикаты *\*jh*: *daji* ‘горит’ при др.-инд. *dahati* ‘горит, жжет’, авест. *dažaiti* ‘жжет’, торв. *daž-*, май. *daz-*, катар. *dazā-* ‘жечь’, кашм. *daz-*, пхал., сави *daj-*, шина *daž-* ‘гореть’ < индоиран. *\*dajh-* < и.-е. *\*dheg<sup>wh</sup>-*. Он четко противостоит приводившимся нами примерам индоарийского развития той же аффрикаты и может быть однозначно проинтерпретирован как дардский по происхождению.

Историко-фонетические процессы, чуждые индоарийскому, но характерные для ряда дардских языков, затронули и систему гласных. К ним относится, например, поведение общеиндоиранского краткого дифтонга *\*ai*, теряющего в ряде случаев второй

<sup>15</sup> Материальная близость языков двух памятников настолько значительна, что мы видим основания рассматривать их именно как два варианта (или стиля) одного идиома.

компонент в преконсонантной позиции: *ak(k)a* '1' < \**aika-*, ср. кашм. *akh*, май., башк. *ak*, пхал. *āk*, др.-инд. *ēka-*, авест. *aēva-*. Как и дезаспирация аффрикаты \**jh*, данный фонетический переход контрастирует с отмеченным нами выше индоарийским развитием той же древней фонемы (стяжением дифтонга \**ai* > *ē*). Интересно, что числительное 'один' встречается в тексте поэмы в двух вариантах (*ēk-* и *ak(k)a*), относящихся, как можно судить по их вокализму, к двум разным этимологическим пластам.

Вероятные кашмирские элементы обнаруживаются в языке *Bāṇāsarakathā* не только в лексике, но и в морфологии. Примечательно, в частности, их наличие среди личных глагольных показателей (-*ā* 1 Sg Prs, ср. кашм. -*ī* (< \**-ā*); -*ēva* 2 Pl Prs, ср. кашм. -*iw*; -*aṇa* 3 Pl Prs, ср. кашм. -*n*)<sup>16</sup>. Весьма частотны в тексте поэмы суффицированные местоимения, этимологически, несомненно, связанные с соответствующими местоимениями в кашмири<sup>17</sup>: -*m* (субъектное мест. 1 л. ед. ч.), -*y* (дательное-объектное мест. 2 л. ед. ч.), -*n* (субъектное мест. 3 л. ед. ч.), -*s* (дательное-объектное мест. 3 л. ед. ч.), ср. кашм. -*m*, -*y*, -*n*, -*s* соответственно.

Среди лексических и семантических изоглосс, объединяющих язык *Bāṇāsarakathā* с дардскими языками, включая кашмири, особый интерес представляют те, что относятся к базисной части словаря: *nār* 'огонь', ср. кашм., говро *nār*, башк. *āngār*, торв. *aṇā*, май. *ṅgār*, пхал. *angōr*, шина *agār* 'огонь', др.-инд. *aṅgāra-* 'уголь'; *van-* 'говорить', ср. кашм. *wan-*, торв., май. *ban-*; *rōda* 'дождь', ср. кашм. *rūd*; *dap(p)-* 'говорить', ср. кашм. *dap-*; *ēr-* 'плыть', ср. кашм. *yīṅin* 'плыть (о предмете)'; *bi* 'я', ср. кашм. *bī* 'я', шина, пхал., май. *be* 'мы'; *cū* 'ты', ср. кашм. *cī*, *tuānō* 'мой', ср. кашм. *m'ōn*, май. *mī*, торв. *mē*, башк. *mā*, тир. *tuāna*, г.-б. *mōna*, паш. *tainā*; *suānō* 'твой', ср. кашм. *ṣōn*, башк. *ṣhā* тир. *ṣāna*.

Приведенные здесь примеры можно было бы расценить как сильный аргумент в пользу принадлежности рассматриваемого языка к дардской группе и его тесной генетической связи с кашмири. Действительно, такие элементы языка, как местоимения и словоизменяемые аффиксы, считаются практически не подверженными заимствованию и в силу этого наиболее показательными при определении генетической принадлежности идиома. Следует, однако, иметь в виду, что данное положение, так же, как и постулат о регулярности фонетических соответствий, было установлено для естественных языков, в то время как в искусственных, по-видимому, ничто не может помешать его нарушению. В этой связи интересно отметить, что и в морфологии, и в местоименной системе языка *Bāṇāsarakathā* имеются примеры дублетов, несомненно, родственных этимологически, но при этом обнаруживающих разное историко-фонетическое развитие и происходящих, вероятнее всего, из разных источников: -*aṇa* и -*anda* 3 Pl Prs (ср. др.-инд., пракр. -*anti*); *cū* и *tū* 'ты' (ср. апабхрэнша *tuḥu*, хинди *tū*)<sup>18</sup>. Данная особенность характерна и для литературных апабхрэнша<sup>19</sup>, где часть подобных дублетов могут рассматри-

<sup>16</sup> Все перечисленные показатели засвидетельствованы в языке *Mahānayaṅprakāśa* в формах -*ā*, -*īva* и -*aṇa* соответственно (Gireson 1929; Коган 2010: 19).

<sup>17</sup> Подробнее о кашмирских суффицированных местоимениях см., например, Хонелия 1976.

<sup>18</sup> Данная ситуация, как представляется, характерна для базисной лексики в целом. Показательно наличие в тексте *Bāṇāsarakathā* дублетов разного, вероятнее всего, индоарийского и дардского (кашмирского) происхождения, выражающих значения из стословного списка Сводеша: *vaśśuna* 'дождь' (ср. др.-инд. *varṣaṇa-*) наряду с *rōda*, *agaṇa* 'огонь' (ср. др.-инд. *agni-*) наряду с *nār*, *jal-* 'гореть' (ср. др.-инд. *jvalati*) наряду с *daj-* (ср. кашм. *daz-*), *vad-* 'говорить' (ср. др.-инд. *vadati*) наряду с *van-* и *dap(p)-*.

<sup>19</sup> Ср., например, сосуществование в парадигме личн. мест. 2 л. форм Acc Instr Loc Sg *taī* и *paī* (Елизаренкова 2004: 130; Tagare 1948: 216) или наличие двух глагольных окончаний 1 Sg Prs -*ami* и -*aīm* (Tagare 1948: 291) и 2 Sg Prs -*ahi* и -*asi* (Елизаренкова 2004: 132; Tagare 1948: 287; Bubenik 1996: 103). Аналогичная ситуация обнаруживается и в позднейшей форме апабхрэнша – авахатха: ср. дублетные формы Instr Sg относительного местоимения *jē* и *jeṇa* (Sen 1973: 83).

ваться как заимствования из санскрита, пракритов или же из живых позднесреднеиндийских диалектов, распространенных в иных географических ареалах.

Из неиндоарийских элементов в языке *Bāṇāsarakathā* помимо дардских следует упомянуть также персидские и проникшие через персидское посредство арабские: *ābād* ‘благоустроенный’ < кл.-перс. *ābād*; *jān* ‘душа’ < кл.-перс. *jān*; *carak* ‘колесо’, ср. кашм. *čar(a)kh* < кл.-перс. *čarχ*; *band* ‘закрытый’, ср. кл.-перс. *band kardan* ‘запирать, подвергать заключению’; *harta* ‘женская половина дома’ < кл.-перс. *har(a)m* ‘запретное, священное место; гарем, женская половина дома’ < араб. Вся подобная лексика обнаруживает соответствия в кашмири и вполне могла быть усвоена через последний.

\* \* \*

Присутствие в языке *Bāṇāsarakathā* многочисленных корневых и грамматических морфем, усвоенных из кашмири или через кашмири, представляется нам вполне объяснимым. Как известно, в индоарийских искусственных литературных идиомах, в частности, в пракритах, выделяется особый этимологический пласт, включающий заимствования из народно-разговорных языков (Вертоградова 2002: 9–10)<sup>20</sup>. Тот факт, что в рассматриваемом тексте этот пласт является преимущественно кашмирским, можно считать свидетельством кашмириязычия всего или, по крайней мере, значительной части населения кашмирского государства в XV в.

Вместе с тем, этимологический состав лексики *Bāṇāsarakathā* обнаруживает специфическую особенность, по-видимому, не характерную для среднеиндийских литературных языков. Эта особенность заключается в том, что уровень *dēśī* включает материал, происходящий не из одного, а, как минимум, из двух источников. Помимо численно преобладающих дардских элементов к нему, как было показано, относятся также вероятные новоиндийские<sup>21</sup>. Данная ситуация, несомненно, нуждается в объяснении. Внести в нее некоторую ясность, на наш взгляд, позволяют результаты недавних сравнительно-исторических исследований по языку кашмири. Согласно им, в период Средневековья Кашмир пережил смену языка: дардский язык – предок современного кашмири – вытеснил бытовавший в регионе ранее новоиндийский язык, ставший для кашмири субстратом (Коган 2011). Наиболее вероятным временем этой трансформации нам представляются XIII–XIV вв.<sup>22</sup> Если литературная традиция на кашмирском *dēśabhāṣā* возник-

<sup>20</sup> Индийские грамматисты относили такие заимствования к особому слою или уровню (*krama*) языка. Этот уровень обозначался термином *dēśī* («местный»).

<sup>21</sup> См. приведенный выше список лексических изоглосс, объединяющих язык *Bāṇāsarakathā* с новоиндийскими языками.

<sup>22</sup> Средневековая история Кашмира до XII в. включительно достаточно хорошо освещена в нарративных источниках, наиболее важным из которых является хроника Раджатарангини («Волны царей»), составленная в 1148–1149 гг. историком и поэтом Кальханой. Ни эта хроника, ни какие-либо другие документы не содержат даже косвенных указаний на события, которые могли бы стать причиной вытеснения индоарийского народно-разговорного идиома дардским. С другой стороны, наличие многочисленных кашмирских элементов в тексте *Bāṇāsarakathā* не оставляет сомнения в том, что в XV в. дардизация населения Кашмирской долины должна была зайти достаточно далеко. Таким образом, XII в. может в данном случае рассматриваться как *terminus post quem*, а XV в. – как *terminus ante quem*. Возможно, благоприятные условия для процесса смены языка создало монгольское завоевание XIII в., вследствие которого Кашмирская долина оказалась политически объединена с дардским языковым ареалом – горными областями Гиндукуша и Каракорума. Характерное для этих областей перенаселение и острый дефицит земли, должны были способствовать массовой эмиграции жителей в Кашмир, частично обезлюдевший в результате как походов монголов, так и предшествовавших им продолжительных междоусобных войн. Подробнее см. Коган 2012; 2014.



