

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

№19/3



Moscow 2021

Российский государственный гуманитарный университет
Институт языкознания Российской Академии наук

Вопросы языкового родства

Международный научный журнал

№ 19/3



Москва 2021

Advisory Board:

H. EICHNER (Vienna) / Chairman
W. BAXTER (Ann Arbor, Michigan)
V. BLAŽEK (Brno)
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
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)

Ф. КОРГЛАНДТ (Лейден)

А. ЛУБОЦКИЙ (Лейден)

Дж. МЭЛЛОРИ (Белфаст)

А. Ю. МИЛИТАРЕВ (Москва)

Л. ХАЙМАН (Беркли)

Редакционная коллегия:

В. А. ДЬБО (главный редактор)

Г. С. СТАРОСТИН (заместитель главного редактора)

Т. А. МИХАЙЛОВА (ответственный секретарь)

А. В. ДЬБО

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., 2021. — № 19/3. — x + 79 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, 2021. — No. 19/3. — x + 79 p.

ISSN 2219-3820

<http://www.jolr.ru/>
gstarst1@gmail.com

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

Подписано в печать 28.09.2021. Формат 60×90/8.
Усл. печ. л. 11,3. Уч.-изд. л. 7,5. Заказ № 1445.
Тираж 1050 экз.

Издательский центр
Российского государственного гуманитарного университета
125047, Москва, Миусская пл., 6
www.rggi.ru

Table of Contents / Содержание

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

Articles / Статьи

<i>Anton Kogan</i> . Towards the reconstruction of language contact in the pre-Tibetan Upper Indus region	153
[<i>А. И. Коган</i> . К реконструкции языковых контактов в регионе верховьев Инда в дотибетскую эпоху]	
<i>Paul Sidwell, Mark Alves</i> . The Vietic languages: a phylogenetic analysis.	166
[<i>Пол Сидвелл, Марк Алвес</i> . Вьетские языки: филогенетический анализ]	
<i>А. С. Крылова</i> . О возможном сохранении реликтов ведийского глагольного ударения в говоре деревни Наггар языка куллуи	195
[<i>Anastasiya Krylova</i> . On the possible preservation of Vedic verbal stress in the Naggar subdialect of the Kullui language]	
<i>А. А. Трофимов</i> . Ударение множественного числа имен с показателем <i>-ān</i> в парачи и его связь с праиранской акцентуацией	210
[<i>Artem Trofimov</i> . Stress position of nominal plurals in <i>-ān</i> in the Parachi language and its connection with the accent in Proto-Iranian]	

Contributors

Mark Alves — professor, Department of English Language for Academic Purposes, Linguistics and Communication Studies, Montgomery College, Rockville, Maryland, Mark.Alves@montgomerycollege.edu

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

Anastasiya Krylova — researcher, Department of Asian and African languages, Institute of Oriental Studies of the Russian Academy of Sciences (Moscow), krylova_anastasi@bk.ru

Paul Sidwell — Honorary Associate, Dept. of Linguistics and Communication Studies of the School of Literature, Art and Media, University of Sydney, paulsidwell@gmail.com

Artem Trofimov — researcher, School for Advanced Studies in the Humanities, RANEPA (Moscow), artemii.trofimov@gmail.com

Сведения об авторах

Алвес, Марк — профессор колледжа Монтгомери (Роквилль, Мэриленд, США), Mark.Alves@montgomerycollege.edu

Коган, Антон Ильич — доктор филол. наук, науч. сотрудник Отдела языков Азии и Африки Института востоковедения РАН (Москва), kogan_anton@yahoo.com

Крылова, Анастасия Сергеевна — научный сотрудник Института востоковедения РАН (Москва), krylova_anastasi@bk.ru

Сидвелл, Пол — почетный профессор Школы литературы, искусств и медиа Сиднейского университета, paulsidwell@gmail.com

Трофимов, Артем Александрович — научный сотрудник Лаборатории востоковедения и сравнительно-исторического языкознания ШАГИ РАНХиГС (Москва), 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. Articles are published preferably in English or Russian, although publication of texts in other major European languages (such as French or German) may be considered by the Editorial Board.

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

For more detailed guidelines on article submission and editorial policies, please see our website (<http://jolr.ru>) or address the Editorial Board directly at gstarst1@gmail.com. Inquiries may also be sent by regular mail to the official address of the Journal:

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

Будущим авторам

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

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

Подробнее о требованиях к оформлению рукописи, редакционной политике журнала и т. п. вы можете узнать на нашем сайте по адресу: <http://www.jolr.ru> или непосредственно от редакционной коллегии по электронной почте (gstarst1@gmail.com). По различным вопросам с редакцией журнала можно также связаться по обычной почте:

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

Towards the reconstruction of language contact in the pre-Tibetan Upper Indus region

The present article attempts to partially restore the overall picture of language contact in the historical region of Zhangzhung (northwestern Tibet) prior to its conquest by the Tibetans in the late 1st millennium A.D. The pre-Tibetan language of Zhangzhung, known chiefly from a number of fragments in medieval Tibetan texts, belongs to the West Himalayan group of the Tibeto-Burman branch of the Sino-Tibetan language family. However, the vocabulary of this language contains a significant Indo-Iranian etymological stratum. The author shows that certain words belonging to this stratum have exact cognates in Tibetan dialects of Ladakh and Baltistan. Based on the analysis of Indo-Iranian elements in the Zhangzhung language, it is concluded that the pre-Tibetan population of Zhangzhung was likely to have been in a fairly close contact with Aryan, probably Dardic-speaking inhabitants of areas further down the Indus.

Keywords: language contact; lexical borrowing; Indo-Iranian languages; Dardic languages; Sino-Tibetan languages; Zhangzhung language.

Introduction

The uppermost part of the Indus basin, nowadays populated mainly by speakers of Tibetan dialects, in the past was, no doubt, a multiethnic and multilingual region. As I have tried to show in my previous publications (Kogan 2019; 2020), the inhabitants of certain areas within this region (primarily, of the present-day Ladakh) must have spoken an Indo-Iranian language, most probably belonging to the Dardic group. It has also been hypothesized that this language was the substratum of Tibetan varieties spoken in Ladakh today. This assumption, however, was made largely by process of elimination. Based on certain historical-phonological features of Indo-Iranian loanwords in Ladakhi, Purik and Balti, as well as on some historical facts, it was demonstrated that the substrate influence on these dialects is much more likely than the adstrate or superstrate one (Kogan 2020). In this kind of situation, it is highly desirable to find some additional positive evidence corroborating the substratum hypothesis.

Technically, search for such evidence can be successful only when the ethnic and linguistic composition of the area in question prior to the Tibetan conquest (i.e. before the 8th century A.D.) is reconstructed in sufficient detail. In our case, valuable material for the reconstruction can be provided not only by borrowed vocabulary but also by foreign fragments available in some medieval Tibetan texts. The language of these fragments is usually considered to be spoken before as well as some time after the Tibetan invasion, in an area to the immediate south-east of Ladakh, up the Indus. This area, nowadays forming a part of the Ngari prefecture of China's Tibet Autonomous Region, in the past was known as Zhangzhung.¹ It is traditionally believed that Zhangzhung was an independent kingdom until it was conquered by the Tibetans in the mid-7th century A.D., although its language of the same name persisted for several centuries after the conquest. Almost all the extant specimens of this language are

¹ Alternative spellings are Zhang-Zhung and Shangshung.

Zhangzhung-Tibetan bilinguals, mostly bilingual titles, occurring in sacred texts of the Bon religion.²

Today it is generally accepted among scholars that Zhangzhung belongs to the Sino-Tibetan language family, where it is classified with the West Himalayish branch of Tibeto-Burman languages (Haarh 1968; Matisoff 2001; van Driem 2001; Widmer 2014),³ its closest relatives being modern dialects spoken in Indian Himalaya: Darmiya, Byangsi, Chaudangsi, Bunan, Pattani (Manchad) and Kinnauri.⁴ That said, however, it has long been noted that Zhangzhung vocabulary contains an Indo-Iranian etymological stratum. The author of the first substantial and systematic description of the Zhangzhung language, the Danish Tibetologist Erik Haarh pointed out that a number of Zhangzhung words are “related to Sanskrit or some North-Indian languages” (Haarh 1968: 13). Sanskrit loanwords are relatively numerous in Zhangzhung fragments, almost all of them being religious and philosophical terms. Their borrowing dates back to the period after the 11th century A.D., when due to the growing influence of Buddhism the Bon religion was actively absorbing Buddhist concepts and terminology.⁵ Such words are, for the most part, easily identifiable and etymologizable. This, however, is not the case with loans from other Indo-Iranian source(s), termed by Haarh “words related to... some North-Indian languages”. On the one hand, in every hitherto published specimen of Zhangzhung lexicon, starting with Haarh’s rather short glossary,⁶ one can find obvious examples of lexical items for which Indo-Iranian origin seems most likely, but which for historical-phonological and/or semantic reasons can hardly be derived from Sanskrit. On the other hand, apart from a few isolated etymological suggestions, that will be discussed below, no systematic research on such items has been thus far conducted with the view to detecting and etymologizing all of them. Meanwhile, recent progress in Zhangzhung linguistic studies, notably the publication of several substantial dictionaries, has made this kind of research feasible.

This article presents an attempt at preliminary etymological analysis of possible non-Sanskrit Indo-Iranian loanwords in Zhangzhung. The obtained results are listed in the following section. My main data source was Dan Martin’s “Zhangzhung dictionary” (Martin 2010). To date, it is the newest and one of the most extensive dictionaries of the Zhangzhung lan-

² Bon is a religious tradition of obscure and still disputed origin, considered to predominate in Tibet before the spread of Buddhism, and to be greatly transformed subsequently under the latter’s influence. It is also believed that the Bon religion was brought to Tibet from Zhangzhung. For this reason, the Zhangzhung language was regarded as sacred by adherents of Bon. A large number of Bon cultic texts are traditionally deemed to be translated from Zhangzhung into Tibetan. Titles of such texts are always given in both languages. The only extant Zhangzhung written document of significant length is the religious scripture named “The innermost treasury of existence” (Tib. *Srid-pa’i mdzod-phug*). It dates back to the 11th century A.D. and is almost entirely bilingual. For more on this text and its importance as a source of Zhangzhung vocabulary see, e.g. Martin 2000; Mitruiev 2016.

³ According to an alternative hypothesis, the population of Zhangzhung originated, at least partly, from Eastern or Northeastern Tibet, and the Zhangzhung language should be classified with the Qiangic branch of Tibeto-Burman (see, e.g. Hummel 1986). This view, however, was found to be ill-founded (Jacques 2008).

⁴ Nowadays the speaking-areas of these dialects are parts of Himachal Pradesh and Uttarakhand states of India.

⁵ Contrary to literary Tibetan, where calquing was the preferred way of adopting Buddhist vocabulary, Zhangzhung seems to have tended to borrow Buddhist terms phonetically from Sanskrit. It is not unlikely that most of such terms appeared in the language after it had ceased to be spoken as a mother tongue. A remarkable fact in this connection is that in the above-mentioned early (11th century) Bon text “The innermost treasury of existence” no undoubted Sanskritisms seem to be present.

⁶ Haarh’s description of the Zhangzhung language, comprising grammatical notes and Zhangzhung-English glossary, is based on data from “Tibetan-Zhangzhung dictionary” (Bon-po Association 1965), a collection of Zhangzhung phrases with Tibetan translations published by Tibetan immigrants in India.

guage with the total number of entries exceeding 3500.⁷ In addition, “A lexicon of Zhangzhung and Bonpo terms” compiled by Pasar Tsultrim Tenzin, Changru Tritsuk Namdak Nyima and Gatsa Lodroe Rabsal (Tenzin et al. 2008) was sometimes consulted because it contains some words, absent from Martin’s dictionary. A particular word was selected for analysis if it defies etymologizing in Sino-Tibetan,⁸ meeting at the same time at least one of the two criteria: (i) presence of a plausible Indo-Iranian etymology without being derivable directly from Sanskrit; (ii) presence of possible cognates in Indo-Iranian lects spoken, nowadays or in the past, in areas contiguous to historical Zhangzhung. These lects include primarily Dardic languages, as well as Indo-Aryan forms of speech current in Western and Central Himalaya.⁹ One should bear in mind that either of the above criteria can be met without satisfying the other one. A lexical item borrowed into Zhangzhung from a certain Indo-Iranian source can ultimately have a non-Indo-Iranian (and non-Indo-European) origin, being, e.g., a Dravidian loan in Indo-Aryan or a Burushaski loan in Dardic. And vice versa: a Zhangzhung word may have doubtless cognates in ancient Indo-Iranian languages and in other branches of Indo-European, but be absent from neighboring Dardic and Indo-Aryan dialects, which will make its immediate source unidentifiable. In both cases, however, we are quite justified in postulating the borrowing of a particular lexeme from a particular Indo-Iranian lect into the Zhangzhung language.

Etymologies on the list are numbered consecutively and arranged in alphabetical order. Following Martin 2010, I usually specify Tibetan equivalents of Zhangzhung lexical items. They are given in brackets after entry words.

A list of probable Indo-Iranian loanwords in Zhangzhung

1. *a da ra* (*lhad med pa’am dag pa*) ‘uncontaminated or pure’ (Martin 2010: 244). It is tempting to compare the first syllable with the Indo-Iranian privative particle (cf. OIA, Av. *a-*, *an-* ‘not, un-, non-’). In case this etymology is correct, the initial *a* of the Zhangzhung word should be considered an equivalent of *med pa* in the Tibetan gloss, whereas its remaining part (*da ra*) must have been semantically similar or identical to Tibetan *lhad* ‘a baser substance mixed with a finer one, an alloy’ (Jäschke 1881: 600). Zhangzhung *da ra* may have been an Aryan loanword reflecting some derivative, e.g. a thematic verbal noun, of PIE **dher-/dhrei-d-* ‘dirt, feces; to defecate’.¹⁰ The reflexes of this root are well attested in East Iranian, where they mostly mean ‘dung, manure’,¹¹ the meaning ‘dirt’, however, also existing in certain languages (cf. Sogdian

⁷ One of the major advantages of this dictionary is that it uses numerous sources of lexical data, including “The innermost treasury of existence”, the earliest as well as the longest Zhangzhung text, composed when the language was probably still alive.

⁸ For Sino-Tibetan etymologies I consulted two online databases: “Sino-Tibetan etymology” compiled by Sergei Starostin (Starostin 2005), and “The Sino-Tibetan Etymological Dictionary and Thesaurus” compiled by a group of scholars led by James A. Matisoff (Matisoff et al. 2015). The above-cited works (van Driem 2001, Matisoff 2001 and Widmer 2014), as well as Martin’s dictionary, were also used as important sources of comparative data.

⁹ Traditionally they are classified with West and Central Pahari subgroups of Indo-Aryan. It should be noted that along with Indo-Iranian and Dardic, Iranian influence on Zhangzhung cannot be ruled out either. This influence may have come, e.g. from some Saka dialects spoken in what is now East Turkestan.

¹⁰ In Pokorny’s dictionary this root is given with the meaning ‘*Unrat, cacāre*’, and its possible etymological connection with PIE **dher-* ‘*trüber Bodensatz einer Flüssigkeit*’ is postulated (Pokorny 1959: 251, 256).

¹¹ Cf., e.g. Shughni *ḍīd* ‘(pressed) dung’, Sariqoli *ḍig* ‘cow dung’, Yazghulami *ḍāg*, Wakhi *ḍart* ‘cow dung, manure’. The prototype for these words is usually reconstructed as **darta-*, **darti-* or **dṛti-*. For details, see Rastorgueva, Edelman 2003: 356-357.

ḍrt'yč(h) ‘dirt, dung’). The root in question seems to have been preserved also in Dardic, cf. Kalasha *darāli gaṇ* ‘anus’¹² (Trail, Cooper 1999: 73).

2. *ag* (*sha za, srin*) ‘rākṣasa’ (Martin 2010: 245-246). Cf. OIA, Av. *aka-* ‘evil; pain, suffering’, Persian *āk* ‘disaster, adversity; fault’. For semantic development cf. OIA *rākṣasa-* ‘demon, ogre, fiend’ being a vṛddhi-derivative of *rakṣas-* ‘injury, harm; malignant demon’.

3. *'bar sha* ‘a sharp weapon’ (Tenzin et al. 2008: 122). Cf. Hindi, Punjabi *barchā*, Nepali *barcho*, Sindhi, Gujarati *barchī*, Marathi *barcī, barsī*, Bengali *barchi* ‘spear’, West Pahari *borcho* (poet.) ‘long stick with an iron tip, spear’, Persian *barča* ‘a kind of spear’. Turner compares the West Pahari form with OIA *baḍīśa-* ‘hook, fish-hook’ (Turner 1966: 514) but phonological correspondences are irregular. Apparently, a Wanderwort of unclear origin.

4. *be ling* (*rlung*) ‘wind, air’ (Martin 2010: 151). The word seems to be a hybrid compound consisting of two synonymous components. The second component is, no doubt, the basic Zhangzhung word for wind.¹³ The first element may be compared either with OIA *vāyu-* ‘wind, air, god of wind’, Av. *vaiiu-* ‘air, atmosphere, name of a god’, Kashmiri *wāw*, Sindhi *vāu*, Lahnda, Gujarati *vā*, Bengali *bāo*, Marathi *vāv* ‘wind’, or with OIA, Av. *vāta-*, Assamese *bā*, Bengali *bā, bāy*, Kumauni *bai*, Persian *bād*, Ossetic *wad* ‘wind’.

5. *bha gi* (*mngon par*) ‘visibly’ (Martin 2010: 154). Cf. OIA *bhāti*, Av. *fra-uuāiti* ‘shines’, OIA *bhāsati* ‘is bright’, *bhāsa-* ‘light’, Kashmiri *bāsun* ‘to seem, appear, become visible’, Punjabi *bhāh* ‘fire, splendor’, Hindi *bhāsnā* ‘to shine’. The exact prototype and phonological development remains unclear.

6. *chal* (*yal*) ‘decrease, lower, cause diminution, fade away’ (Martin 2010: 77). Cf. OIA *kṣarati, kṣalati* ‘flows, trickles; wastes away, wanes, perishes’, Proto-Iranian **xšar-* ‘to flow’ (> Persian *šārīdan* ‘to trickle’, Ossetic *äxsärdzän* ‘waterfall’). The reflex of the same Aryan root with the meaning ‘overflow, spill over’ has been borrowed into Northwestern Tibetan dialects (Balti, Purik, Ladakhi *chal-*).¹⁴

7. *da ra* (*rnam pa*) ‘piece, part, section’ (Martin 2010: 111). Cf. OIA *dala-* ‘fragment, piece, part; party, band’, *dr̥ṇāti, dalati* ‘bursts, cracks’,¹⁵ Av. *dərəṇti* ‘tears, splits’, Shina *dalo* ‘party, group’, Khowar *ḍal* ‘party, section, herd of horses’, Bengali *dal* ‘fragment’, Marathi *daḷ* ‘half’, Singhalese *dalaya* ‘division, part’, Persian *darrīdan* ‘to tear’, Kurdish *dīrīn* ‘to be torn’, Pashto *daṛa* ‘split, crack; splinter, piece of wood; slice’ (< **han-dr̥ta-*, see Morgenstierne 1927: 22; Rastorgueva, Edelman 2003: 342).

8. *dh(r)i lu* (*le lo*) ‘laziness’ (Martin 2010: 127, 128). Cf. West Pahari (Pangwali) *ḍhillā* ‘lazy’, Kumauni *ḍhīlo* ‘loose, slow, lazy’ < **ḍhīla-* ‘loose, slack’ (Turner 1966: 316). Martin, however, notes that Zhangzhung *dhri lu* and *dhi lu* can well be alternative spellings of *lhe lus* ‘lazy’ (Martin 2010: 243).

9. *dī ro, de ro* (*'gyur med*) ‘unchangeable’ (Martin 2010: 113-114, 120). The word is probably etymologically identical to Ladakhi *darak*, Purik *deraḡ* ‘stiff, hard’, all the three being borrowed from an Aryan source. Cf. OIA *dhārayati, dharati* ‘holds, keeps’, Av. *dāraiiehi* ‘(you) hold’, Kashmiri *dor* ‘firm, hard, strong, compact, durable, solid’, *darun* ‘to become steady (of something in motion), to become firm, to stand steady’, Pashai *d(h)ar-*, ‘to remain, stay’, Khowar *dorik* ‘hold back, wait, keep’. For more etymological parallels, see Kogan 2019.

¹² Cf. Kalasha *gaṇ* ‘hole which goes through an object’ (Trail, Cooper 1999: 103).

¹³ This word has an alternative spelling *li* (Martin 2010: 211, 214).

¹⁴ For details, see Kogan 2019. The semantic difference between the Zhangzhung and the Northwestern Tibetan words does not seem a major obstacle to our etymology. As the Old Indo-Aryan cognate clearly demonstrates, the verbal root in question could have been polysemic and combined the meanings ‘to flow’ and ‘to wane, decrease’.