ла до начала этого периода, что представляется весьма вероятным, то в течение определенного времени главным источником пополнения лексики слоя *dēśī* в текстах на данном языке должен был являться распространенный некогда в Кашмирской долине индоарийский диалект. К XV же веку смена языка завершилась или, по крайней мере, была близка к завершению. Вследствие этого ранний кашмири не мог не превратиться в региональный лингва-франка. Он, скорее всего, являлся также родным языком заказчика поэмы – султана Зайн-уль-Абидина, а возможно, и ее автора. В подобной ситуации сильное влияние его на язык *Bāṇāsarakathā* было вполне естественным.

Из вышесказанного, в частности, следует, что дальнейшее изучение рассмотренного нами литературного идиома может представлять интерес как для специалистов по дардским языкам, в частности, кашмири, так и для индологов. С одной стороны, кашмирские вкрапления в тексте поэмы могут считаться самой ранней надежно датированной фиксацией дардских лексических и морфологических единиц. Единицы эти отражают языковое состояние, характерное для XV в., и обнаруживают некоторые историко-фонетические архаизмы, среди которых можно отметить, например, отсутствие умлаута<sup>23</sup> (ср. *tuānō* ‘мой’ при кашм. *m’ōn*; *tuānē* ‘мой’ при кашм. *m’ēn*; *vātō* ‘прибывший’ при кашм. *wōt*; *vācī* ‘прибывшая’ при кашм. *wāc*). Анализ дардских элементов, направленный на выявление других архаичных черт, несомненно, мог бы пролить свет на целый ряд фактов исторической фонетики и грамматики кашмири. С другой стороны, индоарийская лексика в *Bāṇāsarakathā* в некоторой своей части должна обладать особенностями, характеризующими не сохранившийся и не дошедший до нас в письменных памятниках ранний новоиндийский язык<sup>24</sup>. Значение подобного материала для сравнительной индологии трудно переоценить. Нужно отметить, что в работе Ш. Ш. Тошкхани, как нам представляется, кашмирским элементам уделено неоправданно больше внимания, нежели индийским<sup>25</sup>. Это означает, что именно последние требуют наиболее интенсивного исследования в будущем.

### Сокращения

авест. — авестийский	кашм. — кашмири
араб. — арабский	кл.-перс. — классический персидский
асс. — ассамский	кум. — кумауни
башк. — башкарлик	май. — майян
бенг. — бенгали	мар. — маратхи
г.-б. — гавар-бати	неп. — непали
гудж. — гуджарати	пандж. — панджаби
др.-инд. — древнеиндийский	паш. — пашаи
зап. пахари — языки западной подгруппы группы пахари	пракр. — пракриты
и.-е. — индоевропейский	пхал. — пхалура
катар. — катаркалаи	тир. — тирахи
	торв. — торвали
	цыг. — цыганский

<sup>23</sup> О кашмирском умлауте и относительной хронологии его возникновения см. Коган 2009.

<sup>24</sup> Другим источником сведений об этом языке являются индоарийские субстратные заимствования в кашмири. О них см. Коган 2011.

<sup>25</sup> Причиной этого была, по-видимому, априорная убежденность ученого в том, что язык *Bāṇāsarakathā* представляет собой раннюю форму кашмири.

## Литература

- Вертоградова, В. В. 2002. *Пракриты. Издание второе*. Москва: Восточная литература.
- Елизаренкова, Т. Я. 2004. *Языки мира. Индоарийские языки древнего и среднего периода*. Москва: Academia.
- Захарьин, Б. А., Д. И. Эдельман. 1971. *Язык кашмири*. Москва: Наука (Главная редакция восточной литературы).
- Коган, А. И. 2009. К вопросу о ряде фонетических изменений в языке кашмири и их относительной датировке. В: Г. С. Старостин (ред.). *Аспекты компаративистики 4 [Aspects of Comparative Linguistics 4]*: 25–54. Москва: Издательство РГГУ.
- Коган, А. И. 2010. К вопросу о языке несанскритских фрагментов памятника Mahānayaṅprakāśa. Ранний кашмири или поздний апабхранша? *Вопросы языкового родства* 3: 13–21.
- Коган, А. И. 2011. К характеристике индоарийских элементов в языке кашмири. *Вопросы языкового родства* 5: 23–47.
- Коган, А. И. 2012. Еще раз о монгольских завоеваниях и монгольском владычестве в Кашмире. *История и современность* 1: 90–110.
- Коган, А. И. 2014. Кашмир и его соседи в XII–XIII веках. *Восток (Oriens)* 4: 37–47.
- Коган, А. И. 2015. Некоторые вопросы генетической классификации дардских языков по данным исторической фонетики. *Вопросы языкового родства* 13/1–2: 1–21.
- Хонелия, Н. А. 1976. Суффиксированные личные местоимения и некоторые особенности синтаксиса простого предложения в кашмири. В: Н. А. Дворянков (ред.). *Индийская и иранская филология. Вопросы грамматики*: 218–233. Москва: Наука.
- Эдельман, Д. И. 1965. *Дардские языки*. Москва: Наука (Главная редакция восточной литературы).

## References

- Bubenik, Vit. 1996. *The Structure and Development of Middle Indo-Aryan Dialects*. Delhi: Motilal Banarasidass.
- Edelman, Dzhoy. 1983. *The Dardic and Nuristani languages*. Moscow: Nauka.
- Edelman, Dzhoy I. 1965. *Dardskije jazyki*. Moskva: Nauka (Glavnaja redaktsija vostochnoj literatury).
- Elizarenkova, Tatiana Ya. 2004. *Yazyki mira. Indoarijskije jazyki drevnego i srednego perioda*. Moskva: Academia.
- Grierson, George A. 1919. *Linguistic survey of India. Vol. VIII. Pt. 2. Specimens of the Dardic or Piśāca languages (including Kāshmirī)*. Calcutta: Superintendent Government Printing (India).
- Grierson, George A. 1929. The language of the Mahānayaṅprakāśa, an examination of the Kashmiri as written in the 15th Century. *Memoirs of the Asiatic Society of Bengal* XI (2): 73–129.
- Khoneliya, N. A. 1976. Suffigirovannyje lichnyje mestoimenija i nekotoryje osobennosti sintaksisa prostogo predlozhenija v kashmiri. In: N. A. Dvor'ankov (ed.). *Indijskaja i iranskaja filologija. Voprosy grammatiki*: 218–233. Moskva: Nauka.
- Kogan, Anton I. 2009. K voprosu o r'ade foneticheskikh izmenenij v jazyke kashmiri i ikh otnositel'noj datirovke. In: G. Starostin (ed.). *Aspekty komparativistiki 4 [Aspects of Comparative Linguistics 4]*: 25–54. Moskva: Izdatel'stvo RGGU.
- Kogan, Anton I. 2010. K voprosu o jazyke nesanskritskikh fragmentov pam'atnika Mahānayaṅprakāśa. Rannij kashmiri ili pozdnij apabhraṅsha? *Journal of Language Relationship* 3: 13–21.
- Kogan, Anton I. 2011. K kharakteristike indoarijskikh elementov v jazyke kashmiri. *Journal of Language Relationship* 5: 23–47.
- Kogan, Anton I. 2012. Jeshche raz o mongol'skikh zavojevanijakh i mongol'skom vladychestve v Kashmire. *Istorija i sovremennost'* 1: 90–110.
- Kogan, Anton I. 2014. Kashmir i jego sosedi v XII–XIII vekakh. *Vostok (Oriens)* 4: 37–47.
- Kogan, Anton I. 2015. Nekotoryje voprosy geneticheskoi klassifikatsii dardskikh jazykov po dannym istoricheskoi fonetiki. *Journal of Language Relationship* 13/1–2: 1–21.
- Mayrhofer, Manfred. 1996. *Etymologisches Wörterbuch des Altindoarischen*. Bd II. Heidelberg: Universitätsverlag Carl Winter.
- Sen, Subhadra Kumar 1973. *Proto-New Indo-Aryan*. Calcutta: Eastern Publishers.
- Tagare, Ganesh V. 1948. *Historical Grammar of Apabhraṅśa*. Poona: Deccan College.

- Toṣkhānī, Shashi Shokhar. 1972. *Bāṇāsurovadh [Bāṇāsarakathā]. Kāśmīr viśvavidyālay kā pī ḛc ḍī upādhi kē liyē prēṣit śōdh prabandh*. University of Kashmir. Available at: <http://shodhganga.inflibnet.ac.in/handle/10603/56812> [accessed 03.05.2018].
- Turner, R. L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages*. Oxford University Press.
- Vertogradova, Viktorija V. 2002. *Prakṛity. Izdaniye vtoroye*. Moskva: Vostochnaya literatūra.
- Zakharyin, Boris A., Dzhoy I. Edelman. 1971. *Yazyk kashmiri*. Moskva: Nauka (Glavnaya redaktsiya vostochnoy literatury).

*Anton Kogan. On the language of the Bāṇāsarakathā poem and its relation to Kashmiri*

The language of *Bāṇāsarakathā* (“The story of the demon Bana”), a poem written in Kashmir in the mid-15<sup>th</sup> century, is traditionally considered to be an early form of Kashmiri. However, certain facts of historical phonology and lexicon lead us to doubt this viewpoint. In the present article an attempt is made to throw maximum light on the issue. The author concludes that the language of the main body of the poem should be classified as Indo-Aryan, and is furthermore undoubtedly stylized and artificial. At the same time, the text contains numerous fragments in early Kashmiri, which should be regarded as the earliest extant datable written record of Dardic material. A number of Indian elements attested in the poem were most probably borrowed from a New Indo-Aryan language, spoken in the Vale of Kashmir before 13<sup>th</sup>–14<sup>th</sup> centuries.

*Keywords:* Kashmiri language, Dardic languages, Indo-Aryan languages, Prakrits, Apabhraṃśa, artificial languages, language shift.

## Методы определения фокуса в мертвом языке: на пути к реконструкции праиндоевропейского синтаксиса<sup>1</sup>

Работа посвящена проблеме определения фокусных и топикальных элементов в хеттском и древнегреческом языках, двух из древнейших зафиксированных в письменности языков индоевропейской семьи. Выявление информационной структуры предложения дает ключи к определению места ударения, что актуально для хеттского языка, а также необходимо для реконструкции синтаксиса на праязыковом уровне. На данный момент методология работы с информационной структурой в мертвых языках недостаточно разработана, выделение фокусов не обосновывается и может выглядеть произвольным. Мы предлагаем использовать комплексный подход, последовательно применяя несколько методов, разработанных на материале живых языков, но по отдельности не дающих в корпусных языках уверенного ответа на вопрос, почему это фокус: метод фокусного скелета (определения ряда фокусных альтернатив наряду с анализом топикальной структуры предложения) и метод вопроса к предложению, применяемые на фоне контекстного и филологического анализа.

*Ключевые слова:* хеттский язык, древнегреческий язык, праиндоевропейский язык, реконструкция синтаксиса, информационная структура, фокус, топик, семантика фокусов

### 1. Проблемы реконструкции праиндоевропейского синтаксиса

Вопрос, можно ли реконструировать синтаксис праязыка, до сих пор остается дискуссионным – однозначных подходов к методологии пока не предложено. Рассчитать, какие изменения синтаксических характеристик на протяжении достаточно большого периода истории языка, с учетом данных родственных языков и ареальных контактов, привели к состоянию, зафиксированному в древнейших письменных языках, на данный момент выглядит как нерешаемая задача. И всё же без представления о синтаксисе и информационной структуре о полноценной реконструкции ПИЕ говорить не приходится. В принципе, возможной кажется реконструкция определенных синтаксических характеристик, таких как порядок слов и, возможно, принципы организации клаузы (порядок слов в именной группе, группе предглагольных составляющих, в левой периферии, организации топиков и фокусов; порядок слов в предложной группе и т. п.). При этом без восстановления соответствующих характеристик на этапе исторически зафиксированных письменных языков, таких как древнегреческий и хеттский (двух из древнейших зафиксированных в письменности индоевропейских языков), всякая попытка делать обобщение на уровне ПИЕ потерпит неудачу.

---

<sup>1</sup> Исследование информационной структуры хеттского языка проводится в рамках проекта «Типология зарождения стиха: точные и компьютерные методы» и поддержана Российским фондом фундаментальных исследований (проект 17-06-00392).

Говоря о применении компаративистского метода к синтаксису, Даниэль Пети<sup>2</sup> предлагает исходить из шестиуровневой системы реконструкции, которая единственная может отчасти гарантировать исследователей от произвольности в выводах относительно праиндоевропейского состояния: 1) внутреннее описание синтаксиса языка А, 2) внутренняя реконструкция языка А до состояния А1; 3) внутреннее сравнение А и А1, с результатом в виде подробного описания изменений, произошедших с языком за период А1–А; 4) сравнение синтаксических особенностей языка А и языка В, происходящего от языка С; 5) предположительная реконструкция языка С. Последняя производится путем сравнения состояния А, состояния А1, зарегистрированных изменений в период А1–А, состояния В. Сравнительно-исторический метод позволяет на основании этих данных реконструировать предположительное состояние С, если под языками А и В мы имеем в виду доступные нам живые или хорошо описанные мертвые родственные языки, с которых начинается ступенчатая реконструкция, и если при этом учитываются, на деле, не два, а все известные нам языки данной языковой группы. Так, для реконструкции порядка слов необходимо собрать и формализовать информацию о порядках слов во всех родственных языках, выяснить последовательное развитие системы порядков слов в диахроническом аспекте для каждого языка, где это доступно; формализовать полученную информацию в виде правил; далее эти правила экстраполируются на те языки, для которых историческое развитие не зафиксировано, в итоге позволяя выдвинуть предположение о порядке слов в неопisanном языке-предке. Описание реконструируемого языка должно, по Пети, стать шестым шагом, – возможно, недостижимым. Таким образом, чем лучше описаны разные аспекты синтаксиса родственных языков на разных этапах развития, тем больше вероятности, что мы сможем подойти и к реконструкции синтаксиса праязыка. В последние десятилетия интерес к архаическому синтаксису индоевропейских языков в целом растет<sup>3</sup>, и есть надежда, что постепенно объем наших знаний о том периоде их развития, который касается II–I тыс. до н. э., позволит наконец делать уверенные обобщения.

До сих пор, однако, по понятным причинам сравнительно мало внимания уделялось трудноформализуемой области информационной структуры древних индоевропейских языков, а те исследователи, которые брались за данную тему (среди прочих, Naugan 2001, Goedegebuure 2014, Goldstein 2016, Сидельцев 2017, Meelen в печати и др.), к сожалению, часто не занимались проблемами формализации метода – в их работах обычно не обсуждается, каким образом в мертвых языках можно выделять фокусные и топикальные элементы и как стандартная для живых языков методология адаптируется для письменных языков. Цель нашей работы – привлечь внимание к произвольности процедуры фиксации данных о фокусных элементах в древних индоевропейских языках II–I тыс. до н. э. и предложить более формализованные методы, адаптированные к работе с мертвыми языками. Более формализованный подход к данным древнегреческого и хеттского языков позволит закрыть один из пробелов на пути к реконструкции праиндоевропейского синтаксиса.

<sup>2</sup> Daniel Petit, Ecole Normale Supérieure, личн. сообщ.; лекция по историческому синтаксису в рамках 4-й Летней школы по индоевропеистике, Университет Павии, Италия, сентябрь 2017 года.

<sup>3</sup> Достаточно отметить следующие работы: Watkins (1964), Lehmann (1974), Kortlandt (1983), Ginnakis (1997), Bauer (2000), Luraghi (2004, 2012), Barðdal & Smitherman (2009), Barðdal (2013), Kulikov, Lavidas (2015), не забывая, разумеется, труды Дельбрюка и Хирта. На данный момент наибольший интерес исследователей в области исторического синтаксиса привлекают темы порядков слов, эргативного/активного строя, субъектного и объектного маркирования, – по понятным причинам, именно эти характеристики синтаксиса легче поддаются фиксации, формализации и обработке большого количества материала с помощью машинных методов.

### 3. Зачем и как искать фокусы в мертвых языках: постановка проблемы

Информационная структура (ИС) предложения важна, в частности, для объяснения порядков слов, для понимания закономерностей их варьирования, и, с другой стороны, для уточнения роли предложения в дискурсе. Многие вопросы синтаксиса решаются только с привлечением данных об информационной структуре, и древних языков это наблюдение касается в полной мере. При этом теоретические рамки и терминология, используемые для описания и анализа информационной структуры, сильно варьируются даже для живых языков, а для мертвых предпринимаются разнонаправленные попытки приспособить имеющийся терминологический аппарат для анализа корпусного языка.

Любопытно, что древнегреческий, порядок слов в котором принято считать свободным (non-configurational), согласно последним исследованиям, в значительной степени оказывается связан информационной структурой, и в этом смысле фактически не может считаться свободным. Так, Д. Голдстейн в работе, посвященной синтаксису классического древнегреческого (Goldstein 2016: 17–18), применяет к древнегреческому термин *discourse-configurational*, предложенный Н. Dik (1995), то есть зависящий от коммуникативного членения предложения. Наоборот, хеттский язык считается имеющим крайне жесткий порядок слов (Hoffner, Melchert 2008), но при этом он показывает варьирование порядков, которое тоже объясняется только посредством ИС (подробный анализ неканонического порядка слов и его связи с информационной структурой в хеттском языке см. в Молина 2018). В русском языке, имеющем, согласно общепринятому представлению, свободный порядок слов, за выделение фокуса, то есть формулирование тем самым ряда фокусных альтернатив, в значительной степени отвечает интонационный контур (типы ИК в русском языке см., например, Брызгунова 1980, Янко 2008, см. тж. Падучева 2015 о коммуникативной структуре предложения, с литературой), однако и тут термин «свободный» применительно к порядку слов следует считать условным, связанным с информационной структурой.

Переходя от анализа живых языков к анализу мертвых, мы натываемся на методологическую проблему следующего рода. Поскольку ИС непосредственно связана с интонацией («the accent units are the domains of activation for new, accessible and/or given information», E. Couper-Kuhlen 2001:15), стандартным методом ее выявления, то есть определения ее составляющих, таких как фокус/топик/данное, является поиск отрицательных контекстов через эксперимент с носителем – «так сказать можно» либо «так сказать нельзя». Но экспериментальный подход невозможен в языке, все носители которого давно мертвы. Это ставит перед нами сугубо теоретическую задачу разработки метода, пригодного для использования на материале мертвых языков.

### 4. Фокусы с теорией: что такое фокус

Ламбрехт (1994) называет фокусом то, что не входит в пресуппозицию, то есть ее часть, не являющуюся само собой разумеющейся (Lambrecht 1994: 207). Методы анализа информационной структуры *Alternative Semantics* (альтернативная семантика) и *Structured Meanings* (структурированные значения: метод анализа, предложенный еще Р. Джекендоффом (Jackendoff 1972) и поддержанный М. Руттом (1985, 1992)) предполагают, что высказывание делится на две части: значение фокуса и значение фона/бэкграунда (background, см. также Krifka 2007: 39, Beaver, Clark 2008: 25). Фокус, согласно этому

представлению, является отношением между пропозицией в бэкграунде и той переменной, которая заполняет недостающее в пропозиции. «Если  $\phi$  – синтаксическая фраза, а  $C$  – синтаксически выраженная семантическая переменная,  $\phi \sim C$  вводит пресуппозицию, что  $C$  представляет собой фокусное подмножество  $[[\phi]]_{\text{focus}}$ , содержащее  $[[\phi]]_o$  (представленное в высказывании) и как минимум еще один элемент» (Rooth 1996: 279, «Where  $\phi$  is a syntactic phrase and  $C$  is a syntactically covert semantic variable,  $\phi \sim C$  introduces the presupposition that  $C$  is a subset of  $[[\phi]]_{\text{focus}}$  containing  $[[\phi]]_o$  and at least one other element»). При этом фокус может быть широким, на всем предложении или на предикате, и узким, на одной составляющей или фразе. В русле концепции Structured Meanings рассматриваются также понятия о сложном фокусе, – состоящем из вложенных друг в друга частей, – и мультифокусности, то есть сосуществовании в предложении нескольких фокусов разного типа.