¹⁵ On OIA *dr̥ṇāti* and *dalati* as etymologically identical verbs, see Mayrhofer 2001: 262.

10. *du ti* (*dmyal ba*) ‘hell’, (*sdug bsngal*) ‘suffering’ (Martin 2010: 115). Erik Haarh (1968) derives this word from OIA *duḥstha-* ‘bad off, faring ill, wretched’. However, semantically as well as phonologically closer seems to be OIA *duḥsthiti-* ‘bad plight’.

11. *dul pang* (*khu rlang*s) ‘mist and fog’ (Martin 2010: 120). This word has the variant form *du phang*,¹⁶ and its second component can hardly be separated from Zhangzhung *phang* ‘rainy mist, drizzle’ (Martin 2010: 145). The element *dul* may be compared with OIA *dhūli-* ‘dust, mist’, Torwali *dur* ‘mist’, Punjabi (Kangri) *dhūr* ‘thick mist or cloud’. It probably has nothing to do with Proto-Sino-Tibetan **tʃt* ‘dust’.¹⁷ The latter root seems to have reflected in Zhangzhung as *dur* in *slas dur* ‘earth dust’ (Martin 2010: 233).

12. *dus khri* (*gnod sems*) ‘thoughts of harming, malevolence’ (Martin 2010: 120). The element *khri* is nothing but the Zhangzhung word for ‘thought’.¹⁸ For *dus* (evidently, the equivalent of Tib. *gnod* ‘harm, damage’) cf. OIA *dūṣayati* ‘corrupts’, *dōṣayati* ‘spoils’, *dūṣya-* ‘corruptible, wicked’, *dōṣa-* ‘fault, sin’, Khowar *diš* ‘evil, bad, distasteful’, Shina *dišoiki* ‘to speak ill of, harm’, Kashmiri *dišērun* ‘to blame, malign’, Hindi *dūsnā* ‘to censure’.

13. *dzan* (*rims*) ‘contagion, epidemic’ (Martin 2010: 186). Cf. OIA *hanti* ‘kills, strikes’, *han-tar-* ‘killer’, *hana-* ‘killing, slaying’, Av. *ḵaiṇti* ‘kills, strikes’, *ḵaṇtar-* ‘killer’ < PII **ḵhan-* ‘to kill, strike’. For semantic development cf. OIA *māraka-*, *mārī-* ‘plague, epidemic (lit.: “killer, murderer”)’. The prototype of the Zhangzhung word in the donor language seems to have reflected an old lengthened-grade derivative of the cited root (PII **ḵhāna-*),¹⁹ the regular reflex of the formation with the short *a* (PII **ḵhana-*) would probably have been **dzun* (see 16).

14. *dzi ra* (*rig ’dzin*) ‘comprehension, clever’ (Martin 2010: 187). Cf. OIA *ḵīra-* ‘swift, lively, active’, Av. *ḵīra-* ‘swift, active, quick-witted’, Persian *zīrak* ‘clever, understanding’.

15. *gha ha* (*glang po*) ‘ox’ (Martin 2010: 65). The aspiration of the initial consonant in the Zhangzhung word is perhaps non-etymological. It may have appeared as a result of orthographical “Sanskritization”.²⁰ Cf. OIA *gō-* ‘ox, cow’, Pali *gāva-* ‘ox’, Av. *gau-* ‘bull, cow, cattle’, Gawar-Bati *gā*, Bashkarik *gā* ‘ox’, Indus Kohistani *gō-* ‘bull’, Shumashti, Torwali *gā*, Kalasha *gak*, Phalura *ghāu*, Kashmiri *gāw*, Lahnda *gā*, Punjabi *gāi*, Nepali, Assamese, Bengali *gāi*, Hindi *gāy*, Persian *gāv*, Pashto *ḡwā* ‘cow’.

16. *gu na* (*sgo nor*, *sgo phugs* [i.e., *sgo phyugs*]) ‘livestock, flocks’ (Martin 2010: 57). According to Martin, “this could be a Sanskritism, but it might need to be emended to correspond better with Skt. *gaṇa*, *kula* or the like” (Martin 2010: 57). The first of the two Sanskrit words just cited (i.e. OIA *gaṇa-* ‘troop flock’) seems to be the correct etymological parallel to the Zhangzhung lexeme. Cf. also Torwali *gan* ‘a herd’, Indus Kohistani *ghāṛ* ‘a crowd or group of people; group of trees or of houses’. The vowel *u* in the Zhangzhung form may have developed regularly from earlier **a*, cf. the change **a > o* before a nasal in the Aryan language of pre-Tibetan Ladakh (Kogan 2020).

17. *gu ra* ‘to turn (spontaneously, by itself)’ (Tenzin et al. 2008: 34). The Zhangzhung word may be either inherited and related to Tib. *’khor-ba* ‘to turn round’ or borrowed from some Aryan source (cf. OIA *ghūrṇati* ‘moves to and fro, rolls about, is agitated’, Hindi *ghūrṇā* ‘to roll round’, Indus Kohistani *ghuraṅ* ‘to turn, direct (water into a field)’).

¹⁶ The absence of final *l* in the first syllable may be due to a scribal error prompted by the existence of semantically similar expressions containing *du* ‘cloud’. Cf. *du pang spri zhi* ‘cloud and fog’, *du dang phang* id. (Martin 2010: 116).

¹⁷ Proto-Sino-Tibetan **tʃt* > Chinese **drən* ‘dust’, Tibetan *rdul* id., Lushai *dōl* ‘soil, refuse etc. banked up against a fence’ (Starostin 2005).

¹⁸ Cf. *khri* (*sems*) ‘mind, thought’ (Martin 2010: 51).

¹⁹ Cf. Buddhist Sogdian *’wz’n* (*ōžān*) ‘murder’ (Gharib 1995: 81) < Proto-Iranian **aua-jāna-* < PII **aua-jhāna-*.

²⁰ On this process see, e.g. Martin 2010: 12.

18. *ka ya bag* (*bkraḡ mdangs rgyas*) ‘wide and full brightness, lustre’ (Martin 2010: 33). The last component (*bag*) is attested as an independent word with the meaning ‘open’ (Martin 2010: 149). The meaning ‘brightness, lustre’ can thus be ascribed to the syllables *ka ya*. This Zhangzhung syllable sequence may reflect PII **kāśa-* ‘shine’. Cf. OIA *kāśate* ‘shines, is visible’, *kāśila-*, *kāśin-* ‘bright’, Av. *ākasaṭ* ‘(he) saw, beheld’, Khotanese *kas-* ‘to appear, to be seen’, Ashkun *kaširä*, Kati *kašera*, Waigali *kešerá*, Prasun *kašir* ‘white’ < Dardic (Turner 1966: 159). The intervocalic change **ś* > *y*, *i* is noted, e.g. in Pashai (Morgenstierne 1967: 33).

19. *kam* (*sog, sog pa*) ‘shoulder [blade], scapula’ (Martin 2010: 34). Cf. Gujarati *khām* ‘shoulder’, Sindhi *khambhurī* ‘wing’, Lahnda *khambh* ‘shoulder-blade, wing, feather’, Punjabi *khambh* ‘wing, feather’. R.L. Turner reconstructs for the Gujarati word the prototype **khavaka* ‘shoulder-blade’ which he compares with Latin *scapula* (Turner 1966: 203). It does not, however, seem reasonable to separate the Gujarati form from all the rest above-cited Indo-Aryan examples as well as from OIA *skambha-* ‘prop, pillar, fulcrum’. The semantic development was probably ‘fulcrum’ > ‘fulcrum of the arm’ > ‘shoulder(-blade)’ > ‘wing’.

20. *kun* (*rog rog*) ‘black, reddish brown’ (Martin 2010: 36). Cf. OIA *kr̥ṣṇa-*, Shina *kiñū*, Phalura *kišīnu*, Indus Kohistani *kišāṅ*, Bashkarik *kišin*, Torwali *kaṣan*, Kashmiri *kr̥hon*, *kruhun*, Tir. *kəḡən*, Kalasha *kriz̥ṇa* ‘black’. The etymological connection between the Zhangzhung, Old Indo-Aryan and Dardic forms was first postulated in Martin 2010. It should be added that Dardic seems to be the most probable source of borrowing, because in Indo-Aryan this word is no longer used as a color term.²¹

21. *ldem* (*shing*) ‘wood, tree’ (Martin 2010: 129-130). The word is hardly separable from Ladakhi (*l*)*ḍim*, Purik *ḍim* ‘trunk (of a tree), stem (of a plant)’, Balti *ḍim(s)* ‘trunk (of a tree)’ and shows etymological parallels in Burushaski, Dardic and Old Indian (cf. Burushaski *-ḍim* ‘Körper, Person, Selbst, (Nagir) Halm (des Getreides)’ (Berger 1998: 132), Shina *ḍim*, Khowar *ḍim* ‘body, tree-trunk’, OIA *ḍimba-* ‘body’). The Zhangzhung word could have been phonologically and/or orthographically influenced by Tib. *ldem-pa* ‘straight, upright, tall’.

22. *mu la* (*drod*) ‘warm’ (Martin 2010: 169). Probably, etymologically identical to Ladakhi *mala-mule* ‘lukewarm, tepid’, Balti *balbul* ‘tepid’. Cf. also Shina *būbūlū* ‘tepid’, Indus Kohistani *bubūl* ‘luke-warm (water)’, Burushaski *bulbulo* ‘lukewarm (of water)’. The word is not etymologizable in Dardic, and Burushaski seems to be the most likely ultimate source of borrowing.

23. *nel* (*ljang*) ‘green’ (Martin 2010: 136). Cf. OIA *nīla-* ‘dark blue, dark green, black; blue substance, indigo’, Ashkun *nīlestā* ‘green, blue’, Waigali *nyīlā*, *nīrā*, Kati *nīlā*, *ninyīlē*, Prasun *nīl*, *nyīlī* ‘blue’, Tirahi *nīlā*, Gawar-Bati *nīlā*, Kalasha *nīlā*, Bashkarik *nūl*, Torwali *nīlā*, Phalura *nīlō*, Sh. *nīlū*, K. *nyūl* ‘green, blue’, Sindhi *nīrō*, Lahnda, Punjabi, Hindi *nīlā*, Gujarati *nīlū*, Marathi *nīl*, Nepali *nīlō*, Bengali *nīlā* ‘blue’, Singhalese *nīl* ‘green, blue’. In Martin’s opinion, the Zhangzhung word must be a Sanskritism (Martin 2010: 136). This hypothesis, however, fails to explain phonological differences between the Sanskrit and Zhangzhung forms. Most probably, the true Sanskritism would have sounded *nī la*. It is also remarkable that in Old Indian *nīla-* is not at all the basic color term for ‘green’. In later Indo-Aryan this latter meaning is relatively rare, whereas it is very widespread in Dardic.

24. *ra ga* (*dmar po*) ‘red’ (Martin 2010: 203). Haarh and Martin compare this word with OIA *rāga-* ‘color, hue, dye, redness’ and consider it a Sanskritism in Zhangzhung (Haarh 1968; Martin 2010). It is, however, worth noting that Zhangzhung *ra ga*, as its Tibetan gloss clearly

²¹ The original Old Indo-Aryan (and Proto-Indo-European) meaning ‘black’ has been retained only in Pali and literary Prakrits, but borrowing from these languages into Zhangzhung is unlikely. In most New Indo-Aryan lects, where the reflexes of OIA *kr̥ṣṇa-* have been preserved, they function as local forms of the theonym Krishna. For details see Turner 1966: 179.

shows, is an adjective, whereas in Sanskrit *rāga-* was hardly used adjectivally. For this reason, it seems more probable that the source of borrowing was not Sanskrit but some spoken Aryan language where adjectivization had taken place. The possible source-form for the word in question could be either **rāga-* or **ranga-* (cf. OIA *raṅga-*, Persian *rang* ‘color’). The drop of old nasal before voiced gutturals is characteristic, e.g. of the Aryan language of pre-Tibetan Ladakh (Kogan 2019; 2020).

25. *rka* (*mkhal ma*) ‘kidney’ (Martin 2010: 40). It is unclear if this word can be a regular reflex of Proto-Tibeto-Burman **m-kal* as was hypothesized by Helmut Hoffmann (Hoffmann 1972). James Matisoff argues that the absence of final *-l* in the Zhangzhung form makes this etymology problematic (Matisoff 2001). Cf., however, OIA *vṛkka-* ‘kidneys’, Av. *vərəṭka-*, Persian *gurda*, Pashai *ṛakaṭi* (*r* < **wr*), (Areti dialect) *waṭṭawik* (*ṭṭ* < **ṛtk*), Kalasha, Khowar *bruk*, Bashkarik *juk* (*j* < **br*), Phalura *b(h)ruk*, Shina *zūk* (*z* < **br*), Lahnda *bukkī*, West Pahari (Jaunsari) *būkū* ‘kidney’, Bengali *buk* ‘heart, courage, chest’, Marathi *bokā* ‘kidney, breast’.

26. *shan* (*rnam shes*) ‘consciousness’ (Martin 2010: 219). Cf. also *sha shan* ‘memory, one-pointedness, vigilance, concentration’ (Tenzin et al. 2008: 260). The word seems to have Northwestern Tibetan cognates (Balti *shang* ‘wisdom, sense’, Purik *šan* ‘consciousness’, Ladakhi *shang* ‘alertness, awareness, caution, prudence’), which are themselves Aryan loanwords (cf. Shina *šon* ‘care, anxiety; awake, alert’, Khowar *šan* ‘fear, suspicion’, OIA *śaṅkatē* ‘is afraid, distrusts’, *śaṅkā-* ‘fear, distrust’ < PIE **k’enk-*). The same Aryan word was borrowed into Burushaski (*šan* ‘awake, aware; care, heed, attention’). In the Zhangzhung lexeme the final dental *n* (instead of expected guttural *ng*) may have developed due to phonological influence of some semantically close but probably etymologically unrelated words (cf., e.g. *shin* ‘to know’, *she shen* ‘awareness, conscience, intellect, knowledge, idea’). Note, however, the change *ng* > *n* in Zhangzhung *thon* ‘the number one thousand’ (Tenzin et al. 2008: 104) < Proto-Tibeto-Burman **s-tong* ‘thousand’.

27. *ta pi* (*’od*) ‘light’ (Martin 2010: 85). Cf. OIA *tapati* ‘is hot’, *tāpayati* ‘burns’, Pali *tapati* ‘shines’ Av. *tāpaiieiti* ‘makes hot’, Khotanese *pattav-/pattau-* ‘to burn up; illuminate’ (< **pati-tap-*), Sogdian *wyt’’p-* ‘to shine, radiate’ (< **ui-tap-*), Persian *tāftan* ‘to shine’, Kurdish *taw* ‘sun’.

28. *tan* (*ldan, nyid*) ‘having, oneself’ (Martin 2010: 87). Perhaps, we are dealing here with two homonyms. Zhangzhung *tan* ‘having’ may be somehow connected with its Tibetan equivalent *ldan* ‘having, being possessed of, provided with’, whereas the homophonous reflexive pronoun seems to be an Aryan loanword. Cf. OIA, Av. *tanū-* ‘body, person, self’, OIA *tānva-* ‘one’s own’, Pashai *tānuk*, Khowar *tan*, Phalura *tēni* ‘own, self’, Kalasha *tan* ‘self’ < **tānu-*.

29. *tsa mo* (*nya mo, nya*) ‘fish’ (Martin 2010: 179). This word has a monosyllabic variant *tsa*. Since in Zhangzhung, unlike in Tibetan, the element *mo*, apparently, did not exist as a productive derivative morpheme, the monosyllabic variant seems to be historically secondary.²² It may have come into being due to later misconstruction of *tsa mo* as containing the Tibetan feminine marker. Martin seems to be the first to reasonably point out the similarity of the lexeme in question to the word for fish in Shina dialects (Martin 2010: 22, 179). It should be noted in this connection that besides Shina possible cognates of the Zhangzhung word can be found also in Indus Kohistani and Burushaski (cf. Shina *čhīmū* (Gilgiti dialect), *čhumo* (Astori and Gurezi dialects), Indus Kohistani *chīm*, Burushaski *čhumo* ‘fish’). Given that the above-cited lexical items are not etymologizable in Indo-Iranian, Burushaski is likely to be their ultimate source.²³

²² James A. Matisoff (2001: 8) argued that Zhangzhung *tsa* may be either an Indo-Aryan loan or an inherited Sino-Tibetan word related to some Qiangic words for fish. He, however, seems to have discarded the latter alternative later on, since in his Sino-Tibetan database the Zhangzhung form is not cited as a cognate of Qiangic ones.

²³ It is remarkable that Burushaski *čhumo* is attested in the dialect of Yasin. This dialect is not in close contact with Shina, and no apparent Shina loanwords have been detected in it so far.

30. *tsag kor* (*'khor lo*) 'wheel, circle' (Martin 2010: 179). Cf. OIA *cakra-*, *cakrī-*, Av. *čaxra-* 'wheel', Shina *čarkū* 'wheel, spinning wheel', Kashmiri *cakḥir* 'wheel', Sindhi *caku*, Punjabi *cakk*, Gujarati *cāk*, Assamese *sāk*, 'potter's wheel', West Pahari *cakkī* 'mill', Kumauni *cāk(h)o* 'hand-mill', Hindi *cāk* 'millstone', Bengali *cākā* 'wheel', Sinhalese *saka* 'wheel, circle', Persian *čarχ*, Balochi *čark*, Ossetic *calχ* 'wheel'. Zhangzhung *tsag kor* 'wheel, circle' is construed by some scholars as a compound containing clipped forms of OIA *cakra-* and Tib. *'khor lo* (Stein 1971; Matisoff 2001). This, however, does not seem to be the case. If the second component were really borrowed from Tibetan, the voiceless aspirate *kh* would have been preserved as such. Cf. Zhangzhung *ti khor* 'the world, transmigration' (Martin 2010: 179) functioning as an equivalent of Tib. *'khor ba* and, most probably, formed from the borrowed Tibetan root *'khor* 'to go round', i.e. from the same root that we find in Tib. *'khor lo* 'wheel'. However, nothing like **tsag khor* is attested in Zhangzhung. Therefore, it appears more plausible that the Zhangzhung word, borrowed from some Aryan source, where it sounded as **čakkVr*, was later phonologically influenced by the Tibetan one, and due to this influence the vowel *o* has developed in its second syllable. The word in question coexisted in Zhangzhung with two etymologically identical lexemes evidently borrowed directly from Sanskrit, i.e. *rtsa krad* 'wheel, circle'²⁴ (Martin 2010: 182) and *tsa kra* 'knife, dagger' (Martin 2010: 178), cf. OIA *cakra-* 'a sharp circular missile weapon, the discus of Vishnu'.

31. *yos se* (*bud med*) 'woman' (Martin 2010: 203). The element *se* is a widespread Zhangzhung suffix.²⁵ As Martin reasonably notes, the word is, no doubt, connected with OIA *yoṣit-* 'young woman'. However, historical-phonological facts (e.g., the loss of the old second syllable) suggest that we are hardly dealing with a direct borrowing from Sanskrit. Rather, the lexeme in question was adopted from some colloquial Aryan language.

In Martin's dictionary I have come across two words belonging, in all probability, to the inherited Sino-Tibetan stratum but wrongly classified as Aryan loans:

1. *ag sho* (*kha*) 'mouth' (Martin 2010: 247). The Zhangzhung lexeme is hardly separable from words for mouth in closely related West Himalayish languages: Bunan *ag*, Pattani (Manchad) *à(h)*, Byangsi *ā*, Chaudangsi *ak*. For the element *sho* cf. *ting sho* 'river, moisture', *ting* being the basic Zhangzhung word for water.²⁶ The cited West Himalayish forms are traditionally considered to be etymologically connected with Tib. *'ag tshom* 'beard of the chin', Burmese *āk* 'opening, gap', Lepcha *ók* 'open (as door, mouth)' and to descend from Proto-Tibeto-Burman **ak* 'crack open' (Benedict 1972: 36; Matisoff et al. 2015). Martin, however, believes that the Zhangzhung word in question, as well as its West Himalayish cognates, may be of Aryan origin and compares it with Shina *āi*, *āzu*, Brokskat *ūzu* 'mouth' (< PII **ās-*, cf. OIA *ās-*, *ās(i)ya-*, Av. *āh-* 'mouth'). In terms of historical phonology Martin's hypothesis seems to be much less plausible than the traditional view.

2. *yu ti* (*chang*) 'wine, barley beer' (Martin 2010: 200). The second syllable is, no doubt, identical to the word for water in many Tibeto-Burman languages including Zhangzhung. The first syllable Martin compared with Shina *yō* 'barley'. The Shina word, as well as the related lexemes in other Aryan languages, reflects PII **iayā-* (> OIA, Av. *yava-* 'cereal'). However, Zhangzhung *yu* can hardly be separated from Kinnauri (Thebor), Tsangla, Digaro *yu*, Lushai *zu* 'liquor, wine, beer', Chinese **lu* 'wine bottle, wine' < Proto-Sino-Tibetan **jū* (Starostin 2005).²⁷

²⁴ Phonologically probably *tsakra* or *tsakre*. The spelling of this word was Tibetanized.

²⁵ For more details on this suffix see Haarh 1968.

²⁶ In Zhangzhung the word for water is attested in two spelling variants: *ting* and *ti* (Martin 2010: 88, 95–96).

²⁷ For more Tibeto-Burman comparanda see, e.g. Benedict 1972: 32; Matisoff et al. 2015.

Conclusions and discussion

The above list contains 31 etymologies of different reliability. At first glance, this figure may seem to be small, but it should be kept in mind that we are dealing with a poorly attested language. Considering the fact that the sizable proportion of about 3500 entries in Martin's dictionary are cross-references (Martin 2010: 5), it's easy to conclude that the above-cited Indo-Iranian loans make up not less than 1% of the Zhangzhung words identified thus far. It is also worth noting that a significant part, if not the majority, of lexical items in all the extant Zhangzhung dictionaries, including the Martin's one, are religious and philosophical terms, whereas in our list such terms are relatively rare.²⁸ It means that outside religio-philosophical technical lexicon the percentage of Indo-Iranian borrowings must have been much higher than the average for the analyzed dictionary. The above-listed words mostly belong to non-terminological everyday vocabulary and even include several items of the Swadesh list (*kun* 'black', *ldem* 'tree', *mu la* 'warm', *nel* 'green', *ra ga* 'red', *tsa mo* 'fish', *yos se* 'woman'). The mere presence of such lexical items, even though some of them do not seem to be main synonyms,²⁹ clearly indicates that Aryan-Zhangzhung contact must have been fairly intensive.

It would be of interest to try a tentative classification of the loans on the above list, based on their probable immediate sources. A noteworthy fact is the presence of several words possessing etymological parallels in Tibetan dialects of Ladakh and Baltistan (see 6, 9, 21, 22). These parallels, identified in our previous publications (Kogan 2019; 2020) as Indo-Iranian borrowings, show special similarity in phonological development, and in most cases also in semantics, to their Zhangzhung counterparts. This gives us reason to believe that the Indo-Iranian language of pre-Tibetan Ladakh, or some lect closely related to it,³⁰ was at least one of the sources of non-Sanskrit borrowed vocabulary in Zhangzhung.

It is also instructive that some of the above-listed lexical items, even though they lack cognates borrowed into Tibetan, display historical-phonological peculiarities pointing to their theoretical derivability from the same source-lect as Indo-Iranian loanwords in Ladakhi, Purik and Balti. E.g., as noted above (see 24), for Zhangzhung *ra ga* 'red', if we compare it with OIA *raṅga-* 'color', the drop of historical nasal before a voiced guttural may be hypothesized, which was attested in Aryan loan vocabulary of Northwestern Tibetan (Kogan 2019; 2020).

The change **a > u* in Zhangzhung *gu na* 'livestock, flocks' (cf. OIA *gaṇa-* 'troop, flock') may have passed through the stage of *o*, the development *o > u* before *n* being a typologically widespread phenomenon. If this conjecture is true, the Zhangzhung lexeme may well have been borrowed from the Indo-Iranian language of pre-Tibetan Ladakh, where *o* seems to be the regular reflex of earlier *a* before nasals (Kogan 2020).³¹

²⁸ Perhaps, the only two clear examples are *ag* 'rākṣasa' and *du ti* 'hell, suffering'.

²⁹ Zhangzhung *kun* 'black' is used for animals only (Martin 2010), the most common words for 'black' being *kh(w)a*, *mung* and their derivatives. For the two other above-mentioned color terms synonyms of non-Indo-Iranian origin are also attested, cf. *i, li ti* 'green', *pa ru, mang* 'red'. The latter word seems to be the most frequent and has apparent West Himalayish cognates: Bunan *maṅi*, Darmiya *mangnu*, Rongpo *məṅdə*, Byangsi *maṅde* 'red'. Zhangzhung *yos se* 'woman' has an inherited synonym *tsa med*, also possessing etymological parallels in modern West Himalayish languages (cf. Bunan *tsemet*, Kinnauri *chīme*, Byangsi *tsame* 'daughter, girl').

³⁰ Dialectal differences may account for certain phonological discrepancies between Northwestern Tibetan and Zhangzhung lexemes. Cf., e.g. Ladakhi *darak*, Purik *deraq* 'stiff, hard' vs Zhangzhung *di ro, de ro* 'unchangeable'. In this particular case, however, there seems to be an alternative plausible explanation. Absence of the final guttural stop in the Zhangzhung forms can be attributed to the fact that, unlike the Ladakhi and Purik words, they reflect not the old formation with the suffix *-k-*, but some other derivative of PII **dhar-* 'to hold'.

³¹ The change **a > o* did not take place if the nasal was followed by another consonant or the historical vowel **i* (Kogan 2020).

The intervocalic change $*ś > y$, attested in Zhangzhung *ka ya* ‘brightness, lustre’ (< PII $*kāśa-$ ‘shine’, see 18), may represent an intermediate phase of the historical-phonological process in the source-language, finally resulting in the complete drop of intervocalic $ś$, detected in Indo-Iranian loan vocabulary in Ladakhi.³² If this process really followed the scheme suggested here, i.e. $ś > y > 0$, borrowing into Zhangzhung must have taken place before the change had reached its final stage.

Naturally, all the three above scenarios of phonological development and loanword adoption are entirely hypothetical, and additional material is strongly needed for their verification.

Another historical-phonological feature, common for Indo-Iranian elements in Zhangzhung and in Northwestern Tibetan dialects, is the dentalization of earlier palatal affricates.³³ This feature, however, is also characteristic of Zhangzhung words borrowed from Sanskrit and, for this reason, can hardly be considered diagnostic.

As has been shown in Kogan 2020, the source-language of Indo-Iranian vocabulary in Ladakhi, in all likelihood, belonged to the Dardic group. In the light of this, it is noteworthy that at least for some Aryan loanwords in Zhangzhung Dardic origin seems probable. Such loanwords show historical-phonological isoglosses that affect Dardic languages too and even can, with a high degree of plausibility, be regarded “pan-Dardic”. For instance, in Zhangzhung *chal* ‘decrease, lower, cause diminution, fade away’ (see 6), whose prototype is likely to be related to OIA *kṣalati* ‘flows, trickles; wastes away, wanes, perishes’ and Proto-Iranian $*xšar-$ ‘to flow’, one can trace the phonological development PII $*kš > čh$ (= *ch* in Tibetological transcription). This process can hardly be reconstructed in detail, but what can be stated with certainty is that the voiceless palatal aspirate could not have evolved before the dentalization of historical palatal affricates. Otherwise, one should expect $*tshal$ with initial dental. A totally identical picture can be observed in the Indo-Iranian language of pre-Tibetan Ladakh for which I have suggested the affricate shift, similar to the one peculiar to Kashmiri and some other Dardic languages (Kogan 2019; 2020). Such a shift implies the development of palatal affricates from earlier cerebral ones. I see no reason to consider the same scenario impossible for the Aryan lect from which the Zhangzhung root in question had been borrowed. In other words, nothing prevents us from hypothesizing the change PII $*kš > čh$ in this lect. The latter change took place throughout the Dardic group, as well as in Nuristani. Nuristani influence on Zhangzhung, however, seems much less likely than the Dardic one. No examples of Nuristani diagnostic historical-phonological innovations were noted in the above list, while in some words the development is totally different from the Nuristani one (cf., e.g. *sh* (= $š$) < PIE $*k'$ in 26, the Nuristani development being $*k' > c$).

Another Dardic isogloss, found in certain Indo-Iranian loans in Zhangzhung is the deaspiration of the Proto-Indo-Iranian voiced aspirated affricate $*jh$ (< PIE $*gh, *g^wh$ in the palatalizing position). Cf. Zhangzhung *dzan* ‘contagion, epidemic’ < PII $*jhan-$ ‘to kill’.³⁴ This historical-phonological feature, however, is shared by Nuristani and Iranian too. The exact source of the cited loanword therefore remains, technically, an open question, although Iranian origin seems to be less probable for the same reason as the Nuristani one: no instance of typically Iranian phonological development was noted in our material, whereas in some cases historical phonology is definitely non-Iranian.³⁵

³² More on this process in Northwestern Tibetan varieties see Kogan 2020.

³³ See 13, 14, 30. On dentalization in the Indo-Iranian language of pre-Tibetan Ladakh see Kogan 2020.

³⁴ For more details, see 13.

³⁵ Cf. the above-mentioned change of PIE word-initial prevocalic $*k' > š$. The expected reflex of the Proto-Indo-European voiceless palatal dorsal in this position in East Iranian, including Saka dialects that theoretically

To sum up, there are some reasons to assume the presence of the Dardic etymological stratum in Zhangzhung vocabulary, though at the current state of our knowledge this hypothesis cannot be conclusively proven because of the lack of data. What does already seem remarkable and suggestive is the above-mentioned set of lexical isoglosses, shared by the donor language of Aryan loanwords in Zhangzhung and the supposedly Dardic source-lect of Indo-Iranian borrowings in Ladakhi.³⁶ The existence of such isoglosses may be indicative of the fact that the two here mentioned Sino-Tibetan languages had adopted a number of Indo-Iranian words from the same Dardic dialect or dialect group.

As for Indo-Aryan lexical influence on Zhangzhung, it appears to be much more questionable than the Dardic one, if we exclude evident loans from Sanskrit. No instances of typical Indic historical-phonological development have been noted in the above material. The list contains as little as two lexical items, common for Zhangzhung and Indian but absent from other branches of Indo-Iranian, i.e. *dh(r)i lu* ‘laziness’ and *kam* ‘shoulder blade, scapula’ (see 8, 19). The former of these examples, as already noted, may well be a ghost word, while the latter, being fairly reliable in terms of etymology, is remarkably lacking in Indo-Aryan languages spoken adjacent to Zhangzhung in the Himalayan region.

The preliminary picture of historical-phonological and lexical isoglosses, resulting from the above analysis, yields us some hints as to the erstwhile language and ethnic contact in the area under study. The speakers of the Zhangzhung language in pre-Tibetan times seem to have communicated rather intensively with their Dardic neighbors down the Indus. This fact is very much in line with some local historical traditions. The present-day Ladakh, especially its southeastern (“upper”) areas, is sometimes believed to have formed a part of the Zhangzhung kingdom before the Tibetan conquest (Bellezza 2008; Zeissler 2010).³⁷ Regions further to the northwest were regarded by some Bon authors as one of the main components of the Zhangzhung empire and called *Zhang-Zhung bar-pa*, i.e. Middle Zhangzhung. The language of *Zhang-Zhung bar-pa* was deemed to be related to Sanskrit (Haarh 1968).³⁸ In this tradition the Zhangzhung language proper was known as *Zhang-Zhung smar* (lit. “good (or refined) Zhangzhung language”) and its speaking area as *Zhang-Zhung sgo-pa* (lit. “the Zhangzhung of the gate, gateway Zhangzhung”). It is only natural to assume that the lects spoken in the two Zhangzhungs (“middle” and “gateway”) could have been in close contact.

It should, however, be borne in mind that all the above is still a hypothesis. The issue of possible Zhangzhung-Indo-Iranian language contact, no doubt, needs further research for which additional material is required. Potential sources of such material are, first of all, hitherto unstudied Zhangzhung-Tibetan bilinguals, still numerous in Bon religious literature. The Zhangzhung text of these bilinguals may well contain so far unknown Aryan loan vocabulary.

may have once been in contact with Zhangzhung, is s. Borrowing into Zhangzhung from Proto-Iranian, where according to some reconstructions the reflex was *ś, is improbable.

³⁶ In Kogan 2020 it was demonstrated that this lect had probably been influenced by Burushaski. It is notable in this connection, that for some Zhangzhung words (see, e.g. 22, 29) Burushaski seems to be the likely ultimate source. It cannot be ruled out that they were borrowed into Zhangzhung via Dardic.

³⁷ Interestingly, archaeological excavations have shown that in the western areas of the present-day Ngari prefecture, adjoining the Indo-Chinese frontier, there existed a number of important settlements of the Zhangzhung period, some of which probably functioned as religious centers. One of such settlements was discovered near the historic town of Ru-thog (Rudok), situated right near the border with Indian-controlled Ladakh (Bellezza 2008).

³⁸ It cannot be ruled out that this traditional view rested upon linguistic observations. The genetic relationship between Dardic languages and Sanskrit is in certain cases self-evident and, for this reason, could have been noticed by educated Tibetans.

It also cannot be excluded that the Indo-Iranian source-language of this vocabulary is represented in some previously unexplored textual fragments. In this regard, of special interest is the existence of several bilingual titles, whose non-Tibetan parts are written in lects, undoubtedly different from Zhangzhung and as yet unidentified.³⁹ Their decipherment may turn out to be of great help in the reconstruction of the language situation in the pre-Tibetan Upper Indus region.

Abbreviations for language names

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

References

- Benedict, Paul K. 1972. *Sino-Tibetan: a conspectus*. Cambridge University Press.
- Berger, Hermann. 1998. *Die Burushaski-Sprache von Hunza und Nager. Teil III. Wörterbuch Burushaski-Deutsch, Deutsch-Burushaski. Unter Mitarbeit von Nasiruddin Hunzai*. Wiesbaden: Harrassowitz Verlag.
- Bellezza, John Vincent. 2008. *Zhangzhung. Foundations of civilization in Tibet. A historical and ethnoarchaeological study of the monuments, rock art, texts, and oral tradition of the ancient Tibetan upland*. Wien: Verlag der österreichischen Akademie der Wissenschaften.
- Bon-po Association. 1965. *Tibetan Ḥaṅ-ḥuṅ dictionary*. Delhi: Lahore Press.
- van Driem, George. 2001. Zhangzhung and its next of kin in the Himalayas. In: Y. Nagano, R. LaPolla (eds.). *Bon Studies 3, New Research on Zhangzhung and Related Himalayan Languages*. Osaka: National Museum of Ethnology.
- Gharib, Badrozaman. 1995. *Sogdian dictionary (Sogdian-Persian-English)*. Tehran: Farhang Publications.
- Haarh, Erik. 1968. The Zhang-Zhung language: A grammar and dictionary of the unexplored language of the Tibetan Bonpos. *Acta Jutlandica* 40 (1): 6–43.
- Hoffmann, Helmut. 1972. Several Ḥaṅ-ḥuṅ etymologies. *Oriens Extremus* 19 (1-2): 193–201.
- Hummel, Siegbert. 1986. Der Ursprung der Sprache von Zhang-zhung. *Journal of the Tibet Society* 6: 3–16.
- Jacques, Guillaume. 2008. Zhang-zhung and Qiangic languages. In: Y. Nagano (ed.). *Linguistic Substrata in Tibet*: 121–151. Osaka: National Museum of Ethnology.
- Jäschke, Heinrich A. 1881. *A Tibetan-English dictionary with special reference to the prevailing dialects*. London: Routledge and Kegan Paul.
- Kogan, Anton. 2019. On possible Dardic and Burushaski influence on some Northwestern Tibetan dialects. *Journal of Language Relationship* 17/4: 263–284.
- Kogan, Anton. 2020. Notes on the historical phonology of Indo-Iranian loanwords in Northwestern Tibetan dialects. *Journal of Language Relationship* 18/3-4: 261–275.
- Martin, Dan. 2000. Comparing Treasuries: Mental states and other *Mdzod-phug* lists and passages with parallels in Abhidharma works by Vasubandhu and Asaṅga, or in Prajñāpāramitā sūtras: a progress report. In: S. G. Kar-may, Y. Nagano (eds.). *New Horizons in Bon Studies*. Osaka: National Museum of Ethnology.
- Martin, Dan. 2010. Zhangzhung dictionary. *Revue d'Etudes Tibétaines* 18: 5–253.
- Matisoff, James A. 2001. The interest of Zhangzhung for comparative Tibeto-Burman. In: Y. Nagano, R. LaPolla (eds.). *Bon Studies 3, New Research on Zhangzhung and Related Himalayan Languages*. Osaka: National Museum of Ethnology.
- Matisoff, James A. et al. 2015. *Sino-Tibetan etymological dictionary and thesaurus*. Available online: <https://stedt.berkeley.edu/> (accessed 15.07.2021).
- Mayrhofer, Manfred. 2001. *Etymologisches Wörterbuch des Altindoarischen. III. Band*. Heidelberg: Carl Winter Universitätsverlag.
- Mitruiev, B. L. 2016. “*Glubochajshaja sokrovishchnitsa*” kak leksikograficheskij istochnik dl’a izuchenija shangshungskogo jazyka. Sankt-Peterburg: Sankt-Peterburgskij Gosudarstvennyj Universitet. Available online: <https://dspace.spbu.ru/bitstream/11701/5266/1/st043060.pdf> (accessed 15.07.2021).

³⁹ The names of these lects are sometimes obscure and hard to interpret (cf., e.g. *ka-pi-ta gyung-drung lha'i skad* ‘the language of Ka-pi-ta swastika gods’).

- Morgenstierne, Georg. 1927. *An etymological vocabulary of Pashto*. Oslo: Kommisjon Hos Jacob Dybwad.
- Morgenstierne, Georg. 1967. *Indo-Iranian Frontier Languages. Vol. III. The Pashai Language. 1. Grammar*. Oslo: Universitetsforlaget.
- Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern / München: Francke Verlag.
- Rastorgueva, Vera S., Dzhoy I. Edelman. 2003. *Etimologičeskij slovar' iranskix jazykov. T. 2 (b—d)*. Moskva: Vostochnaja literatura RAN.
- Starostin, Sergei. 2005. *Sino-Tibetan etymology*. Available online: <https://starlingdb.org/cgi-bin/response.cgi?root=config&basename=\data\sintib\stibet> (accessed 15.07.2021).
- Stein, Rolf A. 1971. La langue *Ṣaṅ-ṣuṅ* du Bon organisé. *Bulletin de l'École Française d'Extrême Orient* 58: 231-254.
- Tenzin, Pasar Tsultrim, Changru Tritsuk Namdak Nyima, Gatsa Lodroe Rabsal. 2008. A Lexicon of Zhangzhung and Bonpo terms. In: Y. Nagano, S. G. Karmey (eds.). *Bon Studies 11, Senri Ethnological Reports* 76. Osaka: National Museum of Ethnology.
- Trail, Ronald L., Gregory R. Cooper. 1999. *Kalasha Dictionary with English and Urdu*. Islamabad: National Institute of Pakistan Studies, Quaidi-Azam University, Summer Institute of Linguistics.
- Turner, Ralph L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages*. Oxford University Press.
- Widmer, Manuel. 2014. *A descriptive grammar of Bunan. Dissertation submitted at the Faculty of Humanities at the University of Bern to obtain the degree Doctor of Philosophy*. University of Bern.
- Zeissler, Bettina. 2010. East of the moon and West of the sun? Approaches to a land with many names, North of Ancient India and South of Khotan. *The Tibet Journal* XXXIV (3/4) / XXXV (1/2): 371–463.

А. И. Коган. К реконструкции языковых контактов в регионе верховьев Инда в дотибетскую эпоху

В статье делается попытка частично восстановить картину языковых контактов в исторической области Шангшунг на северо-западе Тибета в эпоху, предшествовавшую завоеванию этой области тибетцами во 2-ой половине I тыс. н.э. Язык дотибетского населения Шангшунга, известный, главным образом, по ряду фрагментов в средневековых тибетских текстах, относится к западногималайской группе тибето-бирманской ветви сино-тибетской языковой семьи. При этом, однако, в лексике этого языка имеется значительный по объему индоиранский этимологический пласт. Автор показывает, что некоторые слова, относящиеся к данному пласту, обнаруживают точные соответствия в тибетских диалектах Ладакха и Балтистана. На основании анализа индоиранских элементов в шангшунгском языке делается вывод о том, что дотибетское население области Шангшунг, вероятно, находилось в достаточно тесном контакте с арийскими (возможно, дардскими) по языку жителями областей, расположенных ниже по течению Инда.

Ключевые слова: языковые контакты; лексические заимствования; индоиранские языки; дардские языки; сино-тибетские языки; шангшунгский язык.