В рамках функциональной грамматики, согласно (Dik 1997: 326), категорию фокуса можно описать следующим образом: «Фокусная информация в языковом выражении – информация, которая относительно остального является наиболее важной и значимой (the most important and salient) и говорящим воспринимается как наиболее адекватная для включения в прагматическую картину адресата».

Для работы с мертвыми языками, впрочем, формулировка Н. Дик не очень помогает, равно как и формализация в духе Alternative Semantics и Structured Meanings сама по себе. Возможно, эти методы можно было бы сделать более теоретически приемлемыми при изучении письменных языков с закрытым корпусом, если добавить<sup>4</sup>, что, опуская составляющие, находящиеся в фокусе, мы фактически лишаем выражение информационной ценности, – что позволяет работать с оценкой информационной ценности, опираясь на контекст. Можно также подчеркнуть наличие связи между ударением и важной информацией, что позволяет использовать маркеры ударения (такие как полноударные местоимения вместо, допустим, энклитических, либо данные фонетики, такие как изменения звуков под ударением, фиксируемые в записи) для анализа информационной структуры.

Большую помощь в анализе ИС также оказывает тот факт, что фокус может выделяться в языках не только с помощью интонации, но и с помощью других грамматических средств, в том числе фокусирующих частиц или определенного порядка слов, например клефта, как в ирландском примере (1), где набор фокусных альтернатив подразумевает открытый список пропозиций типа {дом, который [построил X]}:

- (1) An teach [a thóg Seán]CLEFT  
 DEF дом DREL строить Шон  
 ‘Дом, который построил Шон’ (Шкапа 2013:94)

## 5. Фокусы с фокусом: пути решения

Д. Голдстейн (2013) предлагает базово отталкиваться от противопоставления PRESUPPOSED (или ‘данное’) и NON-PRESUPPOSED (или ‘новое’), разработанное Prince (1981), Krifka (2007). Информация в пресуппозиции – то, что было уже эксплицировано в дискурсе, либо является частью общего знания (common knowledge), либо «может быть представлено в физическом или культурном контексте» (Goldstein 2013: 5), последнее

<sup>4</sup> Вслед за обсуждающей этот вопрос П. Худехебюре (Goedegebuure 2009: 101).

условие, к сожалению, довольно тяжело формализовать. Информация не в пресуппозиции (Beaver, Clark 2008) рассматривается как фокус высказывания. Голдстейн придерживается теории речевых актов, в которой утверждение предполагает истинность той части высказывания, в которую входит фокусная составляющая.

Метод, с помощью которого он предлагает выяснить, какая именно часть высказывания оказывается в древнегреческом языке в фокусе, какое значение имеет, – так называемый метод вопроса к предложению (Question Under Discussion, QUD), изначально предложенный (Hatcher 1956), активно применяется для анализа информационной структуры живых языков и хорошо апробирован на них. Согласно QUD, фокус является ответом на такой вопрос к предложению, какой можно задать с сохранением осмысленности высказывания, а коммуникация и дискурс понимаются как формулирование выраженных и невыраженных вопросов. Центральную мысль организации дискурса вокруг QUD много разрабатывал К. Робертс (Roberts 2012), но подобные идеи выдвигались и раньше, начиная с XIX в. (Kvíčala 1870, Paul 1920, Carlson 1983, Ginzburg 1995). Значительные добавления к данной модели были предложены Бюрингом (Büring 1997, 1999, 2003, 2006). Согласно Робертсу, структура дискурса представляется в виде набора QUD, на которые участники коммуникации дают ответы. По Бюрингу (2003), этот набор имеет иерархическую структуру, – чтобы ответить на более общий вопрос, требуется ответить на ряд более детальных. При этом фокус представляет собой переменную QUD (2). Вопрос А в (2) предполагает, что есть человек по имени Джон, которого собеседник В легко узнает по имени (имеет в своем поле знания, *common knowledge*), и что Джон что-то сделал. Фокус ответа В в (2) уточняет значение *x*.

- (2) А: Что сделал Джон?  
В: Он [вымыл посуду]<sub>FOC</sub>
- (3) Данное: Джон *x*, *x* ∈ {действия}  
Фокус: *x* = {вымыл посуду}

Описанным в примерах (2) и (3) образом QUD применяют к материалу живых языков с целью выделения фокуса – и формализации данных об информационной структуре предложения. Структура дискурса может, таким образом, быть описана как структура вопросов к предложению. Д. Голдстейн предлагает использовать QUD и переменную фокуса и для мертвых языков (4).

- (4) QUD: Что это?  
[Ἡροδότου Ἀλικαρνησέος ἱστορίας ἀπόδεξις] FOC ἦδε.  
Ἡροδότου Ἀλικαρνησέος  
Геродот.М.GEN.SG Галикарнас.Ф.GEN.SG  
ἱστορίας ἀπόδεξις ἦδε.  
исследование.Ф.GEN.SG представление.Ф.NOM.SG это.Ф.NOM.SG  
‘Это есть [представление исследования Геродота Галикарнасского]FOC’ (Goldstein 2016:31, Геродот «История»)

Теоретически это разумно: ничего не мешает применять метод QUD к текстам на мертвых языках, если мы способны их понимать. Практически же основной проблемой в применении QUD к материалу на мертвых языках является сама постановка вопроса – в живых языках благодаря интонации мы автоматически понимаем, как устроен вопрос; пример (2) не позволяет двойного толкования QUD, но можно представить себе аналогичный пример на русском или английском, допускающий, если не учитывать интона-



цию, два QUD – с фокусом на субъекте и с фокусом на глаголе. Поэтому выделение фокусного элемента информационной структуры предложения только с QUD, увы, оставляет слишком много пространства для исследовательского произвола.

Пример выделения узкого фокуса в классическом древнегреческом (5, в тексте фокус обозначен жирным шрифтом):

- Когда истинность утверждения  $\acute{\epsilon}\nu\ \delta\acute{\epsilon}\ \tau\acute{\omega}\iota\ \acute{\epsilon}\pi\iota\sigma\chi\epsilon\acute{\iota}\nu\ \acute{\epsilon}\nu\epsilon\sigma\tau\iota\ \alpha\gamma\alpha\theta\acute{\alpha}$  становится очевидной? Она становится очевидной [со временем]<sub>FOC</sub>.
- Данное: Это станет ясно X, X ε {обозначение времени}.
- Фокус: X = {со временем}

- (5)  $\acute{\epsilon}\nu\ \delta\acute{\epsilon}\ \tau\acute{\omega}\iota\ \acute{\epsilon}\pi\iota\sigma\chi\epsilon\acute{\iota}\nu\ \acute{\epsilon}\nu\epsilon\sigma\tau\iota\ \alpha\gamma\alpha\theta\acute{\alpha}$ .  $\epsilon\acute{\iota}\ \mu\grave{\eta}\ \pi\alpha\rho\alpha\upsilon\tau\acute{\iota}\kappa\alpha\ \delta\omicron\kappa\acute{\epsilon}\omicron\nu\tau\alpha\ \acute{\epsilon}\acute{\iota}\nu\alpha\iota$ ,  $\alpha\lambda\lambda'$  [ $\acute{\alpha}\nu\grave{\alpha}\ \chi\rho\acute{\omicron}\nu\omicron\nu$ ]<sub>focus</sub>  $\acute{\epsilon}\xi\acute{\epsilon}\upsilon\rho\omicron\iota=\tau\iota\varsigma=\acute{\alpha}\nu$ .
- |                         |                                   |  |   |   |  |
|-------------------------|-----------------------------------|--|---|---|--|
| $\acute{\epsilon}\nu$   | $\delta\acute{\epsilon}$          | $\tau\acute{\omega}\iota$                                  | $\acute{\epsilon}\pi\iota\sigma\chi\epsilon\acute{\iota}\nu$  | $\acute{\epsilon}\nu\epsilon\sigma\tau\iota$    | $\alpha\gamma\alpha\theta\acute{\alpha}$ . |
| В                       | PTCL                              | DET.N.DAT.SG   | ждать.INF.PRS.ACT   | быть.V.PRS.IND.ACT.3SG                          | хорошо.N.ACC.PL                            |
| $\epsilon\acute{\iota}$ | $\mu\grave{\eta}$                 | $\pi\alpha\rho\alpha\upsilon\tau\acute{\iota}\kappa\alpha$ | $\delta\omicron\kappa\acute{\epsilon}\omicron\nu\tau\alpha$   | $\acute{\epsilon}\acute{\iota}\nu\alpha\iota$ , |  |
| если.COMP               | NEG                               | немедленно.ADV   | казаться.PTCP.PRS.ACT.N.NOM.PL  | быть.INF.PRS.ACT                                |  |
| $\alpha\lambda\lambda'$ | $\acute{\alpha}\nu\grave{\alpha}$ | $\chi\rho\acute{\omicron}\nu\omicron\nu$                   | $\acute{\epsilon}\xi\acute{\epsilon}\upsilon\rho\omicron\iota=\tau\iota\varsigma=\acute{\alpha}\nu$ . |   |  |
| но                      | через                             | время.M.ACC.SG   | обнаружить.AOR.OPT.ACT.3SG=DET.C.NOM.SG=MOD   |   |  |
- ‘Есть хорошее в ожидании. Пусть вначале не очевидно, но [со временем]<sub>FOC</sub> это станет ясно’ (Goldstein 2016: 180; Геродот, «История»).