The Vietic languages: a phylogenetic analysis

We present a new internal classification of the Vietic languages, covering all recognized sub-groups and languages for which we could source suitably comparable data. The analysis reconciles results from two distinct methodologies: (1) computational phylogenetics based on a 116-item basic word list, and (2) historical phonological changes among syllable codas. The analysis identifies five principal sub-groups: Thavung-Malieng (TM), Chut-Arem, Pong-Toum, Cuoi-Tho, and Viet-Muong (VM). While the identification of these subgroups is not original, aspects of the branching structure are. We find that the Vietic tree has a binary structure at the root, splitting between TM and the rest of the branch, which we may call the *Eastern Vietic* clade. Within that Eastern clade, we further find a north-south division, in which the northern VM group is most innovative while those to the south are more conservative. Previously, scholars have tended to group TM and Chut together as archaic lects based on broad typological features of word structure without suggesting specific common phonological innovations to link them. Our findings suggest that syllable restructuring and tonogenesis, which are hallmarks of Vietic languages, have developed largely independently among subgroups, notwithstanding very broad commonalities in phonological change which have contributed to strong areal convergence. Additionally, the Vietic homeland question is briefly discussed, with a northern locus of dispersal arguably supported by old loanword evidence and archaeological data.¹

Keywords: Vietic languages; Austroasiatic languages; linguistic classification; computational phylogenetics.

1. Background and Scope

Our understanding of the branching structure of the Vietic languages, and the position of Vietnamese in particular (as the focus of most scholarly attention), has developed at a frustratingly slow rate, with published studies reporting results that are inconsistent, incomplete, and lacking in transparency. This sits oddly alongside the rest of Austroasiatic², which has seen tremendous progress in classification across most branches in recent years (see the various discussions in Jenny & Sidwell eds. 2014). This progress follows decades of comparative historical work on Austroasiatic, and Vietic has not especially lacked attention in this regard (see references below). The slow development of Vietic classification occurred in part because – for the first half of the 20th century – there was real division among scholars as to whether Vietnamese was even an Austroasiatic language. In part, the uncertainty was a result of the tonal typology of Vietic that resembles Tai and Sinitic languages in its broad character; confusion regarding shared Chinese loanwords in both Tai and Vietic but which were regarded as evi-

¹ We would like to thank Michel Ferlus, Irina Samarina, and Nguyen Huu Hoanh for sharing of significant unpublished Vietic data. Without such data, we would not have been able to reach the extent of analysis and insight provided in this study.

² The term Austroasiatic is used here in preference to *Mon-Khmer*, as we do not regard the latter as a true taxon. It is generally recognized these days that Mon-Khmer is merely a label for Austroasiatic languages other than Munda, the latter tending towards more synthetic typology than most other Austroasiatic branches.

dence of common origins of the latter; and the poor knowledge of closely related Vietic languages that are more Austroasiatic in typology³. In this study, we have put those issues behind us and have pursued a fresh attack at the problem of Vietic internal classification achieved by combining computational phylogenetics and historical phonology. The result includes a family tree model of Vietic which we assert approximates an accurate phylogenetic structure of nested relations among Vietic sub-groups.

The idea that Vietnamese is ultimately related to Austroasiatic, primarily identified through identification of shared basic vocabulary, was first indicated in the mid-1850s (Logan 1854) under the rubric of the *Mon-Annam* language family, and enjoyed broad support until the early 20th century. Strikingly, Vietnamese was left out of discussion of Austroasiatic in Schmidt's (1906) influential study, and through the early decades of the 20th century, some scholars (e.g. Maspero 1912, Blagden 1913) treated Vietnamese as having origins more closely linked to Tai and Chinese. However, sympathy for the Austroasiatic position was maintained in various quarters (e.g. Pryluski 1924), although views differed over the relative priority of lexicon versus typology for several decades. In hindsight, it is understandable that the strong typological convergence of Viet-Muong, Tai, Hmong-Mien, and Sinitic languages was difficult to recognize for what it was. The theory of areality in linguistics was not well developed, and the idea of languages sharing so much that they converge structurally seemed to be just too much to accept in some quarters. Consequently, the robust isolating morphology, monosyllabicity, and complex tone systems, combined with numerous Chinese loanwords in both Tai and Viet-Muong made it seem more likely that Vietnamese shared a common origin with those languages than the Austroasiatic family with its derivational morphology, largely non-tonal character, and robustly sesquisyllabic word structure.

Nevertheless, in the second half of the 20th century, scholars converged on general recognition of Vietnamese and other Vietic tongues as Austroasiatic languages that have been affected structurally by contact with Tai and Chinese. Scholars were persuaded by multiple converging lines of evidence:

- the persistent identification of Austroasiatic core vocabulary in Vietnamese and the consistent results of lexicostatistical studies (e.g. Thomas and Headley 1970, Huffman 1978, Alves 2017);
- the hypothesis of tonogenesis proposed by Haudricourt (1953, 1954a) and numerous subsequent publications supporting it;
- a clearer understanding of the shared early Sinitic loanwords in Tai and Vietnamese (e.g. Haudricourt 1954b), including many different loanwords, thereby demonstrating such words do not demonstrate Tai-Vietic language contact;
- the recognition that Vietnamese must not be considered in isolation, but rather as one of a number of Vietic languages forming a distinct language branch (cf. various studies cited in this article).

It is in view of the typologically transitional Vietic languages spoken by small, isolated groups that the Austroasiatic affiliation of Vietnamese becomes particularly evident (cf. Alves 2003 for discussion). Beginning with initial observations at the beginning of the 20th century (mainly wordlists gathered by French linguists such as Cadière 1905, Chéon 1907, Guignard 1911, and others), scholars began to become aware of very small highland communities in north-central Vietnam and the Vietnam-Laos border lands, speaking languages that have strong lexical affinities with Vietnamese and Muong, but with strikingly different phonologi-

³ See also Gage (1985) and Alves (2006) for historical reviews of the place of Vietnamese in Austroasiatic and Sidwell (2009) and Sidwell (2014) for a summary of Vietic phylogenetic studies.

cal and morphological typological features. Some of these languages are non-tonal, while others have simple tone systems. Some have a complex range of onset clusters, while Muong lects have fewer clusters, and Vietnamese has only a single type with medial *-w-*. All preserve pre-syllables, which are completely absent from Vietnamese and Muong. In addition, those languages have evidence of Austroasiatic prefixes and infixes, which are mostly fossilized and with little to no productivity. These languages tended to be classified together with the Muong lects as minor languages standing alongside Vietnamese, essentially regarded as uncultivated rural dialects unworthy of the attention properly given to the national tongue. One notable exception is Maspero's (1912) major work on Vietnamese language history, a monograph containing dozens of tables of comparative Vietic, Austroasiatic, Tai, and Chinese lexical data and historical phonological observations. The Vietic data he presented was rich, with Vietnamese, multiple Muong lects, Nguon, Cuoi-Tho lects, and the Sach lect of the Chut group. However, he referred to all of these as «dialects», grouped only by three geographic areas. His primary focus was on Vietnamese history, Sino-Vietnamese, and establishing a connection between Vietnamese and Tai, not addressing the issues of the relationships among the «dialects». Nevertheless, the data he amassed and his insights were influential in later works, including those of Ferlus described below.

In the second half of the 20th century, Vietnamese scholars (e.g. Mạc 1964, Phạm 1975, 1979, etc.) noted that these minor upland languages have some typological traits that are more similar to other Austroasiatic languages than Vietnamese or Muong. Consequently, there emerged a better recognition of the extent and complexity of *Viet-Muong* languages as they were primarily known and referred to in linguistic studies and reference works (e.g. Barker 1963, Barker and Barker 1970, Thompson 1976, Ferlus 1974, and others), or *Vietnamuong* in a few publications.

With the 1973 International Conference on Austroasiatic Linguistics (or ICAAL, hosted by the University of Hawaii), new focus and effort fell upon issues of language phylogeny. In particular, Ferlus provided one of the first substantive lists of Vietic language subgroups (Ferlus 1974: 70–71) and the first phylogenetic proposal for Vietic (Ferlus 1979: 81), under the *Viet-Muong* rubric. A little later Hayes (1982) coined the name *Vietic* for the branch, reserving *Viet-Muong* for just the northern sub-grouping of Vietnamese, Muong and Nguon. The new name was arguably needed as with increasing documentation and understanding of the smaller upland languages of the group, the *Viet-Muong* label had become increasingly problematic. Hayes (1992) also offered his own proposal for the internal structure of Vietic based on lexicostatistics and the historical phonology of coda **-h*, although his proposals did not find wider support. Other scholars have also offered classificatory schemes (e.g. Chamberlain 2003, Trần 2011, Sidwell 2014), although these studies, and the classification by Ferlus (1979, discussed in more detail below) apply different methodologies, making them difficult to compare or assess.

In this context, we (Sidwell and Alves) decided to take a fresh look at the problem from first principles. The primary challenge previously was the limited amount of directly comparable data which could be brought to bear on the problem. The past century of research on Vietic followed a familiar pattern in language documentation: there were few overriding principles guiding investigators, who opportunistically and/or idiosyncratically collect lexical and textual materials. Each had his own research priorities which overrode elicitation of standard lists or concern for common glossing of basic vocabulary items. The problem arose in part because researchers came from different national backgrounds and worked in different historical periods. For example, word lists in early French language publications have somewhat limited value, such as the Harème lexicon gathered by Rivière (Malglaive 1902: 285–290) which com-

piles words he collected in the field before the *École française d'Extrême-Orient* devised a standard list. Additionally, Malglaive and others used French-based impressionistic transcription which has to be interpreted with knowledge of both French orthography and the linguistic typology of Southeast Asian languages. More recently, Chamberlain (1998, 2018) has compiled zoological terminology in Vietic languages; while this is especially useful for specific etymological issues, it is difficult to find matching lexical items in sources for related languages compiled by other researchers. These and other issues lead to practical difficulties in the aggregation and semantic alignment of wordlists for comparative purposes, although we are fortunate in that these days, investigators do more or less consistently transcribe their data in broad IPA such that phonetic values can be assumed with reasonable confidence.

Given the real difficulties of aggregating and dealing with the available data, it is understandable that efforts to date have been limited, with scholars generally attempting aggregations based on their own data sets. An example of the latter is Babaev and Samarina (2018) who aggregate 100-word lists for Ruc, Sach, Maleng, Arem, and Kri based on the data compiled over decades of Soviet/Russian-Vietnamese joint expeditions to inform a lexicostatistical matrix. The impressive data aggregation of Ferlus (2007) presents another aspect of the problem: being compiled specifically to identify historical cognates, that collection ignores many early loanwords and low-level lexical innovations that are crucial for statistical analysis, artificially producing gaps in the data that we labor to restore by reference to Ferlus' original source materials. To deal with these issues and make the study manageable, we have attempted to exhaustively aggregate lexicons from across the full diversity of Vietic languages while keeping to a modest list of basic vocabulary items. We assume that this approach maximizes the likelihood of finding items in any list to fill the limited semantic categories, minimizing overall data gaps.

In addition to the aggregation of published and unpublished lexicons, we have also taken advantage of existing comparative studies for recognizing cognates and regularity in correspondences. In this respect, we rely principally on the analytical framework of Ferlus' phonological and lexical reconstruction of proto-Vietic which emerged in a series of works (1991, 1997, 1998, 2007, etc.). Additionally, the wider Austroasiatic context is rendered largely accessible by the publication of Shorto's (2006) *Mon-Khmer Comparative Dictionary*, and the online resources and search tools of the *Mon-Khmer Etymological Dictionary*⁴. Given the access we now have to relevant tools and resources, we took the view that it is now worth pursuing a comprehensive phylogenetic assessment of Vietic by applying parallel computational phylogenetic methods, with related matters of etymology (i.e., retentions, innovations, and loanwords), and analyses of shared historical sound-changes in Vietic, integrating these into a coherent model of phylogeny.

2. Previous studies

Through the first half of the 20th century, researchers were aware of the relationship of Vietnamese with Muong and languages such as Nguon and Sach (e.g. Chéon 1907, Maspero 1912, etc.). In a largely ethnographic book, Vương (1963) assembled core prior studies and raised major ethnohistorical points about Austroasiatic broadly. As part of his supporting data, he presented a substantial list of comparative lexical data (about 100 items in the Appendix to his book) for several major lect groups of Vietic, including Vietnamese, Muong, varieties of Tho, Arem, May, Ruc, and Sach. His discussion did further connect Vietnamese with Austroasiatic,

⁴ <http://sealang.net/monkhmer>

but he did not address the degree of relatedness among these languages. Indeed, sufficient data and analytical tools to properly assess the degree of relatedness among these languages were lacking. Into the 1960s, scholars had no effective sense of nested branching among Vietic languages, and Vietnamese and Muong were conventionally treated by linguists as fair representatives of the branch for comparative and typological purposes. An influential example of such is Thomas and Headley's (1970) lexicostatistical study of Mon-Khmer languages. That study included Vietnamese and Hoabinh Muong lexicons as representatives of Vietic and mentioned as an after-thought that "Arem, Mày (Ruc), and Taypong [...] clearly belong in the Viet-Muong branch" (Ibid. 404). This was consistent with the received view that the branch essentially split between Vietnamese on one side and Muong plus various small languages on the other.

It was not until the 1970s that in a series of articles by Ferlus (1974, 1975, 1979), Vietnamese and Muong were recognised as a sub-group distinct from the other Vietic languages, with the latter forming one or more other clades within Vietic. Subsequently, the first published proposal for a full phylogenetic tree of Vietic with nested sub-branching appeared at the end of that decade (Ferlus 1979, Figure 1).

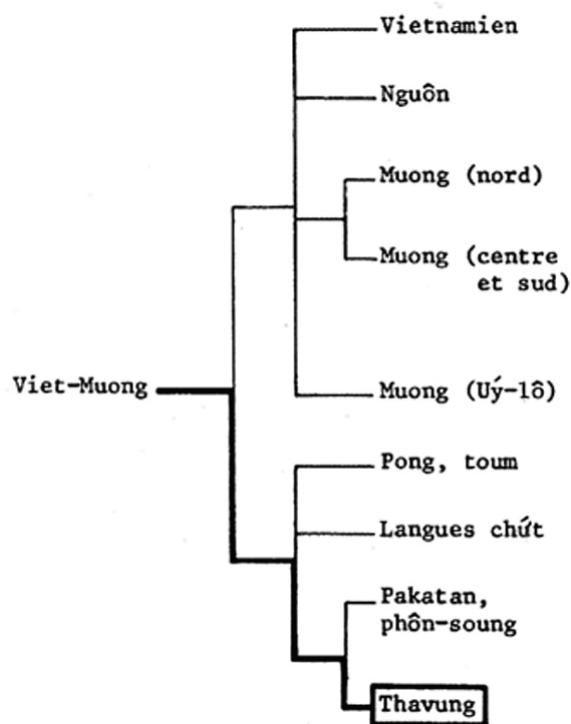


Figure 1. Vietic classification by Ferlus (1979: 81)

Ferlus' (1979) phylogeny split Vietic (*Viet-Muong* in his terminology) into two sub-branches: a northern group of Vietnamese, Nguon, and Muong lects, versus a southern group encompassing all other Vietic languages. The main significance of Ferlus' tree is to make clear that the subgrouping of Vietnamese-Nguon-Muong (or *Viet-Muong* in present day terms) was solidly recognized. This followed nearly two decades of comparative reconstruction focused solely on Viet-Muong by various scholars, in which the relationship with other Vietic lects was left unclear. This separation of Viet-Muong from all other groups lingered in later phylogenetic claims. It is striking that the southern group in this scheme was divided into three equidistant clades (i.e. Pong-Toum, Chut, Pakatan-Thavung), hinting that this was effectively an assumed classification rather than a well justified phylogeny.

2.1. Comparative Reconstruction

Comparative reconstruction is closely related to classification. The method models ancestral states of language at various levels of representation (e.g., segments, morphemes, etc.) and may be dependent on the modelling of tree structure while also informing the tree structure. While this is potentially circular, in practice, one is expected to test various models of branching against hypotheses of shared innovations which are assumed to indicate nodes in the tree. The tree structure that best accommodates the reconstructed changes between the proto-language and daughter languages is then proposed as approximating the real phylogeny of the family.

Serious comparative Austroasiatic studies began with the work of Schmidt in the early 20th century (1903, 1904, 1905), laying serious groundwork for understanding the historical de-

velopment of Austroasiatic phonology, morphology and lexicon. However, Vietnamese or other Vietic data did not figure in that work until Haudricourt (1953, 1954a) showed the regular correspondences between Vietnamese tones and Khmu syllable structure. This permitted the formulation of sound laws that map the development of Vietnamese morphemes from originally Austroasiatic roots. By the 1970s, the historical lexicon and phonology was known well enough by concerned scholars that many Vietnamese roots were uncontroversially linked to proto-Austroasiatic forms (e.g. *nước* ‘water’ < **da:k*, *chó* ‘dog’ < **cɔʔ*, etc.), although this was not reflected well in formal publications of the time.⁵

From the early 20th century, the *École française d’Extrême-Orient* in Hanoi conducted language survey work, with scholars collecting lexicons on the bases of standard elicitation lists, but these lists were limited in extent, and data was not transcribed in structuralist-linguistic terms. The situation began to change from the late 1950s onward as linguists trained in phonemics and structuralist theory began collecting field data on Muong and other minority languages of Vietnam, yielding lexicons readily useful for comparative analyses. This included foreign scholars in South Vietnam who worked with Muong speakers who had relocated from the north after partition, in addition to a modest field linguistic tradition within the Socialist Republic of Vietnam. Consequently, comparative studies of Viet-Muong became viable, and results began to appear from the mid-1960s (Barker 1966, Barker & Barker 1970). Significantly, many of the smaller Vietic speaker groups, such as the Chut, live in highland and border areas which were mostly inaccessible to scholars due to geographical and war-time conditions, a situation that favored the focus on Viet-Muong.

By the mid-1970s, scholars were incorporating data from other Vietic groups, variously using both older and newer sources, working towards a comprehensive Vietic reconstruction. The most important in this tradition is the work of Ferlus extending from the 1970s to the present day. Ferlus (1975) effectively established the paradigm for comparative Vietic in a 33-page paper that reconstructs Vietic phonology using comparative, philological, and loanword evidence. That work is weighted heavily towards Vietnamese and Muong, and it incorporates limited data from Vietic minority languages (mainly Thavung, Sach, and Nguon). The 1970s also saw the beginning of a long-term cooperation between Soviet and Vietnamese scholars, and in particular, Sokolovskaya took an interest in Vietic reconstruction. She marshalled 27 published sources to offer a reconstruction of 600-plus proto-Vietic roots (Sokolovskaya 1978) and later prepared a more substantial comparative dataset, although this was lost after her untimely passing in the 1980s.⁶

Other works of comparative-historical reconstruction related directly to Vietic have appeared, and more than a dozen are referenced and characterized in Table 1. One can see from the table that these are quite diverse, with works focusing on different aspects of phonology or level of structure, with a slow progression towards a coherent and comprehensive view of Vietic history only emerging from decades of investigations. The scholar to whom we are most indebted is Michel Ferlus, who has done most to aggregate, integrate and analytically assess the scholarly history and the ongoing emerging breadth of data to reconstruct proto-Vietic.

By the early 2000s, Ferlus had released an extensive and well justified proto-Vietic reconstruction which establishes phonological correspondences that allow us to account for all ap-

⁵ From the 1973 ICAAL meeting onward scholars were circulating manuscripts and sharing conference presentations which featured well developed models of Austroasiatic phonology and many reconstructed forms. However, much of this work was not published, or only published decades later. Hints of this can be seen in publications such as Shorto (1976), which foreshadowed his posthumous (2006) *Mon-Khmer Comparative Dictionary*.

⁶ A substantial manuscript was passed to the Mon-Khmer Studies editors but did not proceed further. The authors would appreciate any leads that would assist in tracking down a copy of this missing manuscript.

parent sub-groups within the branch. However, this reconstruction is not dependent on a model of nested branching within Vietic; morphemes in each language or sub-group are directly derived from proto-Vietic forms. Thus, the reconstruction is not reliant on sequencing shared innovations among sub-groups. Effectively, each subgroup is treated as direct descendants of proto-Vietic, with their various patterns of lexicon and phonology reflecting coincidental parallels. While this may imply a flat rake-like tree structure, actually this is not the case, and nested relations were simply not treated as analytically relevant above the sub-group level. We have no problem as such with Ferlus' approach as a means of coming to a phonological reconstruction, but it was clearly not his priority that the reconstruction yield a detailed family tree. That said, Ferlus' organization of lexicon and phonological correspondences provides a solid basis for informing our methods and results.

Authors	Years	Level	Notes
Barker	1966	VM	Tonal correspondences demonstrated with some 600 VM lexical comparisons
Barker & Barker	1970	VM	Coda and vowel reconstructions supported by 210 VM lexical comparisons
Ferlus	1975, 1982	Vietic	Reconstructs proto-consonants and tones, and their development to modern Vietnamese values
Thompson	1976	VM	Reconstructs approximately 700 Proto-VM lexemes based on Vietnamese and three varieties of Muong
Sokolovskaya	1978	Vietic	Reconstruction of over 600 Proto-Vietic lexemes supported by data from 27 sources. Proto-Vietic treated as non-tonal, non-glottalized
Ferlus	1991, 1997	Vietic	Reconstructs Proto-Vietic vowels
Ferlus	1992	Vietic	Reconstructs Proto-Vietic onset consonants, connecting modern Vietnamese to Proto-Vietic and Chinese loanwords
Nguyen T. C.	1995	Vietic	Reconstructs Proto-Vietic inventories of onsets, vowels, and codas (but non-tonal), tracing modern Vietnamese segments to Proto-Vietic and two stages of Chinese loanwords
Ferlus	1998, 1999, 2004	Vietic	Reconstructs Vietic tonal history
Nguyen V. T.	2005	VM	Reconstructs VM onsets, vowels, and tones, in addition to one hundred lexemes based on Hanoi Vietnamese, Nghe Tinh (Northern) Vietnamese, and 30 Muong dialects
Shorto	2006	AA	Incorporated Viet. and Muong in his reconstruction (including tone categories), tabling consonant correspondences.
Ferlus	2007	Vietic	Reconstructs more than 1,200 Proto-Vietic lexemes
Ferlus	2014	Vietic	Provides reconstructions of onsets, vowels, and codas in Proto-Vietic (which he calls Proto-Viet-Muong), notes on tonogenesis, and proto-language vowels for his proposed Proto-Pong-Cuoi

Table 1. Comparative reconstructions related to Vietic

A specific example of this problem can be seen in Hayes' (1992: 216) attempt to invoke the loss and/or retention of coda -h for subgrouping among Thavung, Ruc, Muong and Vietnamese. Relying solely on the presence or absence such a segment is quite problematic; in the context of the areal trend to tonicity, it is particularly important to recognise that one cannot just treat coda [-h] as straightforward segment to be retained or lost as is, but whether it remains

present rephonologized as a tonal feature (e.g. phonation features of a tone), and this may or may not be readily recognized in data transcription. It is our considered view that the problem of justifying clustering/splitting arguments in this genetic and areal context is best addressed by identifying multiple types of evidence of linguistic innovations and retentions that align consistently to support a model of phylogeny. In the history of Vietic classifications, authors have relied upon diverse methods and forms of evidence, but none appear to have been able to assemble a sufficient range of types of data to achieve a coherent synthesis.

2.2. Classification

As we have noted above, comparative studies of Vietic are now quite advanced, yet the results have not been successfully applied to the classification problem. This is not to say that classification has been ignored, on the contrary, but our reading of the situation is that scholars have rather relied on a range of rather different, and to some extent incompatible, methodologies when offering their classifications. These include phonological/typological considerations, geographic distribution, and/or lexical data (i.e. lexical isoglosses or lexicostatistics). The approaches are sometimes explicitly stated, while in some cases, the methods must be inferred from the method of presentation. Table 2 shows our characterization of how scholars have prioritized different aspects of language to inform their classifications of Vietic and their resulting hypotheses.

Author	Year	Approach	Key Insights
Ferlus	1979: 81	Typological	Two main branches 1. Viet-Muong (neutral regarding Nguon) 2. Pong & Chut & Thavung-Malieng
Ferlus	1989–90: 52–53	Typological and geographical	Several main groups listed without tree structure 1. Vietnamese 2. Muong 3. Pong-Cuoi 4. Western archaic (Thavung, Pakatan) 5. Eastern archaic (Chut, Arem, Malieng)
Diffloth	1989 (via Chazée 1999)	Typological	Two main branches 1. Viet-Muong & Pong-Cuoi 2. Archaic languages
Hayes	1992: 220–221	Geographical organization, lexicostatistics, and sound change (final *-h)	Three main branches 1. West Vietic (Thavung lects, Pakatan, others) 2. Central Vietic (Pong-Cuoi) 3. East Vietic: Viet-Muong (Nguon with Muong) & Chut
Nguyễn, T. C.	1995: 12	Typological	Two main branches 1. Viet-Muong 2. Pong-Cuoi & Chut & Thavung-Malieng (adopting Ferlus's 1979 model)
Chamberlain	1998: 106	Ethnozoological lexical data and inferrable geographical organization	Two main branches 1. Viet-Muong & Pong-Cuoi & Chut 2. Thavung-Malieng

Table 2. Previous classification of Vietic languages

Table 2. Previous classification of Vietic languages (continued)

Author	Year	Approach	Key Insights
Nguyễn H. H.	1999	Lexicostatistical	Several inferable lect groups 1. Viet-Nguon-Muong 2. Phong-Poong-Tum-Ly Ha 3. Chut 4. Arem 5. Malieng-Kri-Pakatan 6. Thavung-Phon Sung
Chamberlain	2003	Ethnozoological lexical data and inferable geographical organization	Several branches (citing Diffloth 2001) 1. Viet-Muong 2. Pong-Cuoi 3. Thavung 4. Chut-Malieng 5. Malang 6. Kri
Peiros	2004	Lexicostatistical	Several main branches (branching downward from Thavung) 1. Thavung 2. Malieng 3. Ruc-Arem 4. Pong-Cuoi 5. Viet-Muong
Sidwell	2015: 205	Synthesis of multiple studies	Three main branches 1. Viet-Muong 2. Pong-Cuoi 3. Chut (archaic languages divided into groups to the East and West)
Chamberlain	2018	Lexical (fauna terms)	Two main branches 1. Viet-Muong & Pong-Chut 2. Thavung-Malieng lects & Kri-Mlengbrou
Trần	2018: 61	Typological	Two main branches 1. (monosyllabic) Viet-Muong & Cuoi 2. (bisyllabic) All others (Pong, Arem, Chut, Malieng, Aheu)

It is clear from all the literature that the Viet-Muong sub-branch within Vietic is uncontroversial. It consists of Vietnamese, multiple varieties of what are collectively called Muong (30 Muong lects have been documented in Nguyễn V. T. 2005), and Nguon, of which there are at least two documented varieties. Significant phonological diversity is found among Muong lects, although lexically they are rather homogenous, so it is not quite clear to what extent they reflect distinct languages or are better regarded as a close dialect chain or linkage. Recently, Phan (2012) presents phonological evidence and arguments that Muong is a paraphyletic taxon: based on a lack of clear phonological patterning, the Muong lects do not sit unified in opposition to Vietnamese within the tree, but all of them and Vietnamese (and potentially Nguon) descend from proto-VM more or less equally. Consequently, we take as a given that Vietnamese, Nguon, and the Muong lects form a unified *Viet-Muong* node somewhere within Vietic, and do not investigate the specific question of Viet-Muong internal branching further

as it is assumed that this is not relevant to the higher tree structure. We have utilized Vietnamese, Muong, and Nguon data in our analysis, and our results also confirm this grouping. Finer clarification of the historical linguistic relationships within the Viet-Muong sub-branch will require an assembling and analysis of phonological and lexical data of the few dozen Muong doculects and thus is beyond the capacity of the current study.

The approaches that have been used to group/divide the Vietic languages can be summarized as follows, with our brief assessments of their consequences:

- Phonological: Hayes (1992: 216), based on Ferlus (1974, 1975), posits limited sub-branching based on the retention or loss of proto-Vietic *-h. Noting various shared phonological retentions, Nguyễn T. C. (1995: 257) acknowledged that shared phonological innovations to determine sub-branching among the smaller Vietic lect groups were insufficient to provide finer phylogenetic distinctions.
- Typological: Restructured monosyllabic Viet-Muong languages are most often considered to be sub-grouped in contrast with the more conservative sesquisyllabic Chut, Arem, Thavung, Malieng/Pakatan, Kri, and others. Also, both the Viet-Muong and Cuoi groups have complex systems of tonal phonemes, while others have smaller tonal inventories or none. However, this aspect is challenging to implement for phylogenetic purposes as the timing of the development of tonogenesis may vary among the sub-branches.
- Lexical: Lexical isoglosses can be referred to in considering sub-grouping membership. Chamberlain (2003, 2018) refers solely to fauna terms, utilizing a type of isogloss approach, to suggest detailed phylogenetic trees of Vietic. However, his methodology and assumptions in use of such data to make these determinations are not clear, making them difficult to assess.
- Geographic: Hayes (1992) relies at least partly on geographic grouping as suggested by regional labels (i.e. east, west, and central). Ferlus in his various works refers to a broad north-south division. While geography does sometimes correspond to phonological, lexical, and/or typological factors, we assume linguistic factors must be considered first, with geography as a secondary indicator of language history (e.g. population movements and language contact).

The twelve publications in Table 2 each show different versions of a Vietic Phylogenetic structure, with overlap in various cases, but otherwise distinct claims in each proposal. A recurring hypothesized division is between Viet-Muong versus all else (e.g. Ferlus 1979, Nguyễn T. C. 1995). This classificatory division appears to be influenced by the typological distinction of the strictly monosyllabic Viet-Muong languages versus the sesquisyllabic structure of other Vietic languages. However, the degree of variation of the dozen approaches makes it clear that no consistent position has dominated. Key to the discussion is the position of Pong-Toum and Cuoi, which have variously been grouped with VM, with sesquisyllabic lects, or as a distinct Pong-Cuoi clade.

Contradictory claims are evident: some proposals put the sesquisyllabic Chut sub-group into a closer relationship to Viet-Muong than other archaic Vietic lects, while other proposals suggest grouping Chut more directly with other archaic varieties. A north-south division seems to emerge at the extremes of Viet-Muong in contrast with the archaic Vietic languages, but there is otherwise little consistency among the posited nested branchings. All things being equal, we might expect that classifications based on different data types and methodologies would converge on common points that underlie the real shape of the family tree. However, beyond identifying the Viet-Muong node, studies have offered strongly diverse results, indicating fundamental problems with the nature of the evidence used and/or the assump-

tions underlying its analysis. The variation in methodologies also means that proposals lack comparability, which we attempt to overcome in this study by the methods outlined in Section 3.

3. Methods and data

Given the above, we decided to proceed with a fresh analysis on first principles, rather than attempt to revise or repair any existing analysis. For this, we compiled a 116-item basic word list (described below), extracting the data from the most complete aggregation of sources available to us in some kind of readily processed form. The initial method is the same as applied in classical lexicostatistical or glottochronological studies, with all items scored for their cognate values, and apparent and likely loans (primarily from Chinese and Tai) identified and scored as zero items, according to our assessment as experienced comparative Austroasiatic specialists with understanding of phonology and etymology in the region. SplitsTree (4.14.2 for Windows) was used to generate neighbor nets and phylograms, and these were further assessed by comparing the computed pattern of branching nodes with shared phonological changes implied by comparative reconstruction. The data we considered was not limited to the 116-word list already mentioned; a more extensive lexical database was built as follows: the data from Ferlus (2007, available online at sealang.net/monkhmer) was laid out in a spreadsheet, and then augmented by integrating additional wordlists, aligning items with Ferlus' etymologies. The advantage of using Ferlus' analysis as an organising principle is that we could rapidly identify reflexes of specific etyma, without having to accept all of Ferlus' reconstructions. Columns with our spreadsheet were created for sorting etymologies by onsets, nuclei, and codas, so we could rapidly identify multiple members and reflexes of specific proto-segments and thus discuss shared changes/innovations. The result is a dynamic working file of Vietic lexical data and phonological analyses which not only informed this project but is also the basis for a larger effort to revise and extend Vietic comparative reconstruction.

The SplitsTree software used is widely applied in phylogenetic studies in fields such as genetics, and is one of a number of programs that linguists have applied in recent years to basic vocabulary lists to test phylogenies (see Greenhill et al. 2020 for a broad discussion about about uses and limits of phylogenetics in respect of cultural traits). We selected splitsTree for its ease of use: we do not have expertise in computer programming, but we have ensured that the analysis is transparent and repeatable. We invite other researchers to download our 116-word dataset and experiment with other software packages as they feel appropriate. Our use of lexical data is strictly predicated on our understanding of inheritance versus borrowing, based on the comparative reconstruction literature reviewed above, and our own knowledge gained over years of research on languages in the region. Lexical matches are scored as cognate (i.e. inherited from a common proto-form) and not for phonological similarity (as in approaches such as the ASJP (Automated Similarity Judgment Program)⁷ and allied methods). The splitsTree approach does not simply reproduce traditional lexicostatistics. Rather, it is Bayesian, running an algorithm to identify all patterns of shared cognates, and recalculating repeatedly through all 116 entries, eventually generating tree structure on the basis of bottom-up hierarchical clustering. The effect is to force grouping based on lexical retention, while division is indicated by semantic change / lexical replacement.

⁷ <http://email.eva.mpg.de/~wichmann/ASJPHomePage.htm>

3.1. Vietic Lects Used

Lexical data for 29 lects were assembled, 12 of which were already aggregated in the Ferlus 2007 file and the later Ferlus 2017 aggregation of Pong lects, and the remainder added from other published and unpublished sources.⁸ The full listing follows:

Languages/Lectal groups	Data Sources
Vietnamese (Hanoi)	Standard dictionaries
Muong lects	
Muong Hoa Binh	Ferlus 2007
Muong Son La	Ferlus 2007
Muong Muong Thanh Hoa	Ferlus 2007
Muong Bi	Nguyễn V. K. et al. 2002
Nguon lects	
Nguon Co Liem	Nguyễn P. P. 1997
Nguon Yen Tho	Nguyễn P. P. 1997
Cuoi lects	
Tho (Cuoi Cham)	Ferlus 2007
Tho (Lang Lo)	Ferlus 2007
Cuoi Thai Hoa	Nguyễn H. H., n.d.
Cuoi Tan Hop	Nguyễn H. H., n.d.
Phong lects	
Phong	Ferlus 2007, 2017
Toum	Ferlus 2007, 2017
Liha	Ferlus 2007, 2017
Liha	Peiros 2004 (data sourced from Nguyễn Văn Lợi)
Chut lects	
Sach	Ferlus 2007
Ruc	Ferlus 2007; Nguyen, Tran & Ferlus 1998; Nguyễn V. L. 1993
May	Babaev & Samarina 2018
Arem	Kasuga 2008
Other archaic lects	
Thavung lects	
Thavung	Ferlus 1979
Thavung Phon Soung	Ferlus 2007
So Thavung	Premssirat 2000
Ahoe	Enfield 2011
Malieng lects	
Malieng (Quang Binh province)	Samarina n.d.
Malieng (Ha Tinh province)	Samarina n.d.
Malang	Ferlus 1997
Malang Pakatan	Nguyễn V. L. and Nguyễn H. H. (2001)
Malieng Bro	Ferlus 1992
Kri (Ha Tinh Province)	Samarina n.d.
Kri Phoong	Nguyễn V. L. and Nguyễn H. H. (2019)

⁸ We take note that an anonymous reviewer correctly points out that extensive relevant lexicons were collected by the late Nguyễn Văn Lợi, and use of these would materially contribute to this study. Indeed, we were able to obtain two such word lists of Malang Pakatan and Kri Phoong, but other materials were not made available to us.

Some of the sources (particularly for Aheu-Ahlao and Harème) lack sufficient lexical data to provide confidence in their use for statistical analyses (i.e. provide less than 80% coverage of the lexical entries). Their membership in their respective lect groups in such cases is determined by key isoglosses, as described below.

1) Harème (listed in Ferlus (1996: 12) as part of the Malieng group) data includes only 50 items in our 116-word list. Nonetheless, within the Harème list, there are notable alignments with lexical items seen uniquely in Malieng, Malang, and Kri ('big/large', 'black,' 'cloud,' 'drink,' 'earth/soil,' 'mountain,' 'rain', etc.), thereby supporting the original claim of affiliation in this group. Ferlus (2014: 1) notes earlier confusion of Harème with Arem, and similarly provides evidence in the numerals 1 to 10 related it with Maleng Bro.

2) Available Aheu-Ahlao data is particularly limited: Chamberlain's study focuses on faunal terminology so there are hardly any useful lexical alignments with our 116-item list. However, Chamberlain's data do include etyma noted in other studies to be members of the Thavung group, indicating that this lect is closely related to the Thavung lects in our database. Our working assumption is that Aheu-Ahlao straightforwardly pairs with Thavung. Selected examples of Aheu-Ahlao-Thavung isoglosses are listed below:

- 'mouse' [ʔiik] occurs in all Thavung varieties, and 'porcupine' is seen in So Thavung [ɲi:⁴³] and Ahoé-Ahlao [jji], with a shared loss of Vietic *-m (Proto-Vietic *k-ɲi:m?).
- 'buffalo' So Thavung [k^huaj⁴³] and Ahoé [khwaay] are like borrowed from Lao k^hwáj 'buffalo'.
- 'duck' So Thavung [ʔa³³|tɔ:³²] and Ahoé [ʔatɔ:] and Ahlao [ʔatɛ:] are likely shared loanwords from Katuic (e.g. Proto-Katuic *ʔadaa).

Beyond the issue of Harème and Aheu-Ahlao being absent from the computational analysis, we have taken the view that the data we assembled for this experiment are reasonably representative of the diversity within Vietic, sufficient to justify proceeding in the manner we have done, and any limitations arising from these circumstances do not constitute fair reason not to proceed or report as we have done. On the contrary, we assert that the present study sets a standard for transparency and testability for Austroasiatic phylogenetic studies that is unsurpassed in print.

The creation of the 116-item list was the outcome of beginning with the Swadesh 100 and 200 lists and reconciling these with the available data with the aim of achieving at least 80% coverage for each lect in the analysis. Procedurally, we went forward as follows: sources were selected and lexicons aggregated in a spreadsheet, with rows identified with Swadesh 100 and 200 items, subject to semantic and phonological adjustments as we judged necessary. For most of the languages, full coverage of the Swadesh 100 categories was not possible, with 20 or more gaps being common. Some 40 additional categories were added from the Swadesh 200 list, based on the 40 best represented items in the aggregated data, seeking to achieve a 120-item list with at least 100 items coverage for all lects. Ultimately, achieving a 120-item list was quite problematic given internal problems with the Swadesh categories, which do not always map neatly to the semantics of mainland Southeast Asian languages, and we settled on 116 items. A number of items that we investigated had to be excluded as the data included confounding and/or limited evidence; these items and reasons for their exclusion are listed in Appendix 1.

The spreadsheet we created, with cognate assignments, and some etymological notes, and the derived nexus file, were converted into a PDF file and archived with zenodo.org.⁹ The nexus file was input to SplitsTree running default settings for generating neighbor-nets and phylograms. Phylograms were created in both UPGMA (unweighted pair group method with

⁹ See: <https://zenodo.org/record/5263195> (DOI: 10.5281/zenodo.5263195).

arithmetic mean) and NJ (neighbor-joining) analyses, producing similar branching results (discussed below). Jahai (Aslian) and Khmu (Khmuic) were included as out-groups to root the Vietic tree, with Jahai set in SplitsTree as the first out-group, consistent with received views on Austroasiatic classification.

4. Primary results

4.1. Neighbor-Net

The neighbor-net (Figure 2) results are regarded as important for identifying low-level sub-groups and to provide indications of whether the tree-structure is strongly branched or ambiguous in some way. The latter may point to contact interference or errors in coding cognates. In this case our interpretation of the neighbor-net is that the results are fairly clean, pointing to six groupings as follows: (1) Viet-Muong, (2) Cuoi-Tho, (3) Pong-Toum, (4) Chut (Ruc, May, Sach), (5) Arem, and (6) Thavung-Malieng. The precise status of Arem is unclear; while there are weak indications that it groups with the Chut lects, this could reflect a couple of unknown borrowings in the dataset, or could be a statistical artifact of missing items. Otherwise, the apparent groupings are firm.