Относительно данной методологии возникают следующие вопросы. Почему в качестве фокуса рассматривается только составляющая  $\acute{\alpha}\nu\grave{\alpha}\ \chi\rho\acute{\omicron}\nu\omicron\nu$ ? Если переменная фокуса предполагает обозначение времени, пару  $\acute{\alpha}\nu\grave{\alpha}\ \chi\rho\acute{\omicron}\nu\omicron\nu$  составляет  $\pi\alpha\rho\alpha\upsilon\tau\acute{\iota}\kappa\alpha$ , также указывающее на время. В живом языке мы бы проанализировали интонационную структуру предложения и проверили, не следует ли выделять параллельный фокус, включающий в себя и  $\pi\alpha\rho\alpha\upsilon\tau\acute{\iota}\kappa\alpha$ . Кроме того, под вопросом остается фокус на отрицании и на значении «очевидно». Есть ли фокус на этих составляющих в рассматриваемом примере, остается невыясненным. Разумеется, Д. Голдстейн, из процитированной выше работы которого взят приведенный пример, оставляет в тени эти сомнения, даже если они и возникали.

П. Худехебюре (Goedegebuure 2009 и 2014) в своих работах о контрастивных фокусах в хеттском языке эти сомнения вполне высказывает, но от окончательного решения уходит. Как уже упоминалось выше, она предлагает исходить из анализа контекста, вычленив наиболее значимую информацию; однако практически ее анализ сводится к выделению нескольких контрастивных фокусов, наиболее распространенных в используемом материале (дейктические местоимения), и описанию их условий и окружающего контекста. В целом, Худехебюре отмечает: «Наш метод обнаружения фокусных выражений в письменных текстах должен строиться на анализе предыдущего контекста. Если клауза содержит именную группу в фокусе, остальная часть клаузы должна быть в значительной степени предопределена. Таким образом, мы должны изучать предыдущий контекст, ища предопределенную часть клаузы, содержащей фокус» («Our method for finding Focus expressions in written texts should be based on an analysis of the preceding context. If a clause contains a noun phrase in Focus, the rest of the clause should to a large extent be presupposed. We therefore need to search the preceding discourse for content that matches the presupposed part of the clause containing the Focus», Goedegebuure 2009: 100–101.) Кроме того, нам, как уже упоминалось, доступен анализ маркеров ударения, которые можно соотносить с важной информацией, что в случае хеттского языка включает использование полноударных местоимений вместо энклитик.

В примере (6), таким образом, фокус должен как минимум включать ударное местоимение *apēdani* ‘той’.

(6) NH (СТН 86) KUB 21.17 ii 8

Контекст: «Затем, во время правления моего брата я разделил (культ богини) Саусги из Самухи. Я построил храмы для нее в (городе) Урикина

*ni kī É<sup>md</sup>SIN.dU apēdani ADDIN*

CONN этот.ACC.SG дом Арматархунда тот.DAT.SG дать-1SG.PRS

и я подарил ей (= ‘именно той’) тот дом Арматархунды» («Эдикт Хаттусили III», Ünal 1974: 22f<sup>5</sup>).

Впрочем, П. Худехебюре сама задает вопрос: «Только ли *apēdani* входит в фокус? Означает ли, что наиболее важная информация данного предложения – что дом не был отдан никому другому?». На эти вопросы очевидного ответа нет, кроме напрашивающегося замечания, что для дальнейшего рассуждения требуется найти некоторый способ формализовать имеющиеся у нас данные. Представляется, что ни одна из описанных методик не дает однозначного решения, однако при комплексном подходе возможен более качественный анализ. Таким образом, мы предлагаем объединить методы анализа контекста, в том числе анализ фокусных альтернатив и QUD (настолько, насколько контекст позволяет как выделять ряд альтернатив, так и уточнять вопрос к предложению).

## 6. Фокусные альтернативы и анализ контекста

Худехебюре (2014) предлагает использовать для анализа мертвых языков метод фокусного скелета (*focus skeleton*), разработанный М. Руттом в рамках формальной семантики (Alternative Semantics, Rooth 1985, 1992). Однако она лишь ссылается на литературу, практически не описывая метод и не показывая, каким образом адаптирует его для работы с хеттским материалом. В работах (Молина 2016, Молина 2018) мы формализуем эту методику для анализа информационной структуры хеттского предложения. Описанную в примере (3) выше переменную фокуса можно представить в виде набора альтернатив, которые и являются семантическим значением фокуса высказывания. Именно этот ряд пропозиций в дальнейшем мы (вслед за Rooth 1985, 1992, 2007; Goedegebuure 2014) и будем называть *фокусным скелетом*. Данный метод также предполагает *контекстный анализ* с целью определения *ряда фокусных альтернатив* на основе *данного и нового* в предложении, а также *установленных и выводимых топиков* в ближайшем контексте.

Так, в предложении (7а) набор альтернатив включает в себя имена девушек, которых может любить Петя, а в предложении (7б) – имена молодых людей, которые любят Машу:

(7) а. Петя любит [Машу]<sub>ФОС</sub>.

б. [Петя]<sub>ФОС</sub> любит Машу.

Методика анализа клаузы с выделением фокусного скелета в мертвых языках означает, что каждый контекст должен быть проанализирован и сравнен текстуально с предыдущими и последующими контекстами, чтобы отыскать слова, повторяющиеся в ближайшем контексте и являющиеся топиками, выводимыми из контекста; должна быть проанализирована информация, известная нам о рассматриваемом тексте и его референтах, чтобы определить, насколько возможно, *common ground* для данного текста и

<sup>5</sup> Цит. по: Goedegebuure 2009: 98.

опирающиеся на него установленные топики. Данный подход позволяет выделить путем обозначения топиков новое и данное, восстановить набор альтернатив, используемых фокусом, понять, является ли полученный ряд фокусных альтернатив закрытым или открытым. Согласно имеющимся у нас теоретическим знаниям и разработкам в сфере информационной структуры, см., к примеру, Krifka 2007, открытый ряд альтернатив всегда бывает у информационного фокуса, закрытый – у контрастивного, бинарный – у фокусов подтверждения и отрицания.

Далее прописывается список возможных вариантов для данной пропозиции, в процессе чего мы проверяем сделанный анализ. Описанная процедура позволяет достаточно уверенно выявить семантику фокуса и далее уточнить слова, находящиеся в фокусе, даже для мертвых языков<sup>6</sup>. Анализ широкого контекста помогает также установить сферу действия фокуса и в целом выявить информационную структуру предложения. Фокусный скелет пропозиции может быть представлен в формате:

Topic [Focus1: alternative1, alternative2, ..., alternativeN] ([Focus2: alternative1, alternative2, ..., alternativeN]).

По приведенной записи видно, что предложение способно иметь более одного фокуса разной семантики: «В одном и том же предложении одно выражение представляет альтернативы, используемые одним образом, а другое выражение представляет альтернативы, которые используются другим образом» (Krifka 2007). Мультифокусность следует учитывать при определении фокусной переменной, как, например, в (10) далее.

Пример анализа хеттского контекста с учетом фокусного скелета и QUD:

(8) MH/MS (СТН 186) НКМ 17 obv. 12

[*nu=tta apiya ŪL ištamaššer*] *nu=tta kinun=pat ištamaššer*

*nu=tta kinun=pat ištamašš-er*

CONN=ты.DAT.2SG сейчас=EMPH слышать-3PL.PST

«Только сейчас они о тебе услышали» («Письма из архива Машат», Hoffner 2009: 124).

QUD: Когда о тебе слышали враги? Ответ: Сейчас, а не зимой.

Предыдущий контекст: «Враги о тебе зимой не слышали, они о тебе сейчас услышали». Фокусный скелет: «Враги услышали о тебе когда? – X», X = [*зимой, весной, осенью, летом, в X году, тогда, сейчас*].

В контексте возникает противопоставление: «Сейчас, а не зимой (зимой они о тебе не слышали)»; таким образом, мы наблюдаем фокус контраста. В непосредственном контексте уже присутствует глагол из этой клаузы, относящийся к той же референтной группе: *ištamaššer* (предыдущая клауза, *nu=tta apiya ŪL ištamaššer*), отсюда следует, что в рассматриваемом предложении он находится в топике. В фокусе, таким образом, остается только *kinun*, подчеркнутое фокусной частицей *=pat*. Контрастивность дополнительно доказывается наличием закрытого ряда альтернатив (Krifka 2007), и семантика контраста – замещающий фокус.

<sup>6</sup> Справедливости ради, Т. Е. Янко в разговоре с автором выразила сомнение, что интонационные особенности, характеризующие звуковой контур предложения и относящиеся к информационной структуре, можно выявить описанной процедурой в таких семантических деталях, как разновидности фокуса. Тем не менее наш опыт описания ИС на хеттском материале заставляет думать, что можно уверенно, то есть с обоснованием контекстом и синтаксическими особенностями, выделить не менее 5 разновидностей идентификационных фокусов (не-информационных), см. Молина 2018.

Пример анализа древнегреческого контекста с учетом фокусного скелета и QUD:

- (9) οὐ μὴν οὐδὲ λέληθε αὐτούς – εἰ γὰρ τινες καὶ ἄλλοι τὰ Περσέων νόμιμα ἐπιστέαται καὶ Αἰγύπτιοι – ὅτι [πρῶτα=μὲν]CT [νοθον]FOC [οὐ=σφι νόμος ἐστὶ βασιλεῦσαι γνησίου παρεόντος]TOP.

οὐ	μὴν	οὐ-δὲ	λέληθε	αὐτούς	εἰ
NEG	PTCL	NEG-PTCL	избегать.PERF.IND.ACT.3SG	3PL.M.ACC	если.COMP
γὰρ	τινες	καὶ	ἄλλοι	τὰ	Περσέων
EXPL	INDF.C.NOM.PL	также.ADV	другой.M.NOM.PL	ART.N.ACC.PL	перс.M.GEN.PL
νόμιμα	ἐπιστέαται	καὶ	Αἰγύπτιοι	ὅτι	
обычай.N.ACC.PL	знать.PRS.IND.MP.3PL	также.ADV	египтянин.M.NOM.PL	это.COMP	
[πρῶτα=μὲν]CT	[νοθον]FOC	[οὐ=σφι	νόμος		
первый.N.ACC.PL=PTCL	бастард.M.ACC.SG	NEG=3PL.DAT	обычай.M.NOM.SG		
ἐστὶ]TOP	[βασιλεῦσαι	γνησίου			
быть.PRS.IND.ACT.3SG	быть.царем.INF.AOR.ACT	законный.M.GEN.SG			
παρεόντος]TOP					
быть.рядом.PTCP.PRS.ACT.M.GEN.SG					

«(Египтяне) не могли не заметить, – ибо, если кто другой и знает также обычаи персов, то это египтяне, – что, во-первых, в их обычаи не входит, чтобы бастард становился царем, когда рядом есть законный наследник» (Goldstein 2016: 166; Геродот, «История»).

Во-первых, *πρῶτα=μὲν* ‘во-первых’ выступает как контрастивный топик: с одной стороны, контраст задается эмфазой, интонационным подчеркиванием, которое выделяет все подобные выражения в языках мира, и семантически «во-первых» входит в закрытый ряд альтернатив, ограниченный числом возможных доводов, из которого выбирается рассматриваемое значение. С другой стороны, это не та информация, которая является важной в предложении, то есть мы ни в коем случае не можем рассматривать смысл предложения как определяющийся выражением «во-первых». Топикальность подчеркивается частицей *=тен*, дающей дополнительные доказательства того, что мы верно выделили контрастивный топик. Отметим, неконтрастивные топики в предложении не выделяются топикальными частицами.

Далее, выражение *οὐ=σφι νόμος ἐστὶ* «не в их обычае» также находится в топике, выводимом из ближайшего контекста – мы видим выше и рассуждение о египтянах, и рассуждение о персидских законах (доступный топик). По рассуждению очевидно также, что описание самого обсуждаемого события, *βασιλεῦσαι γνησίου παρεόντος* «быть царем, когда есть законный наследник», тоже следует выводить из контекста, но более удаленного, не находящегося в том же предложении (выводимый топик). Отсюда становится ясно, что в фокусе может находиться лишь *νοθον* «бастард».

Вопрос к предложению (QUD) в данном случае не становится первым и единственным средством вычленять фокусные элементы, однако QUD нужно достраивать, чтобы проверить, нет ли в нашем рассуждении об информационной структуре предложения какого-либо нарушения смысла:

QUD: Что не входит в обычаи персов?

Ответ: Чтобы бастард становился царем, когда рядом есть законный наследник.

По ответу можно видеть, что сам по себе QUD не в состоянии автоматически выделить имеющийся в данном предложении узкий фокус на *νοθον*. Д. Голдстейн в данном контексте считает, что узкий фокус на *νοθον* является очевидным без дополнительного

анализа, и он, разумеется, в конкретном случае прав, – данный пример действительно имеет очевидную информационную структуру; в общем же случае мы вправе ожидать от автора объяснения, почему конкретное слово он считает находящимся в фокусе. Такое объяснение мы получаем, построив ряд фокусных альтернатив:

Торіс: «Во-первых», «в обычаи не входит», «рядом» [бастард, законный наследник]FOC – закрытый ряд альтернатив, однако не бинарный (в ряд может входить, например, «чужеземец» или какое-то еще противопоставление «законному наследнику»).

Заметим, что только по ряду фокусных альтернатив, как предлагала Худехебюре, мы также не можем восстановить узкий фокус на *voθov*. Поэтому мы полагаем, что только анализ топикальной структуры с учетом контекста, вкупе с формальной проверкой через QUD и набор фокусных альтернатив позволяет сузить рассуждение до относительно бесспорного выделения фокуса в предложении, взятом из текста на одном из корпусных языков. Построение карты фокусных альтернатив, в свою очередь, позволяет довольно детально понимать семантику рассматриваемого фокуса. В данном случае по закрытому ряду очевиден контраст, а не, допустим, фокус подтверждения или отрицания; это избирательный фокус «именно бастард, а не кто-то другой». С другой стороны, именно по ряду альтернатив мы можем увидеть, что «законный наследник», входящий в этот ряд, должен был бы быть в фокусе (и тогда следовало бы обсуждать сопоставительное выделение, или параллельный фокус), не будь разговора о нем в предыдущем контексте.

Итак, ряд альтернатив и семантическое значение фокуса мы определяем, построив ряд фокусных альтернатив, добавив к нему вышеописанный анализ QUD и, что важно, добавив к этому анализ набора топиков в предложении, анализируя их семантику и учитывая фиксированные позиции для разных видов топиков. Чтобы показать, как построение ряда альтернатив работает в случае бинарных рядов (фокуса подтверждения), обратимся снова к материалу хеттского языка. На примере (10) кажется очевидным, что, несмотря на отмеченные Т. Янко (личн. сообщ., Янко 2008) сложности с детализацией фокусных значений в условиях, когда анализ интонации невозможен, мы способны легко выделить такие разновидности фокусов, как контраст (см. пример 9 выше), фокус отрицания, фокус подтверждения (*verum*, пример 10):

(10) NH/NS (СТН 176) KUB 21.38 obv. 11'–12'

*ni* <sup>m</sup>Urḫi-<sup>d</sup>U-*upaš* *kuit* *apiya* *n=an* [*punuš*]<sub>focus</sub> *mān* [*kišan*]<sub>focus</sub> *mān* [*ŪL kišan*]<sub>focus</sub>

1. *ni* <sup>m</sup>[Urḫi-<sup>d</sup>]U-*upaš* *kuit* *apiya*

CONN Урхи-гешшуб поскольку там

«Поскольку Урхи-гешшуб там»

2. *n=an* *punuš-∅*

CONN=PRON.ACC.SG спросить-IMP.2SG

«Спроси его»

3. *mān kišan*

если так

«Так ли это»

4. *mān ŪL kišan*

если NEG так

«Или не так»

«Раз уж Урхи-гешшуб там, спроси его: так это или не так» («Письмо царицы Пудухебы фараону Египта Рамзесу II»; Hoffner 2009: 282;).

В этом контексте можно видеть два ряда фокусных альтернатив: {спроси его, не спрашивай его} {дело обстоит именно так; дело обстоит иначе}. Оба ряда закрыты, поскольку это контекст с идентификационным фокусом верификации/подтверждения, где ядро значения формируется вокруг подтверждения или опровержения предлагаемой семантической переменной. Клаузы 3 и 4 примера (10) обращаются к одному и тому же ряду альтернатив. При наличии такого контекста выделение фокусов и анализ их семантики выглядит однозначным.

## 7. Фокус и данное

Говоря о фокусе, мы не можем обойти вниманием проблему его взаимодействия с данным (*givenness*), которая, помимо прочего, создает много трудностей при разметке фокусных элементов для баз данных мертвых языков, а в целом, значительно затрудняет и без того непростой анализ информационной структуры при невозможности языковых экспериментов. Проблематика данного тщательно исследована уже Чейфом (Chafe 1976), а также многочисленными последующими исследователями; языки мира демонстрируют большой спектр средств обозначения того, что нечто, попавшее в непосредственный лингвистический контекст, уже ранее обсуждалось. Крифка (2007: 41) отмечает, что в языках мира обычно есть специальные анафорические средства выражения, имеющие анафору в составе своего лексического значения (такие как нулевые формы, клитики, личные местоимения, демонстративы, определенные артикли и т. п.); кроме того, анафора может выражаться изменением порядка слов или эллипсисом.

Поскольку составляющие, находящиеся в фокусе, обычно не являются данным и имеют большую значимость, в литературе предлагалось считать фокус комплементарным по отношению к данному. Однако фокус может оказаться в том числе и на данном, в том числе на личных местоимениях или демонстративах (последнее для хеттского материала хорошо показывает П. Худехебюре (2014)), особенно если в выражение, находящееся в фокусе, входит составляющая, являющаяся данным. При этом грамматические средства выражения фокуса, такие как акцент, в живых языках могут сместиться на другую составляющую, как в (11); деакцентуация, которую невозможно засечь при работе с материалом мертвого языка, оказывается в таком случае значимой, позволяя указать на данность находящегося в фокусе аргумента:

- (11) Перемещение акцента с аргумента на глагол, если аргумент оказывается данным:  
 A: I know that John stole a cookie. What did he do then?  
 B: He[reTURNed [the cookie]<sub>given</sub>]<sub>focus</sub>  
 (Krifka 2007:43)

В хеттском языке использование демонстративов непосредственно после первого упоминания референта, пишет П. Худехебюре (Goedegebuure 2014: 83, 208), должно восстановить связь с предыдущим контекстом, даже если референт имеет высокую значимость для дискурса и становится сентенциальным топиком. При этом демонстратив, отсылая к предыдущему контексту, указывает на наличие некоторого разрыва между контекстами.

- (12) MH/MS (CTH 262) IBoT 1.36 i 60–63  
 1. LÚ.MEŠMEŠEDU-TI=ma=kan DUMU<sup>MEŠ</sup> É.GAL-TI GAL-iaz KÁ.GAL-az katta ŪL paiškanda  
 2. n=at=kan lušdaniyaz katta paiškanda

3. *nu* 1 <sup>LÚ</sup>MEŠEDI *kuiš šarkantin widāizzi*
4. UGULA DUMU<sup>MEŠ</sup>.KIN-*za kuin parā piyēškezzi*
5. *nu=kan GAL-iaz katta apaš paišketta*

«1. Телохранители (и) дворцовые служащие, однако, не идут вниз из главных ворот. 2. Они идут вниз из бокового выхода. 3. Телохранитель, который привел жалобщика, (т. е.) 4. Тот, которого назначает глава курьерской службы. 5. (Именно) он проходит вниз через главные ворота» («Инструкция для царских телохранителей»; Miller 2013).

В примере (11.5), хорошо проанализированным Худехебюре (2009), с одной стороны, можно выделить ряд альтернатив, содержащий перечень всех видов телохранителей и слуг хеттского дворца, которые могут пройти через главные ворота, и переменную фокуса в клаузе 5 следует обозначать как «Через главные ворота проходит X». При этом QUD: «Кто проходит через главные ворота из дворцовых слуг и телохранителей?» – то есть ответ на вопрос фактически лежит в предыдущем контексте. Таким образом, демонстратив, находящийся в фокусе, отсылает к предыдущему контексту и является данным. Тем не менее можно видеть, что, несмотря на данность, здесь мы бы услышали интонационное выделение, будь это живой язык. Поэтому нельзя говорить, что такое выделение фокусом не является. Кроме демонстративов, в хеттском языке анафора может выражаться частицами, такими как *=(m)a*, маркирующими как топик, так и фокус (Sideltsev 2016: 78, с библиографией), и иногда строго фокусными частицами, такими как *=pat* (Молина 2018: 144, с библиографией). Последнее только лишний раз подтверждает, что анафора сама по себе, указывая на данность выражения, находящегося в фокусе, не является поводом считать, что фокуса на выражении нет.