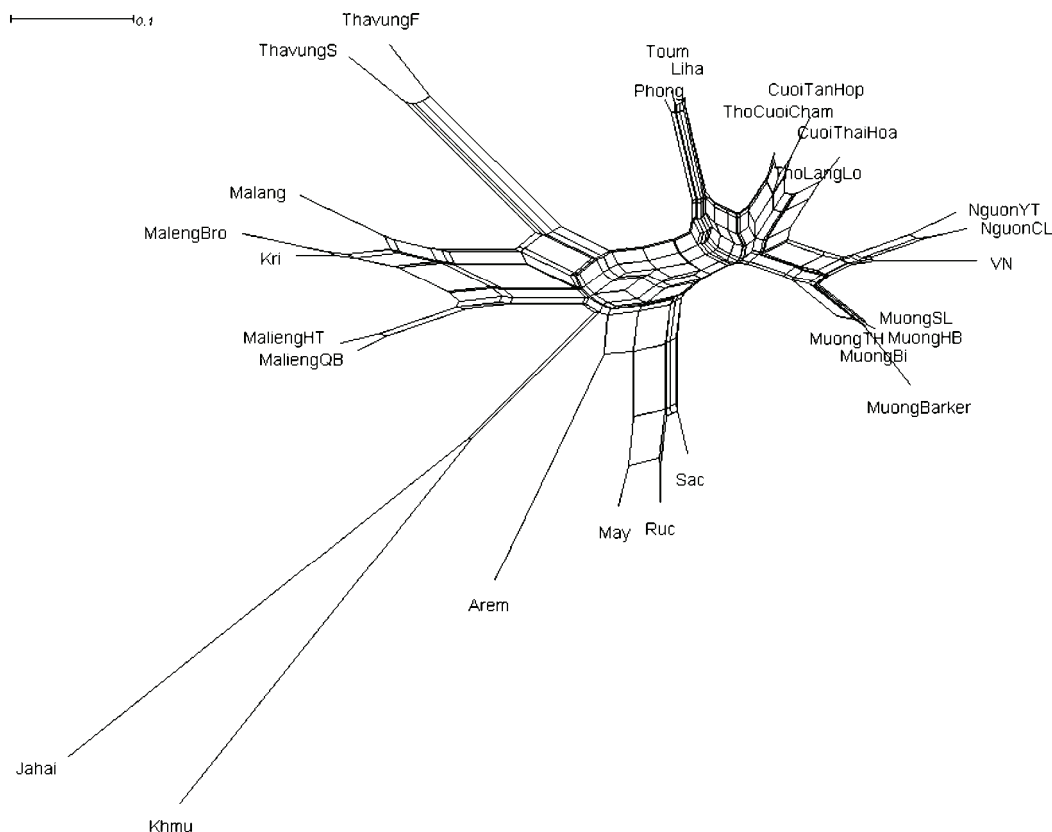


Figure 2. Neighbor-Net results for 28 Vietic doculects.

4.2. Phylogram

The UPGMA phylogram generated with SplitsTree is reproduced in Figure 3; it is strongly consistent with the neighbor-net results, while adding some clarity in terms of the apparent

nested branching. We also generated NJ and BioNJ phylograms, but effectively the same results were replicated, so those trees are not given here for space considerations.

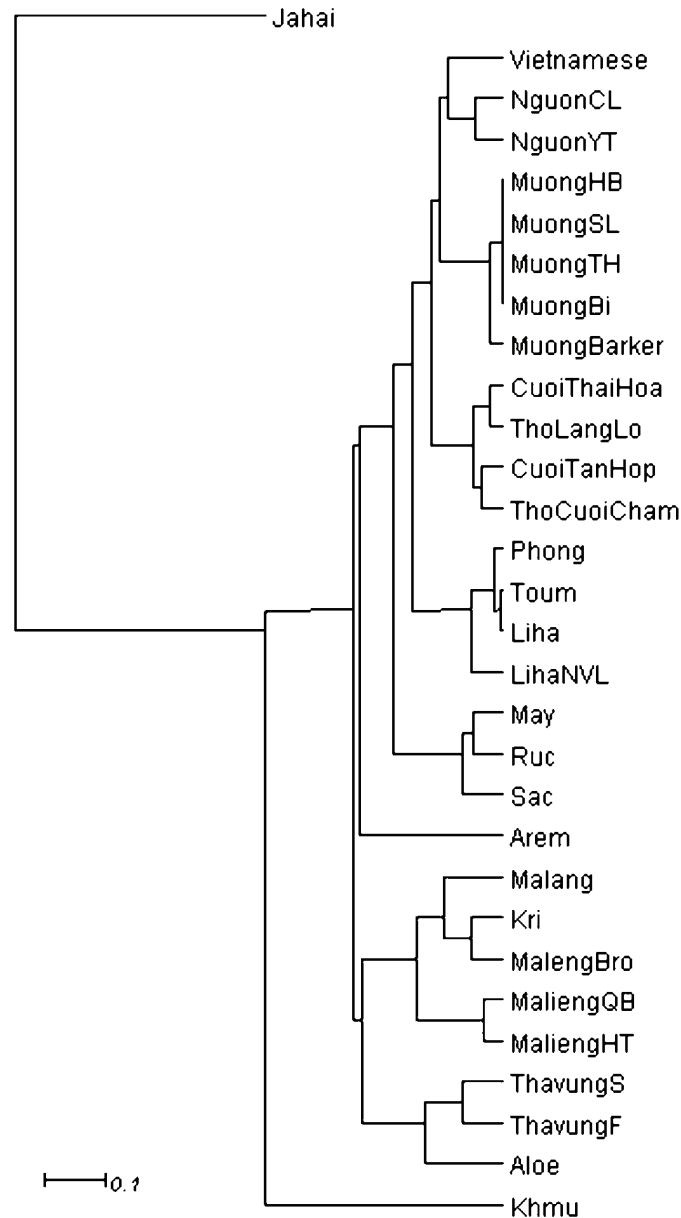


Figure 3. Phylogram results for 29 Vietic doculects, UPGMA.

4.3 Historical phonology and classification

If our computational results do approximate the real tree structure of Vietic, there ought to be correlations in terms of phonological developments reconstructable for the indicated branchings, in addition to typological features, such as syllable structure and suprasegmental features, that have been noted in specific groups, as discussed above. The distinction of the presence or absence of sesquisyllables among Vietic languages is a prominent feature by which relatedness among the sub-branches can potentially be evaluated, and previous approaches have used this as a key criterion. However, based on this feature alone, a finer analysis cannot be achieved, and segmental features must be considered.

Unfortunately, the issue of identifying shared or contrasting segmental developments in Vietic is quite problematic. Ferlus' (2007) reconstruction is a starting point, and we largely

concur with his assignment of proto-language values to segments. As is often the case when applying the comparative method, it is apparent that segmental correspondences fall broadly into two types: (1) a core of relatively regular correspondences each attested in multiple etymologies, and (2) a mass of unique and/or apparently irregular correspondences that are difficult to explain or have competing explanations which are difficult to discriminate.¹⁰ Confining our consideration to the first type, we have found no strong bases to identify nested branching relations in terms of syllable onsets or vowels. Each of the six identified subgroups mostly show onset and vowel reflexes that are variously quite stable, or they show variation within each subgroup that does not link across multiple sub-groups.

proto-Vietic	*-h	*-s	*-r	*-l
Viet-Muong	-∅	*-lh > -j Viet -j/-c/-n Muong -t/-n Nguon	*-l > -j Viet -j/-l/-n ¹¹ Muong -n Nguon	*-l > -j Viet -l/-n/-u Muong -n Nguon
Cuoi-Tho	-∅	*-l > -l/-n/-j	*-l > -l/-n	*-l > -l/-n/-∅
Pong-Toum	-∅ [-h only in loans]	*-c > -t/-c/-k	*-l > -l/-n	*-l > -l/-n
Chut	-h	*-lh > -h/-l/-lh/-rh	*-l > -l	*-l > -l
Arem	-h	*-h > -h	*-l > -l	*-l > -l
Thavung-Malieng	-h/-ʔ ¹²	*-s > -s/-j/-jh/-uqh/-jʔ	*-r < -r/-uq/-l/-n	*-l > -l/-n

Table 3. Vietic reflexes of codas *-h, *-s, *-r, *-l

However, close examination of coda correspondences allows us to identify the evolution of proto-Vietic *-h, *-s¹³, *-r, *-l as potentially relevant to this question. There is considerable variety—and seeming instability—both among and within the lectal groups in the modern reflexes of these sounds, including a range of segments, phonation features, and tones. The data are summarised in Table 3; we include our branch-level reconstructions to highlight the common patterns.

The reconstruction indicates that Vietic *-r and *-l remained in contrast within T-M while merging to /-l/ in the rest of Vietic (notwithstanding some later mergers with -l in some T-M lects). This is consistent with the hypothesis that T-M branched directly from the proto-Vietic node and is the most conservative subgroup in respect of this segment.

Proto-Vietic *-h was already discussed as an indicator of branching structure by Hayes (1992), but we now have a clearer understanding of its significance. Hayes' phylogram is simple (see Figure 4), treating only Thavung (TV), Ruc (RU), Muong (MU), and Vietnamese (VN). He found that the loss of *-h is shared within Viet-Muong, distinguishing it from Ruc and

¹⁰ Shorto (1976) discusses this problem at length in relation to Austroasiatic reconstruction, although we do not necessarily endorse specific proposals for dealing with these issues.

¹¹ One example of -n was found by us in one Muong word list.

¹² Some apparent instances of -ʔ reflexes of *-h are noted for Maleng. Otherwise, -h is general in TM lects.

¹³ The reconstruction of *-s is Ferlus' reconstructed value. We take the view that this segment was probably a laminal fricative with post-alveolar or pre-palatal constriction given the tendency for palatal reflexes in some Vietic lects, but retain the *-s formalism for consistency and simplicity.

Thavung. The more complete data we have aggregated indicates that the loss of *-h is also shared with Pong-Toum and Cuoi-Tho. This is potentially a shared innovation at the VM-P-C node in the tree, yet it may not be so straightforward. Vietic *-h was not simply lost as a segment, but rephonologised as glottal tension in the syllable nucleus, yielding the *hỏi* and *ngã* (traditionally considered Category-C tones) tones of Vietnamese and their equivalents in other lects. While modern Vietnamese dialects vary in phonetic realizations of tones, in addition to contour, it is common throughout regions of Vietnam for these two tones to exhibit phonation features (i.e., glottalization or breathiness, again varying according to local varieties) that are likely residual evidence of the earlier segments. Such is the case as well in Cuoi in Tan Ki district, in which this tone category is glottalized (Nguyễn and Nguyễn 2019:lxii).

Effectively the VM, PT, and CT groups share the rephonologization of *-h as a Category-C tone. However, it is not presently determined whether these tonal developments were one or multiple independent events. This is a complex topic whose details are beyond the scope of this study, but for the present purposes, we will suppose that the single rephonologization of *-h at the VM-PT-CT node is likely and represents a relevant finding which is consistent with the identification of this node in the computational analysis.

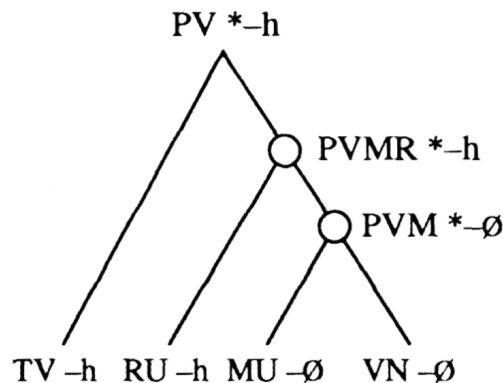


Figure 4. Hayes (1992: 216) phylogram based on development of Vietic *-h.

Also relevant is the development of Vietic *-s. As indicated in Table 3, each subgroup has reflexes of *-s which appear to be distinct, and in most cases show multiple secondary developments within each group. Areally, coda [-s] is unstable, and across multiple language families shifts to [-t] and [-h] are not uncommon.¹⁴ The shift of [-s] to [-l] and [-n] are also unproblematic; the shift of [-s] to [-l] is assumed to be via a voiceless lateral, written *lh* here. A subsequent change of [-l] to [-n] is similarly common in Tai,¹⁵ and it is unsurprising that it happened in Vietic considering the shared typological and historical context.

The phonological reconstruction indicates that, like *-r, Vietic *-s was unchanged in proto-Thavung-Malieng and later developed a range of mostly voiceless approximant articulations. As other subgroups branched off on the Eastern side of Vietic, the *-s underwent independent developments in Arem, Chut, and Pong-Toum. Given the similarity of the reflexes of *-s to the reflexes of *-l in Viet-Muong, and the obvious merger to *-l in Cuoi-Tho, we propose that *-s shifted to *-lh in proto-Cuoi-Tho-Viet-Muong, becoming voiced in Cuoi-Tho. The various /-t, -c/ reflexes in Pong-Toum clearly suggest a sequence *-l > *-lh > *-ç > *-c with lateral friction shifting to palatal friction before hardening to a stop coda.

¹⁴ For example, in Austroasiatic, *rma:s 'rhinoceros' > Sre *rəmis*, Khmer *rəmiəh*; *ʔas 'to swell' > Sre *ʔas*, Khasi *ʔa:t*; *ris 'root' > Mlabri *re:lh*, Chong *rè:t*, Car *ʔeh* (reconstructions from Shorto 2006).

¹⁵ For example, Proto-Tai *bil^A 'to fly' > Siamese *bin^{A1}* (Pittayaporn 2009).

Overall, we can say that our analysis of the Vietic codas *-h, *-s, *-r, *-l is readily reconciled with our computational phylogenetic results, and we take this as pleasingly confirmatory. Below, we first assess the resulting individual lectal groups and characterize the overall phylogenetic findings.

4.4. Summary findings

Figure 5 provides a schematic representation of our overall findings. The computational and phonological results converge on the identification of five subgroups in the branching configuration indicated in the figure. We have a high level of confidence that this is a good approximation of the real family history of Vietic. We are not in a position to make strong claims about the internal configurations of each sub-group; generally, the lexical differences between individual sub-group members are moderate and may not be strongly significant in the context of the 116-item list (with the exception of Arem) although other factors may also be brought to bear in discussing the subgroup internal relations. Below we discuss each subgroup separately.

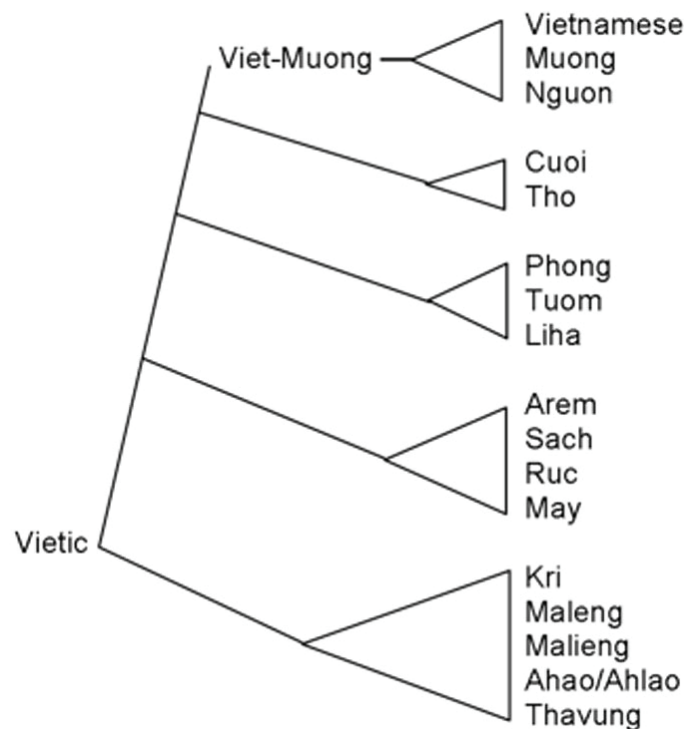


Figure 5. Vietic classification proposed in this study.

Viet-Muong (including Nguon): Our results confirm the general view that the Vietnamese, Muong, and Nguon form a coherent VM sub-branch, what can be considered the Northeast clade of Vietic. While Vietnamese as a national language is spoken throughout Vietnam, Muong lects are largely concentrated in northern Vietnam. Nguon is spoken in north-central Vietnam in Quang Binh Province close to the Chut group. Interestingly, Vietnamese is grouped closer to Nguon than to Muong in the tree, with the latter so tightly clustered that the Muong lects could be regarded as one language lexically. This is contra Phan's (2012) results based on phonological changes, but unremarkable given their real lexical uniformity. The present results are based on counting cognates regardless of phonological changes, so as a lexical result this should not be controversial, and we note again that the question of VM internal

structure has no bearing on the larger question of the higher order branchings. We note here also that in addition to the grouping based on lexical data, Vietnamese, Muong, and Nguon are typologically unified and distinct from most other Vietic groups. VM languages have complex tone systems, generally 5 or 6 tones based on a 3-by-2 tone pattern, with mergers resulting in the loss of a tone. They have only monosyllabic morphemes, and syllables have at most $C_1(C_2)V(C_3)$ structure in which the C_2 can only be a liquid or glide (and in Vietnamese only /w/). They are also both recipients of much larger quantities of Chinese loanwords than any other Vietic lects.

Cuoi-Tho: The results of our study confirm the grouping of the Cuoi-Tho lects, which are geographically farther north than other non-VM lects, and closer to VM. The results of the phylogenetic calculations divide the four lects in an unexpected way: one each with Ferlus' and Nguyen's fieldwork lects. In other words, those authors' two lects do not appear to group more closely to each other, although these differences are marginal for our purposes. Just as Cuoi-Tho are lexically closest to VM languages, they are also typologically closest: they are monosyllabic and have tone systems that parallel those of VM structurally without being neatly cognate. Of particular interest is that, in our data, Cuoi-Tho and Pong-Toum do not form a single node, in contrast with some previous analyses.

Pong-Toum: The available lexical data clearly shows that Pong, Toum, and Liha form a group, but also, crucially, are distinct from the Cuoi-Tho group. In contrast to the clear results of the lexical data, the lack of consistent, complete phonological descriptions presents challenges. As for syllable shapes, Nguyen T. L. (1992: 101) posits that approximately ten percent of Pong words are sequisyllabic (i.e. bisyllabic words with iambic stress, resulting in the pre-syllables that are unstressed and a limited number of consonants and neutralized vowels), and Ferlus (2014: 1) mirrors this point: it is a point of distinction between PT and CT. However, while Pong has been described as having a 4-way system divided by oppositions in registral height and phonation (Nguyen T. L. 1992: 99), in the current lexical data, Liha (in Peiros 2004 from Nguyen V. L.) has six tonal contours. If this is the case, developments in the suprasegmental systems must have occurred after this subgroup separated from Vietic. These lects do appear to be typologically transitional between the monosyllabic, tonal VM and CT groups and the remaining archaic varieties, the latter having higher rates of sesquisyllabic roots and simpler tone systems or register-phonation systems, as described below.

Chut-Arem: To the southernmost extent of the Eastern Vietic clade, Sach, Ruc and May are phonologically archaic lects that clearly form a compact sub-group lexically, generally called 'Chut'.¹⁶ Arem is similarly archaic, but lexically divergent, such that it does not obviously fall into Chut based on the computational results. However, we do identify some five Chut-Arem isoglosses in our 116-item list, and these are listed in the first five rows of table 4, along with representative glosses for the other Vietic sub-groups. We propose that these are best regarded as common inheritances from proto-Chut rather than borrowings of Chut lexical innovations into Arem, as the latter seems unlikely to us.

Additionally, an anonymous reviewer suggested that possible Katuic influence on Arem and Thavung-Malieng should be investigated in this regard. In our 116-item list, we find only one possible shared Katuic loan into Arem and T-M, 'horn', although on phonological grounds it is striking that the Arem form more closely resembles Bahnaric reflexes, while only the T-M

¹⁶ Nguyễn Phú Phong et al. (1998) note this as a Chamic loanword. Thurgood (1999: 315) reconstructs Proto-Chamic *cət 'mountain range', but suggests that the vowel is indicative of an Austroasiatic loanword into Chamic. It does not appear to be a wider Austronesian word or even Malayic. However, as we cannot locate any comparable forms in Austroasiatic, including neighboring Bahnaric and Katuic languages, it does seem to be a borrowed exonym.

forms match closely the Katuic etymon. One other Arem-T-M isogloss is evident in our 116-item list, namely ‘earth/soil’. The form resembles Katu *katiek*, but if those forms are related, the vowel discrepancy suggests independent borrowing or formation of the word in Arem and T-M versus Katuic. The origin of this ‘earth/soil’ etymon is obscure, although it may be a phonological deformation of the Vietic etymon, whereby the coda of *tət became velar, and a pre-syllable was acquired. While an exhaustive etymological investigation might potentially reveal a very differently proportioned distribution of isoglosses, restricting ourselves to the present dataset, we tend to the view that Arem is an aberrant Chut lect.

Gloss	Arem	Sach	Viet.	Cuoi-TanHop	Liha	So-Thavung	Notes
‘beast/udder’	nɔ̃ː	nɔ̃ː ³	vú	ʔu ¹²	now	təmʔək	Lao ʔək ‘chest’
‘nail/claw’	ɲtɔŋ	katoːŋ ¹	móng	săm ²²	sam	kasâm	
‘fat/grease/oil’	tlùŋ	tluŋ ²	mō	mɔ̃ ³³⁷	mə	ʔatûː	
‘navel’	uɬuh	duduːl ³	rón	sun ¹²	suːɲ	cuʔbîː	
‘night’	lum	lúm	đêm	tem ³²	tɛːm	ʔamáh	
‘earth/soil’	atāk	bʔnʔ	đát	tʂt ¹²	tət	ʔaták	Katu <i>katiek</i>
‘horn’	takaːj~təkeː	ʔəŋ ²	sùŋ	khʂjŋ ⁵⁴	k ^h lɔŋ	takôj	Katu takɔj, Bahnar ʔəkeː

Table 4. Arem isoglosses

Thavung-Malieng: These lects, some of which are generally tonal and highly registral (making lexical use of creaky and breathy phonation), appear to unambiguously stem together from the highest branching node in the tree. Phonologically it appears that the TM languages are conservative in regard to the outcomes of proto-Vietic coda *-r, with the ancestor of the Eastern clade undergoing a general merger of *-r and *-l to *-l, while reflexes of *-r diversified later within TM. Additionally, there appear to be Tai loans (e.g. ‘good’, ‘to know’, ‘small’ from Tai) within TM that are not shared with the rest of Vietic (and thus likely within more recent centuries), which speaks to the unique history of the subgroup after the initial split of Vietic into Western and Eastern clades.

Given the considerations discussed above, the tree structure we propose has a compelling logic to it. Taking the West-East split as primary, the TM languages have their own history of lexical, segmental and tonal developments; on the Eastern side, the highest branching nodes reflect archaic lects that retain sesquisyllables and show only limited tonogenesis. Below them in the Eastern branch are the more innovative subgroups, with VM lects nested at the extreme north with the most innovative lexical and phonological histories.

4.5 Previous studies with coinciding results

Vietic phylogenies proposed by other scholars also reproduce specific aspects of our findings, although the bases for these are not always clear. We have also presented a cursory overview of a dozen previous classifications of Vietic languages in Table 2. Below, a few of these are presented with complete trees and discussion of the overlap with our current model.

James Chamberlain has authored several studies analyzing Vietic vocabulary, primarily faunal lexicon, and this has informed several classifications. Phylograms are offered in both 1998 and 2018 papers (Figures 6 and 7), and both anticipate our primary result of a binary division between TM and a clade covering the rest of the Vietic languages, notably indicating a closer connection between Chut and VM. In Chamberlain’s (1998) model, neither primary

clades are named, while the lower sub-groups have mostly neutral geographical names. The structure of each primary clade is flat, one with three groups: Viet-Muong, Pong-Toum, and Chut and the other with four TM lect groups.

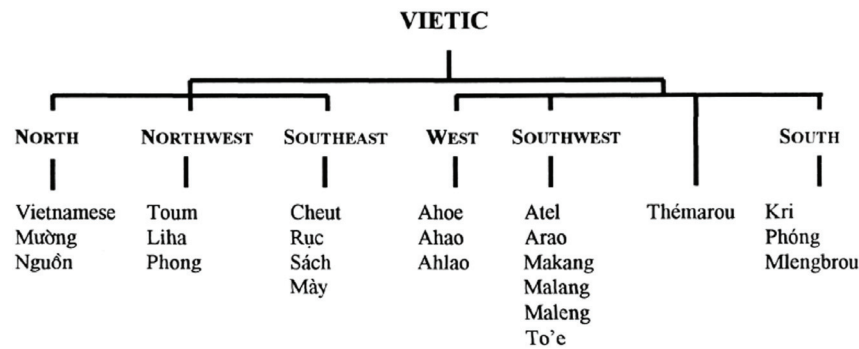


Figure 6. Chamberlain (1998: 106) “Suggested modifications to Vietic subgrouping”.

Chamberlain (2018) offers a somewhat refined classification (Figure 7) that indicates some nested branching. Additionally, he invents unique names for language sub-groupings and nodes in the tree which are unrelated to equivalent commonly used terms. Specifically, Pong-Toum and Chut are grouped in opposition to Viet-Muong, and multiple levels of nested branching are indicated within the TM clade, which is called “Nrong-Theun”. While showing closer affiliation of Chut with VM, as opposed to the other bisyllabic Vietic lects, as our model does highlight the closer connection of Pong-Toum and Cuoi-Tho lectal groups with VM, all of which are monosyllabic (with a small number of sesquisyllables noted in Pong-Toum) and highly tonal.

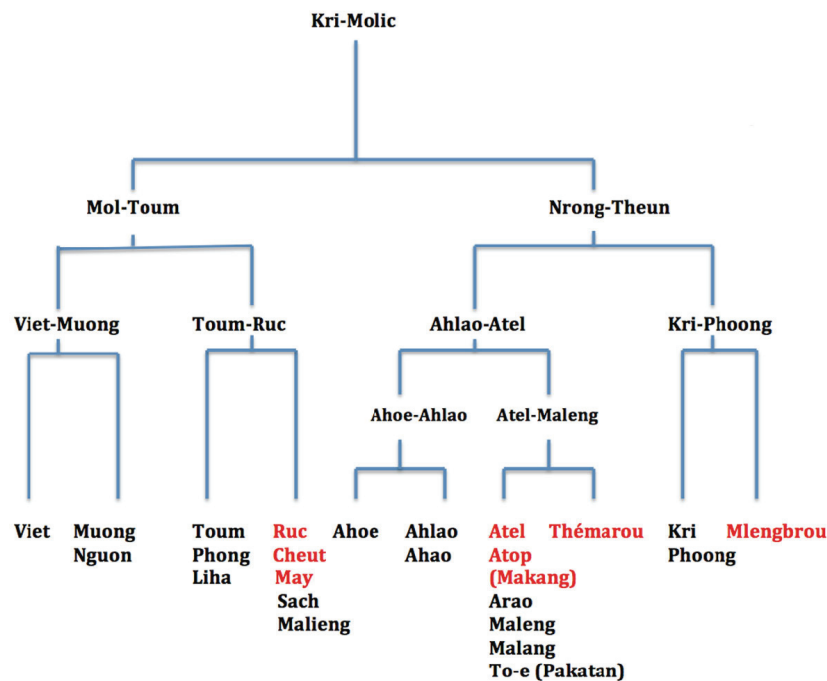


Figure 7. Chamberlain (2018: 12) Kri-Molic (Vietic) phylogram based on faunal lexicon.

The other proposal of note is the lexicostatistical analysis by Peiros (2004), which is based on the Swadesh 100 items, and processed with the StarLing software package (see Figure 7). Peiros’ results closely parallel ours in respect of the lower branchings, notably offering similar

configurations of Viet-Muong and the Pong-Toum and Cuoi-Tho clades (although Liha is displaced by one node). At the higher levels, Peiros does place Thavung on a primary branch opposed to the rest of Vietic, as we do, but he does not group Thavung with Malieng, and this is difficult to account for given the rather strong sub-grouping of these evident in our dataset, which is very similar to that of Peiros. We assume that the main reasons for the differences between Peiros' results and ours relate to the shorter wordlists (100 versus 116 items) and his particular assessments of cognates and loans. Overall, we take this comparable precedent as reinforcing the overall results we have obtained from both computational data as well as the comparative method in terms of segmental and prosodic issues.

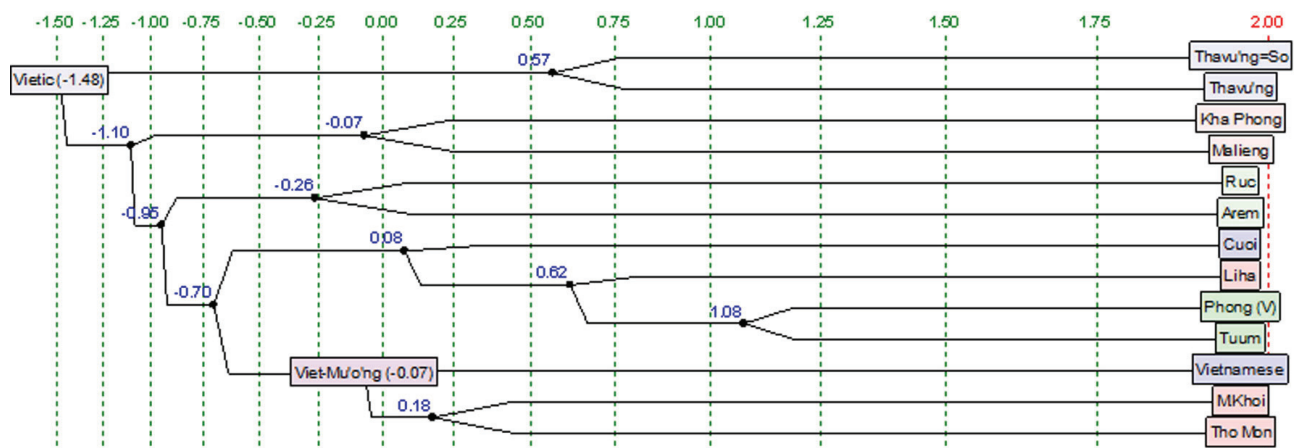


Figure 8. Peiros (2004) Vietic phylogram based lexicostatistics (100 word list).

5. Historical Geographic Considerations

The modern position of Vietic groups is telling of the geographic history of the greater Vietic speech community. Until the late 1440s, the southernmost extent of Vietic speakers was in north-central Vietnam, and the great expansion southward only began in the late 1400s in the period known as the *Nam Tiến*, or the “Southward Advance,” after the fall of the Cham polities following centuries of Cham-Vietnamese military conflicts.¹⁷ This expansion led to the movement primarily of those speaking Vietnamese, which was already a fully distinct language group from the rest of Vietic. The southernmost reaches of Vietnam (e.g. the Mekong Delta) were populated by Vietnamese speakers only in the past few hundred years. Thus, the modern map of Vietic languages from north-central to northern Vietnam (see Figure 9) may be a reasonable representation of the geographic distribution of Vietic groups in the pre-expansion period.

Archaeological, genetic, historical, and linguistic data altogether show that the heartland of Vietnamese is the Red River delta, going back to the Han period of Chinese colonization (cf. Alves 2021), and consistent with this, the Muong lects are spoken inland not far to the west of the delta. The Nguon speakers are located much further south, but this may reflect the relocation of a garrison recruited in the Viet-Muong area under Chinese command (Ferlus pers

¹⁷ It is striking that this follows the fall of Angkor, traditionally dated to 1431 after being sacked by Ayutthaya, and these days considered to have been aggravated by climate change that made the Mekong Delta a more attractive locus for agriculture (Penny et al. 2018). Regardless of the causes, this period is one of considerable political change and migration in the region, and thus one in which social conditions contributed to a dynamic linguistic environment.

com.). Even if this specific historical scenario is not the case, the close affiliation of Nguon with the rest of VM suggests that this outlying geographic location is the result of migration of a VM language, and so this does not affect the identification of the Viet-Muong historical locus in the north.

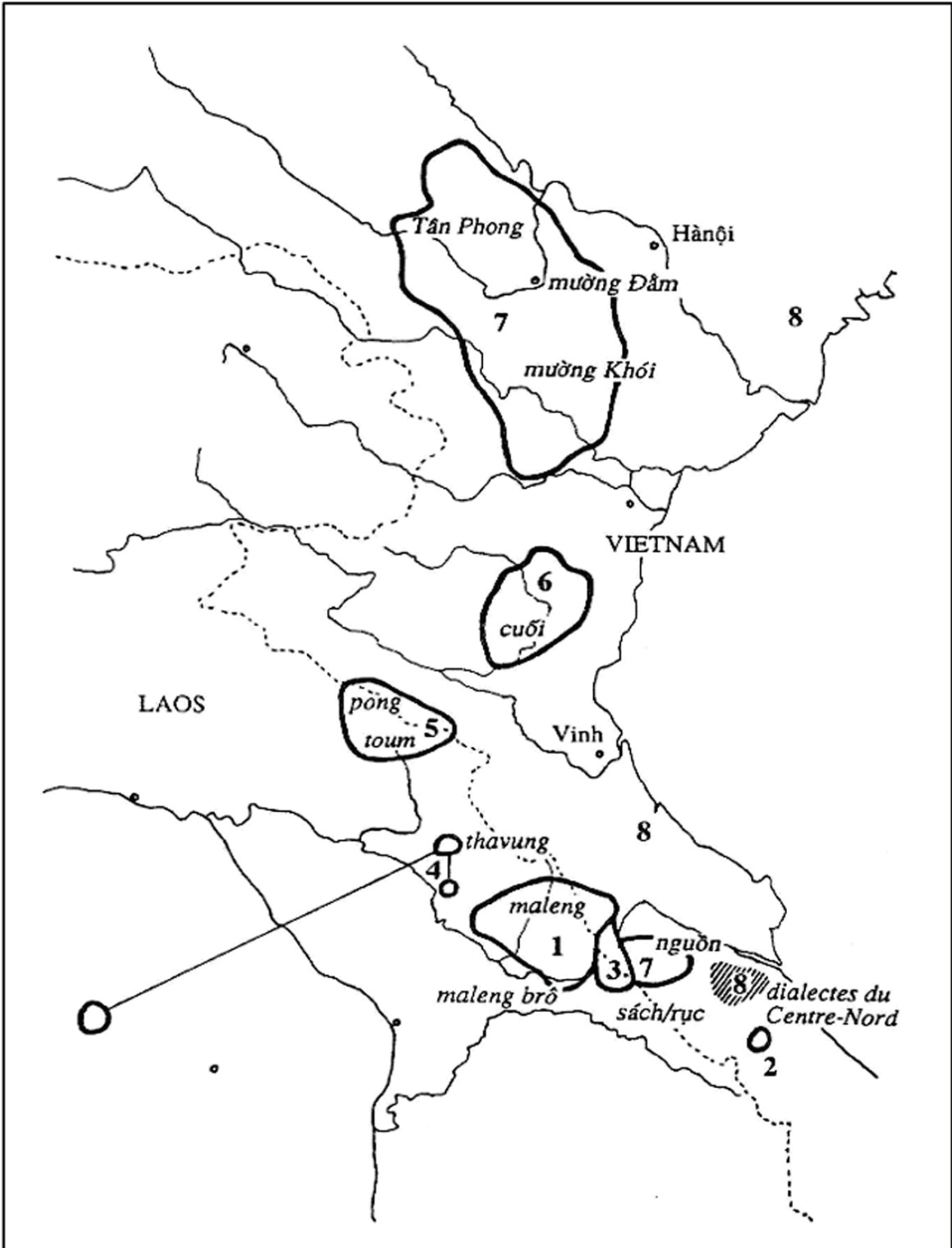


Figure 9. Map of Vietic languages approximate locations (Ferlus 1998: 27)

Other Vietic subgroups are all located south of the delta, mainly in the hills of the Annamite Range, plus some migrations further afield, such as the Thavung in Thailand documented by Premrirat (2000). The geographical distribution broadly follows our proposed phylogeny; the Pong-Toum and Cuoi-Tho subgroups are located closer to Viet-Muong, being roughly in the centre of the Vietic range. The more archaic subgroups Chut, Arem, Thavung-Malieng are further south, with Thavung-Malieng to the west of Chut and Arem, mostly on the Lao side of the border. Assuming a least-moves analysis, this places the locus of genetic diversity in Vietic approximately in the north-west of Quang-Binh Province, implying that Viet-Muong moved northwards into its present range. However, ancient loanword evidence suggests otherwise.

It is apparent that there was a much older phase of Tai-Vietic contact before proto-Vietic began to disperse, potentially dating to the BCE-era Dongson period, evidenced by a number of proto-level items, e.g. ‘duck’, ‘winnow basket’, ‘drum’, ‘water pipe of bamboo’, ‘water spinach’, ‘bush-knife’, etc. (Alves 2021: 25). This is also consistent with genetic material in human remains from the Dongson-era Nui Nap archaeological site in the Red River Delta, showing shared alleles of both Tai and Kinh (i.e. ethnic Vietnamese) populations (Lipson et al. 2018: 2).¹⁸ This was also the period prior to the dispersal of Southwestern Tai, so we expect that Tai was more geographically constrained to the northern parts of Vietnam and bordering areas of southeastern China.

If our interpretation of these old loans is correct, a Vietic homeland in the Red River Delta or proximal region of Northern Vietnam is supported, the issue being to explain the distribution to the south of the other Vietic subgroups. Indeed, we find it provocative that the Man Bac site (in the Red River Delta) of the Phung Nguyen culture (c. 4000–3500 BP) is among the earliest sites with evidence of incoming agriculturalists from the north (e.g. Matsumura et al. 2008), and that there is a demonstrated sequence of archaeological cultures from the Phung Nguyen culture to the Dong Son culture (e.g. Kim 2015: 106). Thus, in the larger picture, the spread of Vietic is tied to and may follow from the original Austroasiatic dispersal. Moreover, it is surely relevant that the range of Vietic languages approximates the reach of Chinese administrative control in the Han period and some Vietic groups may have moved south and upland to avoid Chinese control, or may have been moved southward by the Chinese to help provide a buffer to the Chams. Another possibility is that, when Sinitic speakers arrived, Vietic speakers were already spread from the Red River Delta south into Thanh Hoa and areas of north-central Vietnam and bordering parts of modern-day Laos. In this situation, the extent of the more typologically innovative Viet-Muong shows the approximate range of a substantial Sinitic presence and language contact from the early first millennium CE, but whether there was Sinitic language contact with other Vietic languages in that early period is as yet unclear. However, to determine which scenario is most viable, we need clearer indications of the disposition of Vietic groups before Han period, and in this respect, perhaps all we have to go on are the archaeological continuities in the Red River delta before and after Chinese domination, hinting that the Viet-Muong have a deeper history of settlement in the north.

There remains the question of to what extent language contact is relevant to the higher branching of Vietic, which may be a potentially rich area for subsequent study. There are clear Tai loanwords among TM languages, and relevant items from our 116-word list are aggre-

¹⁸ Lipson et al. (2018) also note the shared alleles encompass Austronesian speakers. However, this is harder to resolve as (a) Austronesian is a huge language family with many branches, making it difficult to interpret in a meaningful way, (b) there are no archaeological indications of an Austronesian community in northern Vietnam in the Dongson period, and (c) the connection between Kra-dai and Austronesian is increasingly accepted, making a claim of both Austronesian and Tai from the mid-1st millennium possibly redundant.

gated in table 5. Comparisons are made with Lao, Thai and Shan to establish the Tai origins, although local Phu Thai¹⁹ may be an important source of influence. Additional Tai loanwords can be found among these lects (cf. Hayes 1982 on Tai loanwords in Thavung), but in terms of our 116-item list, they are not relevant. The Thavung spoken in Northeast Thailand has in particular a large number of probable recent Lao (Isaan Thai) loanwords, including grammatical ones, which is apparent from examining the dictionary of Premsrirat (2000) and the Thavung So grammar of Srisakorn (2008).

Gloss	Lect	Item	Compare
breast/udder	Thavung (S)	təmʔók~təpʔók	Lao ʔók ‘chest, breast’
chest	Thavung (F)	ʔək¹	
cloud	Thavung (F), Malang	me:k	Thai mē:k ‘cloud’
good	Thavung (S), Thavung (PS), Thavung (F), Pakatan, Malang	di:	Thai, Lao di: ‘good’
hard/solid	Maleng Bro	lɛ:ŋ¹	Thai rɛ:ŋ ‘forcefully, hard’
to know (facts)	Thavung (S); Thavung (F); Kri	húʔ; huʔ; ru¹¹ / zu¹¹	Lao hú:, Thai rú: ‘to know’
mountain	Thavung (S)	phú:	Lao pʰú: ‘hill, mountain’
person	Malang	khon	Lao kʰón ‘person’
small	Thavung (S)	ʔi:t	Shan ʔit⁴ ‘small’
to speak/say	Thavung (F) Kri vǎ¹¹	vaw⁴	Lao wâu ‘to speak/say’
walk	Thavung (S)	ʔaŋa:ŋ	Lao p̄a:ŋ ‘to walk, stride’

Table 5. Tai loans in Thavung-Malieng basic vocabulary

Another indicator of language contact is seen in noun phrase structure: while Chut languages show noun phrase structure that parallels that of Vietnamese, So Thavung mirrors the Lao/Thai type, while Kri has flexible word order, allowing for either type, all of which suggests substantial language contact with Vietnamese and Laos respectively (Alves 2020). In contrast to the TM lects, the Chut lects in our small-scale study show no Tai loanwords in their basic vocabulary. Not surprisingly, in our larger data sets, likely Vietnamese loanwords have been discovered, including some grammatical lexemes, likely to have been borrowed in recent decades.

Broadly, we can say that Tai influence is strongly evident in the TM lects, while Vietnamese influence (especially recent influence) is apparent in Chut lects. This speaks to the geographical history, indicating that TM has been isolated from the rest of Vietic on the western slopes of the Annamite Range for a significant period of time, but the relevant chronology is unclear without a more detailed study.

6. Concluding observations

In this short paper, we report on our phylogenetic analysis of the Vietic languages, summarising our results in the family tree given in Figure 5. The computational phylogenetics based on a 116-item basic word list, and the historical phonology of Vietic codas *-h, *-s, *-r, *-l, converge on the identification of five subgroups, Thavung-Malieng, Chut-Arem, Pong-Toum, Cuoi-Tho, and Viet-Muong. The tree has a simple descending binary branching structure, with

¹⁹ Phu Thai is a broad term used to refer to diverse upland Tai, and available lexicons do not readily permit accurate localization of specific lexical forms.