Порядок слов также является одним из грамматических средств выражения фокуса. Если наличие данного в фокализованной группе вызывает передвижение акцента в живых языках, то для мертвых языков проверить интонационный контур мы не в состоянии; в то же время можно ожидать, что если в принципе при включении данного в группу фокуса может измениться грамматическое средство маркировки фокуса, то естественно ожидать изменений в порядке слов, с тем чтобы маркировать данность одной из составляющих в фокусе. Поэтому в работе (Молина 2018:79–114) мы уже высказывали предположение, что изменение порядка слов SO → OS может, в частности и в каких-то определенных случаях, быть вызвано необходимостью маркировать данность фокализованного компонента. Во всяком случае, такая идея не противоречит наблюдениям за живыми языками. Более того, М. Крифка в своей работе (2007) даже приводит примеры нейтрального порядка слов в английском как одного из вариантов маркирования данного.

## 8. Анафоричность фокуса

Из наблюдений за неоднозначным отношением фокуса к данному формулируется *гипотеза анафоричности фокуса*, рассмотренная М. Руттом (Rooth 1992, 2007) и поддержанная некоторыми другими исследователями (см. Schwarzschild 1999), согласно которой семантика и прагматика фокуса включают отношение к контексту, напоминающее анафору. Антецедентом фокуса в таком анализе фактически является QUD. Фокус – или оператор, интерпретирующий фокус, – лицензируется вопросом, потому что *фокусный скелет* непосредственно соотносится с QUD. Антецеденты фокуса могут быть разной степени эксплицированности. К примеру, на месте фокусной переменной в ближайшем

контексте может находиться X, заменяемый фокусом; и тогда Рут (2007) называет это фокусом замены (*substitution focus*, который легко соотносится с семантикой замещающего контраста):

13. [John wrote the report]<sub>2</sub>  
[No, [Mary<sub>focus2</sub> wrote it]]<sub>focus1</sub>

Семантическое значение QUD – набор отвечающих на этот вопрос пропозиций. Если этот набор содержится в дискурсе, мы получаем антецедент фокуса, эксплицированный в виде набора пропозиций (*alternative-set focus*), но он может быть имплицитным:

14. [Mary spoke to someone about his problems]<sub>2</sub>  
[Yeah, [she spoke to John<sub>focus2</sub> about his problems]]<sub>focus1</sub>

Набор X пропозиций, отвечающих на QUD, генерируется путем создания всех возможных значений для переменной в *фокусном скелете*. Ограничение на антецедент *alternative-set focus* таково, что множество фокусных альтернатив должно включать не менее 2 единиц. Вопрос, формирующий антецедент фокуса, задает набор Y пропозиций так, чтобы  $Y \subseteq X$ , и если оно удовлетворяется, высказывание лицензировано.

Другими словами, гипотеза анафоричности фокуса сводится к тому, что фокус можно назвать *анафорой, антецедентом которой является вопрос к предложению (QUD), задающий ограничение на выбор из фокусных альтернатив*. Здесь возникает естественный вопрос, чем фокус отличается от топики. Мы предполагаем, что ответ на данный вопрос лежит в степени известности (данности) адресату, и существует некая шкала перехода от более известного к менее известному, в определенной точке которой топики и фокусы семантически сходятся. К сожалению, такие тонкости на материале мертвых языков вычленять уже не представляется возможным. Однако гипотеза анафоричности фокуса позволяет работать со сложным материалом, позволяя разграничить понятия «анафора» и «фокус», «данное» и «фокус», которые иногда смешиваются работах по синтаксису мертвых языков.

Так, (Sideltsev 2018) предполагает для хеттского раннюю стадию лексикализации выражений типа *apāš=pat* «тот же», опираясь на частотность встречаемости данного сочетания в хеттском корпусе дипломатических текстов и на соображение, что выражение безусловно имеет анафорическую функцию, хотя и маркируется фокусной частицей. Демонстративы, несомненно, задают ряд фокусных альтернатив, соответствующий избирательному либо заместительному фокусу («именно этот/тот», «не тот, а этот»). Эти контрастивные фокусы чаще всего и маркируются с помощью *=pat*, чем и обусловлена отмеченная частотность данных сочетаний. Однако на наших примерах, как только фокус перемещается с указательных местоимений на другое слово, частица *=pat* смещается вместе с фокусом, как показано в примерах (15, 16, 17 – Hoffner 2009: 283, 307, 326).

- (15) NH/NS (СТН 181) KUB 14.3 ii 28  
*apā-š=ma nūwa=pat mem[i-ške-t ...]*  
тот-NOM.SG=же всё.еще=же говорить-IMF-3SG.PST  
«Тот же по-прежнему продолжал говорить».

- (16) NH/NS (СТН 178) KUB 23.103 rev. 22  
*ni apā-š kinun=pat kuit memi-ške-zzi [...]*  
CONN тот-NOM.SG сейчас=же поскольку говорить-IMF-3SG.PRS  
«Поскольку он (даже) сейчас продолжает говорить».



(17) NH/NS (СТН 176) KUB 21.38 obv. 20'

*ni ap-ē[daš ammuk]=pat LÚ.MEŠ ṬE₄ME DUB.BA.A=ya AŠPUR*  
 CONN тот-DAT.PL я=же курьер.ACC.PL табличка.ACC.PL=и послать.1SG.PST  
 «Я (сам) послал им курьеров и таблички».

В примере (15) =*pat* подчеркивает узкий контрастивный (counterexpectant) фокус на *nīwa*: «Он *говорил* и *всё ещё* продолжает *говорить*». Данный контекст взят из так называемого «Письма Тавагалавы», где обсуждаются отношения Пиямарады, деятеля эпохи хеттских царей поздней империи – Хаттусили III (ок. 1267–1237 г. до н. э.), Мурсили III (Урхи-Тешшуб) и Муваталли II (ок. 1295–1272 г. до н. э.), см. подробнее (Hoffner 2009: 298–301), – с Аххиявой и собственно Хаттусой. Предыдущий контекст говорит нам о том, что Пиямарада писал о своих страхах, и ему обещали прислать «одного правителя или брата» (*ni=wa kāša* 1-EN BEL[U...] *našma=wa ŠEŠ uiyami*). Однако тот, вопреки ожиданиям, продолжает писать о своих страхах (*nīwa=pat mem[išket...]*). Семантику данного контекста предлагается рассматривать в следующем виде:

- а) обычное семантическое значение (semantic value)  $[[p]]^o$ ,  $p$  = («он говорит»);
- б) фокусное значение  $[[p]]^f$ , вводимое чувствительной к фокусу (focus sensitive) частицей *nīwa* ‘всё ещё’ и подчеркиваемое фокусной частицей =*pat*;
- в) ожидаемая пресуппозиция  $q$  = («он перестал говорить»);
- г)  $[[p]]^f$ -  $[[p]]^o$  возвращает значение ложности  $q$ , вопреки ожидаемому (counterexpectant focus).

В примере (16) выделяется избирательный фокус на *kinun* ‘сейчас’ («именно сейчас», ряд фокусных альтернатив представляет собой временные значения). В примере (17) узкий фокус на местоимении 1-го л. *ammuk* ‘я сам’: использование эксплицитно выраженного местоимения, как уже говорилось выше, подчеркивает наличие фокуса. Кажется, что приведенные примеры вполне доказывают отсутствие лексикализации выражения *apaš=pat* на новохеттском материале. Принятие гипотезы анафоричности фокуса, таким образом, снимает необходимость объяснять, почему анафора маркируется фокусными частицами, то есть совпадает с фокусом.

## 9. Выводы

Итак, очевидно, что и древнегреческий, и хеттский материал вполне поддаются анализу на предмет выявления фокусных элементов и определения их семантики. Мы показали, что такой анализ возможен лишь при комплексном подходе, включающем в себя как анализ фокусного скелета (выделение ряда фокусных альтернатив, определение его объема и типа), так и анализ QUD, в том числе для проверки кажущихся очевидными утверждений о содержании фокусов. Что более существенно, такой анализ должен предваряться анализом топикальных элементов, исходя из а) доступного, непосредственно предваряющего предложение контекста (доступные топики); б) контекста, находящегося в common knowledge собеседников (выводимые топики); для такого анализа требуется привлечение связанного материала, например, предыдущих серий писем в случае с хеттским языком, либо трудов, приписываемых тому же автору, в случае с древнегреческим; в) контекста, находящегося в common background (установленные топики; информация, которая по умолчанию является известной собеседникам); для такого анализа необходимо учитывать культурологические и/или исторические комментарии к исследуемым текстам. Именно детальный анализ фокусных элементов в текстах на хетт-

ском и древнегреческом языках, как наиболее древних зафиксированных в письменности языках индоевропейской семьи, приблизит нас к пониманию того, как могла выглядеть информационная структура предложения в праиндоевропейском.

### Литература

- Брызгунова, Е. А. 1980. Интонация. В: Н. Ю. Шведова (ред.). *Русская грамматика*: 96–122. Москва: Наука.
- Молина, М. А. 2016. Эмфатическая энклитика =*rat* в хеттском языке: анализ функций и семантика маркируемых фокусов. В: Н. Н. Казанский (ред.). *Индоевропейское языкознание и классическая филология – XX. Чтения памяти проф. И.М. Тронского, 20–22 июня 2016 года*: 739–754. Санкт-Петербург: Наука.
- Молина, М. А. 2018. *Порядок слов в хеттском языке: корпусный подход и анализ в типологической перспективе*. Дисс. исследование. Москва: Институт языкознания РАН.
- Падучева, Е. В. 2015. *Коммуникативная структура предложения*. Материалы для проекта корпусного описания русской грамматики (<http://rusgram.ru>). На правах рукописи. Москва.
- Сидельцев, А. В. 2017. *Синтаксис простого предложения в хеттском языке*. Москва: Языки славянской культуры.
- Шкапа, М. В. 2013. Клефт в ирландском языке: к типологии клефта и тетических предложений. *Вопросы языкознания* 5(2013): 89–105.
- Янко, Т. Е. 2008. *Интонационные стратегии русской речи в сопоставительном аспекте*. Москва: Языки славянских культур.

### References

- Barðdal, Johanna. 2013. Construction-based historical-comparative reconstruction. In: T. Hoffmann, G. Trousdale (eds.). *Oxford Handbook of Construction Grammar*: 438–457. Oxford University Press.
- Barðdal, Johanna, Thomas Smitherman. 2009. Typological changes in the evolution of Indo-European syntax? *Diachronica* 26(2): 253–263.
- Bauer, Brigitte L.M. 2000. *Archaic syntax in Indo-European: the spread of transitivity in Latin and French*. Berlin: Mouton de Gruyter.
- Beaver, David I., Brady Clark. 2008. *Sense and sensitivity: how focus determines meaning*. Malden, MA / Oxford: Blackwell.
- Bryzgunova, Elena A. 1980. Intonatsiya. In: N. Yu. Shvedova (ed.). *Russkaja grammatika*: 96–122. Moskva: Nauka.
- Büring, Daniel. 1997. *The meaning of topic and focus. The 59th Street Bridge accent*. London: Routledge.
- Büring, Daniel. 1999. Topics. In: P. Bosch, R. van der Sandt (eds.). *Focus – Linguistic, Cognitive and Computational Perspectives*: 142–165. Cambridge University Press.
- Büring, Daniel. 2003. On D-trees, beans and B-accent. *Linguistics and Philosophy* 26: 511–545.
- Büring, Daniel. 2006. Intonation und Informationsstruktur. In: H. Blühdorn, E. Breindl, U. H. Wassner (eds.). *Text – Verstehen. Grammatik und darüber hinaus*: 144–163. Berlin: De Gruyter.
- Carlson, Lauri. 1983. *Dialogue games. An approach to discourse analysis*. Dordrecht: Reidel.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: Charles N. Li (ed.). *Subject and Topic*: 25–55. New York: Academy Press.
- Couper-Kuhlen, Elizabeth. 2001. Intonation and discourse: current views from within. In: D. Schiffrin, D. Tannen, H. E. Hamilton (eds.). *The Handbook of Discourse Analysis*: 13–34. Oxford: Blackwell.
- Dik, Helma P. 1995. *Word order in Ancient Greek: a pragmatic aspect of word order variation in Herodotus*. Amsterdam: Gieben.
- Ginnakis, Giorgios. K. 1997. *Studies in syntax and semantics of the reduplicated presents of Homeric Greek and Indo-European*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Ginzburg, Jonathan. 1995. Resolving questions. *Linguistics and Philosophy* 18: 459–527.
- Goedegebuure, Petra. 2009. Focus structure and Q-words questions in Hittite. *Linguistics* 47: 945–969.
- Goedegebuure, Petra. 2014. *The use of demonstratives in Hittite: deixis, reference and focus* (StBoT 55). Wiesbaden: Harrassowitz.

- Goldstein, David. 2012. Iterated modal marking and polarity focus in Ancient Greek. *Transactions of the Philological Society* 110 (2012): 1–25.
- Goldstein, David. 2012. *Classical Greek syntax: Wackernagel's law in Herodotus*. Leiden / Boston: Brill.
- Hatcher, Anna G. 1956. Theme and underlying question: two studies of Spanish word order. *Word (New York) Supplement, Monograph 3*: 234–250.
- Haugan, Jens. 2001. *Old Norse. Word order and information structure*. PhD Diss. Trondheim: Norwegian University of Science and Technology.
- Hoffner, Harry A. 2009. *Letters from the Hittite Kingdom*. Atlanta: Society of Biblical Literature.
- Hoffner, Harry A., Craig H. Melchert. 2008. *A grammar of the Hittite language*. Eisenbrauns.
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, Mass.: MIT Press.
- Kortlandt, Frederik H. H. 1983. Proto-Indo-European verbal syntax. *Journal of Indo-European Studies* 11: 307–324.
- Krifka, Manfred. 2007. Basic notions of information structure. In: C. Fery, G. Fanselow, M. Krifka (eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS)* 6. Potsdam, Universitätsverlag Potsdam: 13–56.
- Kulikov, Leonid, Nikolaos Lavidas (eds.). 2015. *Proto-Indo-European syntax and its development*. Amsterdam / Philadelphia: John Benjamins.
- Kvičala, Johann. 1870. Untersuchungen auf dem Gebiete der Pronomina, besonders der Lateinischen. *Sitzungsbericht Der Österreichischen Akademie Der Wissenschaften. Philosophisch-Historische Klasse* 65: 77–155. Wien: Gerold.
- Miller, Jared L. 2013. *Royal Hittite instructions and related administrative texts*. Atlanta: Society of Biblical Literature.
- Molina, Maria A. 2016. Emfaticeskaja enklitika =pat v khettskom jazyke: analiz funktsij i semantika markirujemykh fokusov. In: N. N. Kazanskiy (ed.). *Indojeuropejskoje jazykoznanije i klassičeskaja filologija – XX. Čtenija pam'ati prof. I. M. Tronskogo, 20–22 June 2016*: 739–754. Sankt-Peterburg: Nauka.
- Molina, Maria A. 2018. *Por'adok slov v khettskom jazyke: korpusnyj podkhod i analiz v tipologičeskoj perspektive*. Diss. thesis. Moscow: Institute of Linguistics (Russian Academy of Sciences).
- Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. 5 ed. Tübingen: Niemeyer.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In: P. Cole (ed.). *Radical Pragmatics*: 223–256. New York: Academic Press.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Lambrecht, Knud. 1994. *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge University Press.
- Luraghi, Silvia. 2004. Null objects in Latin and Greek and the relevance of linguistic typology for language reconstruction. In: K. Jones-Bley, M. Huld, A. Della Volpe, M. R. Dexter (eds.). *Proceedings of the 15<sup>th</sup> Annual UCLA Indo-European Conference (=JIES Monograph 49)*: 234–256. Washington DC: Institute for the Study of Man.
- Luraghi, Silvia. 2012. Basic valency orientation and the middle voice in Hittite. *Studies in Language* 36(1): 1–32.
- Mallory, James P., Douglas Q. Adams. 1997. *Encyclopedia of Indo-European culture*. London: Fitzroy-Dearborn.
- Meelen, Marieke. (In print). Annotating Middle Welsh: POS tagging and chunk-parsing a partial corpus of native prose. In: E. Lash E. (ed.). *Proceedings of the Chronologicon Hibernicum workshop on Morphological and Syntactic Variation*.
- Paducheva, Elena V. 2015. *Kommunikacionaja struktura predlozhenija. Materialy dl'a proekta korpusnogo opisanija russkoj grammatiki (<http://rusgram.ru>)*. Ms.
- Riester, Arndt, Lisa Brunetti, Kordula De Kuthy. 2018. Annotation guidelines for questions under discussion and information structure. In: E. Adamou, K. Haude, M. Vanhove (eds.). *Information structure in lesser-described languages: Studies in prosody and syntax. (Studies in Language Companion Series 199)*: 403–444. Amsterdam: John Benjamins.
- Roberts, Craige. 2012. Information structure: towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5: 1–69.
- Rooth, Mats. 1985. *Association with focus*. PhD diss. Amherst: University of Massachusetts.
- Rooth, Mats. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1: 75–116.
- Rooth, Mats. 1996. On the interface principles for intonational focus. In: T. Galloway, J. Spence (eds.). *Proceedings from the Sixth Conference on Semantics and Linguistic Theory*: 202–226. Ithaca: CLC Publications.
- Rooth, Mats. 1999. Association with focus or association with presupposition? In: P. Bosch, R. van der Sandt (eds.). *Focus: Linguistic, cognitive, and computational perspectives*: 232–244. Cambridge University Press.

- Rooth, Mats. 2007. Notions of Focus Anaphoricity. In: C. Féry, G. Fanselow, M. Krifka (eds.). *Interdisciplinary Studies on Information Structure, Working Papers of the SFB 632 (Universitätsverlag)*: 57–68. Postdam: Universität-Verlag.
- Schwarzschild, Roger. 1999. Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2): 141–177.
- Shkapa, Maria V. 2013. Kleft v irlandskom jazyke: k tipologii klefta i teticheskikh predlozhenij. *Voprosy jazykoznanija* 5(2013): 89–105.
- Sideltsev, Andrej V. 2016. Hittite anaphora: two case studies. *Archiv Orientalni* 84: 75–104.
- Sideltsev, Andrej V. 2017. *Sintaksis prostogo predlozhenija v khettskom jazyke*. Moskva: Jazyki slav'anskoj kul'tury.
- Sideltsev, Andrej V. 2018. Hittite *-pat* and *imma* as focus particles: lexicalisation. *Altorientalische Forschungen* 45(2): 1–13.
- Ünal, Ahmet. *Hattušili III. Vol. II: Quellen. Texte der Hethiter* 4. Heidelberg: Universitätsverlag Winter.
- Watkins, Calvert. 1964. Preliminaries to the reconstruction of Indo-European sentence structure. In: H. G. Lunt (ed.). *Proceedings of the 9<sup>th</sup> International Congress of Linguists*: 1035–1045. Cambridge / The Hague: Mouton.
- Yanko, Tatiana E. 2008. *Intonatsionnyje strategii russkoj rechi v sopostavitel'nom aspekte*. Moskva: Jazyki slav'anskoj kul'tury.

*Maria Molina. On the methodology of defining focus in an extinct language: towards a reconstruction of Proto-Indo-European syntax*

In this paper, I cover the problem of topic and focus revealing in the sentences of Hittite and Ancient Greek, two of the earliest Indo-European languages attested in writing. Information structure revealing gives us clues not only to the placement of accent (which is highly relevant for Hittite phonology), but also to the reconstruction of Indo-European syntax. Current methods of information structure analysis in dead languages remain insufficiently developed, with seemingly arbitrary procedures of focus revealing that lack proper argumentation. The paper suggests using a complex approach, subsequently applying several methods that have been tested on living languages, but so far have not yielded positive results on data from extinct languages on their own: namely, the method of Focus Skeleton (identification of a row of focus alternatives together with topic analysis), and the method of Question under Discussion, both of them applied in combination with contextual and philological analysis.

*Key words:* Hittite language, Ancient Greek language, Proto-Indo-European, syntax reconstruction, information structure, focus, topic, semantics of focus