Thavung-Malieng and Eastern Vietic being the first split, and Viet-Muong the last. The configuration of language relations within the subgroups is largely untouched by our study as the limited data set permits only rather low-resolution results, and we would welcome more richly detailed analyses. While the geography of the languages is suggestive of a locus of dispersal in the southern end of the Vietic range, indications from old loanwords, archaeological studies, and consideration of the history of Chinese occupation equally suggest a northern locus of dispersal, although further research is needed to clarify this important question. The data on which our conclusions have been reached is available for download, we offer this as an example of transparency and will gladly cooperate with any researchers wishing to reproduce or augment this study.

Appendix 1: Excluded words and rationale for exclusion

Item	Remarks
‘all’	The ‘all’ gloss is frequently associated with the meaning ‘all gone/ nothing left’ rather than the concept of ‘every item in a set’.
‘cold’	The languages of the area often distinguish between internal sensation of ‘cold’ versus objects being ‘cold’, and the sense of ‘cool’ in contrast with ‘cold’ adds further challenges.
‘to come’	The meaning is not clearly distinguished between ‘arrive’ and being on the way to a destination.
‘man/husband’	Languages vary as to distinguishing ‘man’, ‘husband’, or ‘male’.
‘many/much’	The meaning was not well represented in the sources, and items present many are suspected of being Vietnamese loans.
‘name’	The meaning was not well represented in the sources, and where items were present many were suspected of being Vietnamese and Tai loans.
‘not’	Multiple negators were typically listed and no clear basis for identifying comparable items was apparent.
‘seed’	Lists did not clearly distinguish ‘seed for planting’ versus ‘seed found within edible fruit’.
‘sun’	‘Sun’ is commonly encoded by a compound ‘eye (of) day’.
‘that’	Languages vary in their demonstrative systems, with cases of multiple distal forms rendering comparison difficult.
‘we (incl.)’	There were too many gaps and inconsistencies in the lists of relevant pronominal forms.
‘woman/wife’	Languages vary as to distinguishing ‘woman’, ‘wife’ and ‘female’.
‘yellow’	Most lists simply had reflexes of Chinese 黃 <i>huáng</i> .

References

- Alves, Mark. 2003. Ruc and Other Minor Vietic Languages: Linguistic Strands Between Vietnamese and the Rest of the Mon-Khmer Language Family. In: Karen L. Adams et al. (eds.). *Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society*: 3–19. Arizona State University, Program for Southeast Asian Studies.
- Alves, Mark J. 2006. Linguistic research on the origins of the Vietnamese language: An overview. *Journal of Vietnamese Studies* 1.1–2: 104–130.
- Alves, Mark J. 2017. Etymological research on Vietnamese with databases and other resources. In: *Ngôn Ngữ Học Việt Nam, 30 Năm Đổi Mới và Phát Triển (Kỷ Yếu Hội Thảo Khoa Học Quốc Tế)*: 183–211. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Alves, Mark. 2020. Initial steps in reconstructing Proto-Vietic syntax. In: Mathias Jenny, Paul Sidwell, Mark Alves (eds.). *Austroasiatic Syntax in Areal and Diachronic Perspective*: 46–81. Boston: Brill.
- Alves, Mark J. 2021. The Đông Sơn Speech Community: Evidence for Vietic. *Crossroads* 19 (2021): 1–41.
- Babaev, Kirill V., Irina V. Samarina. 2018. *Materials of the Russian-Vietnamese linguistic expedition: May language* (in Russian). Moscow: Yask Publishing House.

- Barker, Milton E. 1963. Proto-Vietnamuông Initial Labial Consonants. *Văn-hoa Nguyệt-san* 12.3: 491–500.
- Barker, Milton E., Muriel A. Barker. 1970. Proto-Vietnamuong (Annamuong) final consonants and vowels. *Lingua* 24.3: 268–285.
- Blagden, Otto. 1913. The classification of the Annamese language. *Journal of the Royal Asiatic Society*: 427–432.
- Cadière, Léopold. 1905. Les hautes vallées du sông Gianh. *Bulletin de l'École Française d'Extrême Orient* 5: 349–367.
- Chamberlain, James R. 1998. The origin of Sek: implications for Tai and Vietnamese history. In: S. Burusphat (ed.). *The International Conference on Tai Studies*: 97–128. Institute of Language and Culture for Rural Development, Mahidol University, Bangkok, Thailand.
- Chamberlain, James R. 2003. Eco-Spatial History: a nomad myth from the Annamites and its relevance for biodiversity conservation. In: X. Jianchu, S. Mikesell (eds.). *Landscapes of Diversity: Proceedings of the III MMSEA Conference, 25–28 August 2002*: 421–436. Lijiang: Center for Biodiversity and Indigenous Knowledge.
- Chamberlain, James R. 2018. *A Kri-Mol (Vietic) Bestiary: Prolegomena to the Study of Ethnozoology in the Northern Annamites*. *Kyoto Working Papers on Area Studies No. 133*. Kyoto University.
- Chazée, Laurent. 1999. *The Peoples of Laos: Rural and Ethnic Diversities*. Bangkok: White Lotus.
- Chéon, A. 1907. Note sur les Dialectes Nguon, Sac Et Muong. *Bulletin de l'École Française d'Extrême-Orient* 7.1–2: 87–99.
- Enfield, Nicholas J., Gérard Diffloth. 2009. Phonology and sketch grammar of Kri, a Vietic language of Laos. *Cahiers de linguistique Asie orientale* 38.1: v–69.
- Ferlus, Michel. 1974. Le groupe viet-muong (Recherches dans le cadre de l'Atlas Ethnolinguistique). *Asie du Sud-Est et Monde Insulindien* 5.1: 69–77.
- Ferlus, Michel. 1975. Vietnamien Et Proto-Viet-Muong. *Asie du Sud-Est et Monde Insulindien* 6.4: 21–55.
- Ferlus, Michael. 1979. Lexique Thavùng-Français. *Cahiers de Linguistique, Asie Orientale* 5: 71–94.
- Ferlus, Michel. 1991. *Vocalisme du Proto-Viet-Muong*. Paper circulated at the Twenty-fourth ICS-TL&L. Chiang Mai University, Oct. 10–11, 1991.
- Ferlus, Michel. 1996. Langues et peuples Viet-Muong. *Mon-Khmer Studies* 26: 7–28.
- Ferlus, Michel. 1997. Problèmes de la formation du système vocalique du vietnamien. *Cahiers de Linguistique, Asie Orientale* 26.1: 37–51.
- Ferlus, Michel. 1998. Les systèmes de tons dans les langues viet-muong. *Diachronica* 15.1: 1–27.
- Ferlus, Michel. 2007. *Lexique de racines Proto Viet-Muong (Proto Vietic Lexicon)*. Unpublished Ms. Available online at: <http://sealang.net/monkhmer/database>.
- Ferlus, Michel. 2014. *Proto Viet-Muong (Proto-Vietic)*. Unpublished Ms. Available online at: <https://halshs.archives-ouvertes.fr/halshs-02490370>
- Ferlus, Michael. 2017. *Phong Toum comparatif*. Unpublished Ms.
- Gage, William W. 1985. Vietnamese in Mon-Khmer Perspective. In: Suriya Ratanakul et al. (eds.). *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt*: 493–524. Institute of Language and Culture for Rural Development, Mahidol University.
- Greenhill Simon J., Paul Heggarty, Russell D. Gray. 2020. Bayesian Phylolinguistics. In: R. D. Janda, B. D. Joseph, B. S. Vance (eds.). *The Handbook of Historical Linguistics, Volume II*: 226–253. Wiley-Blackwell: New Jersey.
- Guignard, Théodore. 1911. Note sur une peuplade des montagnes du Quảng-Binh: les Tàc-cúi. *Bulletin de l'École française d'Extrême-Orient* 11: 201–205.
- Haudricourt, André-Georges. 1953. La Place Du Viêtnameien Dans Les Langues Austroasiatiques. *Bulletin de la Société de Linguistique de Paris* 49.1: 122–8.
- Haudricourt, André-Georges. 1954a. De L'Origine Des Tones En Viêtnameien. *Journal Asiatique* 242: 69–82.
- Haudricourt, André-Georges. 1954b. Comment Reconstruire Le Chinois Archaïque. *Word: Journal of the International Linguistic Association* 10: 351–364.
- Hayes, La Vaughn H. 1982. The Mutation of *R in Pre-Thavung. *Mon-Khmer Studies* 11: 83–100.
- Hayes, La Vaughn H. 1992. Vietic and Viêt-Muong: a new subgrouping in Mon-Khmer. *Mon-Khmer Studies* 21: 211–228.
- Huffman, Franklin E. 1978. *On the centrality of Katuic-Bahmaric to Austroasiatic*. Paper presented at the 2nd International Conference on Austroasiatic, Mysore (India), December 18–21, 1978. Available online at: <https://sites.google.com/view/paulsidwell/the-sical-papers>.
- Jenny, Mathias, Paul Sidwell (eds.). 2014. *The handbook of Austroasiatic languages (2 vols.)*. Leiden / Boston: Brill.
- Kasuga, Atsushi. 2008. *Arem Vocabulary*. Ms., organized in order Vietnamese-Arem-English.
- Kim, Nam. 2015. *The Origins of Ancient Vietnam*. (Oxford Studies in the Archaeology of Ancient States.) Oxford University Press.

- Lipson, Mark et al. 2018. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* 361.6397 (2018): 92–95.
- Logan, James Richardson. 1856. *Ethnology of the Indo-Pacific Islands*. Part II, Chap. VI, App. A: Comparative Vocabulary of the Numerals of the Mon-Anam Formation. App. B: Comparative Vocabulary of Miscellaneous Words of the Mon-Anam Formation. *Journal of the Indian Archipelago* 1.
- Mạc, Đường. 1964. *Các dân tộc miền núi Bắc Trung Bộ*. Hà Nội: Khoa Học Xã Hội.
- Malglaive, Joseph de. 1902. Vocabulaire Hang-Tchek, Khas Xos, Harème (by M. Rivière). In: J. de Malglaive. *Voyages au centre de l'Annam et du Laos et dans les régions sauvages de l'est de l'Indo-Chine*: 285–290. Paris: E. Leroux.
- Maspéro, Henri. 1912. Etude sur la phonétique historique de la langue annamite: les initiales. *Bulletin de l'Ecole Française d'Extrême Orient* 12: 1–127.
- Matsumura Hirofumi et al. 2008. Morphometric affinity of the late Neolithic human remains from Man Bac, Ninh Binh Province, Vietnam: Key skeletons with which to debate the ‘two layer’ hypothesis”. *Anthropological Science* 116.2 (2008): 135–148.
- Nguyễn, Hữu Hoàn. 1999. Về sự phân định các ngôn ngữ của nhóm Việt-Mường [On the distribution of the Viet-Muong languages]. *Ngôn Ngữ* 5 (1999): 35–42.
- Nguyễn, Hữu Hoàn. n.d. *Wordlists of about 4,000 words in two Cuoi lects (Thái Hòa & Tân Hợp)*. Unpublished Ms.
- Nguyễn, Hữu Hoàn, Văn Lợi Nguyễn. 2019. Tones in the Cuoi Language of Tan Ki District in Nghe An Province, Vietnam. *Journal of the Southeast Asian Linguistics Society* 12.1: lvii–lxvi.
- Nguyễn, Phú Phong. 1997. *Le Parler Nguồn: Langue d'une Minorité ethnique des hautes vallées du Sông Gianh Quảng Bình, Việt Nam*. *Cahiers d'Etudes Vietnamiennes*. Paris: Université Paris 7 – Denis Diderot.
- Nguyễn, Tài Cẩn. 1995. *Giáo trình lịch sử ngữ âm tiếng Việt (sơ thảo) [Textbook of Vietnamese historical phonology]*. Hà Nội: Nhà Xuất Bản Giáo Dục.
- Nguyen, Tuong Lai. 1992. Poong language – The first contact of languages between Viet and Thai. In: *Pan-Asiatic linguistics: proceedings of the Third International Symposium on Language and Linguistics, Bangkok, Thailand*: 98–107. Bangkok: Chulalongkorn University.
- Nguyễn, Văn Lợi. 1993. *Tiếng Rục [The Rục language]*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Nguyễn, Văn Lợi, Hữu Hoàn Nguyễn. 2001. *Field data of wordlists for Malang Pakatan*. Unpublished Ms.
- Nguyễn, Văn Lợi, Hữu Hoàn Nguyễn. 2019. *Field data of wordlists for Kri Phoong*. Unpublished Ms.
- Nguyễn, Văn Khang, Bùi Chi, Hoàng Văn Hành. 2002. *Từ điển Mường-Việt [A Mường-Vietnamese dictionary]*. Hà Nội: Nhà Xuất Bản Văn Hoá Dân Tộc.
- Nguyễn, Văn Tài. 2005. *Ngữ âm tiếng Mường qua các phương ngôn [The phonetics of the Mường language across its various dialects]*. Hanoi: Nhà Xuất bản Từ điển Bách khoa.
- Peiros, Ilia. 2004. *Genetičeskaja klassifikacija avstroaziatskix jazykov*. Doctoral dissertation. Moskva: RSUH.
- Penny, Dan, Cameron Zachreson, Roland Fletcher, David Lau, Joseph T. Lizier, Nicholas Fischer, Damian Evans, Christophe Pottier, Mikhail Prokopenko. 2018. The demise of Angkor: Systemic vulnerability of urban infrastructure to climatic variation. *Science Advances* 4(10).
- Phạm, Đức Dương. 1975. Về mối quan hệ thân thuộc giữa các ngôn ngữ thuộc nhóm Việt-Mường miền tây tỉnh Quảng Bình [On the close relationship between languages in the Viet-Muong group in Western Quang Binh province]. In: Bê Việt Dăng (ed.). *Về Vấn Đề Xác Định Thành Phần Các Dân Tộc Thiểu Số ở Miền Bắc Việt Nam [On the problem of defining the social position of the minority groups in northern Vietnam]*: 500–517. Hanoi: Nhà xuất bản Khoa học Xã hội.
- Phạm, Đức Dương. 1979. Về mối quan hệ nguồn gốc của các ngôn ngữ nhóm Việt Mường [On the original relationship among the Viet-Muong languages]. *Ngôn Ngữ* 1979.1: 46–58.
- Phan, John D. 2012. Mường is not a subgroup: Phonological evidence for a paraphyletic taxon in the Viet-Muong sub-family. *Mon-Khmer Studies* 40: 1–18.
- Pittayaporn, Pittayawat. 2009. *The Phonology of Proto-Tai*. PhD thesis. Cornell University.
- Premrirat, Suwilai. 2000. *So (Thavung) Preliminary Dictionary*. Salaya / Melbourne: Institute of Language and Culture for Rural Development, Mahidol University – University of Melbourne.
- Przyluski, Jean. 1924. Les langues Austroasiatiques. In: Antoine Meillet, Marcel Cohen (eds.). *Les Langues du Monde (Collection linguistique publiée par la société de linguistique de Paris)*, 16: 335–403. Paris: Librairie Ancienne Edouard Champion.
- Samarina, Irina. n.d. *Spreadsheet of comparative Vietic data*. Unpublished Ms. (early version of the lexicon reproduced on pp. 259–263 in Babaev & Samarina 2018).

- Schmidt, Wilhelm. 1903. The Sakai and Semang languages in the Malay Peninsula and their relation to the Mon-Khmer languages. *Journal of the Straits Branch of the Royal Asiatic Society* 39: 38–45.
- Schmidt, Wilhelm. 1904. Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayerischen Akademie der Wissenschaft* 1.22.3: 677–810.
- Schmidt, Wilhelm. 1905. Grundzüge einer Lautlehre der Mon-Khmer-Sprachen. *Denkschrift der Akademie der Wissenschaften, Wien, Philologisch-Historische Klasse* 51: 1–233.
- Schmidt, Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie, Braunschweig*, 5: 59–109.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra: Pacific Linguistics.
- Sidwell, Paul. 2009. *Classifying the Austroasiatic languages: history and state of the art*. Munich: Lincom Europa.
- Sidwell, Paul. 2014. Austroasiatic classification. In: Matthias Jenny, Paul Sidwell (eds.). *The Handbook of Austroasiatic Languages*. Vol. 1: 144–220. Leiden / Boston: Brill.
- Sokolovskaja, N. K. 1978. Materialy k sravnitel'no-ètimologicheskomu slovar'u vjetmyongskix jazykov. In: V. M. Alpatov (ed.). *Issledovanija po fonologii i grammatike vostochnyx jazykov*: 126–180. Moskva: Nauka.
- Srisakorn, Preedaporn. 2008. *So (Thavung) grammar. Doctoral dissertation*. Thailand: Mahidol University.
- Thomas, David D., Robert K. Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25: 398–418.
- Thompson, Laurence C. 1976. Proto-Viet-Muong Phonology. In: Philip N. Jenner, Laurence C. Thompson, Stanley Starosta (eds.). *Austroasiatic Studies*, Vol. 2: 1113–1204. Honolulu: University of Hawaii Press.
- Trần, Trí Dõi. 2011. *Một vài vấn đề nghiên cứu so sánh - lịch sử nhóm ngôn ngữ Việt - Mường [A historical-comparative study of Viet-Muong group]*. Hà Nội: Nhà xuất bản Đại Học Quốc Gia Hà Nội.
- Trần, Trí Dõi. 2018. Một tham chiếu về nguồn gốc của tiếng Việt [A reference regarding the origin of Vietnamese]. In: Đinh Văn Đức (ed.). *Tiếng Việt Lịch Sử: Một Tham Chiếu Hồi Quan [Vietnamese Language History: A Reference]*: 10–86. Hanoi: Nhà Xuất Bản Văn Học.
- Vương, Hoàng Tuyên. 1963. *Các dân tộc nguồn gốc Nam Á ở miền Bắc Việt Nam [Ethnic groups of Austroasiatic origin in northern Vietnam]*. Hanoi: Nhà Xuất Bản Giáo Dục.

Пол Сидвелл, Марк Алвес. Вьетские языки: филогенетический анализ

В работе представлена новая внутренняя классификация вьетских языков, включающая все общепризнанные подгруппы и языки, для которых доступны языковые данные, необходимые для сравнения. В ходе анализа согласованы результаты двух различных методик: (1) вычислительная филогенетика, основывающаяся на 116-словном сравнительном списке и (2) сравнительный анализ диахронических изменений в финалях слоговых морфем. Анализ позволяет выделить пять основных подгрупп: тхавынг-малиенг (ТМ), тьыт-арем, поонг-гум, куой-тхо и вьет-мыонгскую (ВМ). Хотя идентификация этих подгрупп сама по себе не оригинальна, ряд особенностей ветвления дерева выявлен впервые. Так, установлено, что вьетское дерево имеет бинарную структуру, разделяясь на ТМ и все остальные языки, которые можно в совокупности назвать *восточновьетской* кладой. Внутри этой восточной кланды языки подразделяются на северные и южные, из которых северная группа (ВМ) оказывается более инновативной, а южная – более консервативной. В прошлом исследователи склонялись к тому, чтобы объединять ТМ и тьыт языки на основании общей для них архаичности структуры словоформы, не предлагая при этом каких-либо общих фонетических инноваций. Наши результаты показывают, что перестройка структуры слога и тоногенез (отличительные черты вьетских языков) в значительной степени развивались в разных подгруппах независимо друг от друга, несмотря на многочисленные сходства в фонологических изменениях, которые способствовали появлению ареальной конвергенции. В работе также вкратце обсуждается вопрос о возможной прародине вьетской группы; в пользу ее локализации на севере свидетельствуют как старые лексические заимствования, так и археологические данные.

Ключевые слова: вьетские языки; австроазиатские языки; классификация языков; вычислительная филогенетика.

О возможном сохранении реликтов ведийского глагольного ударения в говоре деревни Наггар языка куллуи

В работе рассматривается проблема рефлексов общесреднеиндийского **a* в ударной позиции в языке куллуи на материале глаголов. Ударение в куллуи соответствует среднеиндийской системе ударения. Общесреднеиндийский ударный гласный **a* в куллуи в основном дал *ɔ*, однако в ряде случаев ему соответствует *ε*. Этимологический анализ глаголов куллуи с *ɔ* и *ε* в корне показывает, что *ɔ* соответствует ударному, а *ε* – безударному *a* в ведийском языке. Таким образом, можно сделать предположение о сохранении следов ведийского ударения в глаголах куллуи.

Ключевые слова: язык куллуи; ударение; индоарийские языки; ведийский язык; глагольная акцентуация.

Введение

Куллуи – индоарийский язык, относящийся к группе идиомов химачали (в классификации индоарийских языков Дж. А. Грирсона «западные пахари»), распространённых в Западных Гималаях в индийском штате Химачал-Прадеш. Используемые в статье данные были получены автором и его коллегами в ходе экспедиций в округ Куллу штата Химачал-Прадеш в 2014–2018 гг.¹ Автором данной статьи проводится работа по этимологизации лексики куллуи (Крылова, Ренковская 2020) и систематизации исторической фонетики этого идиома. В общих чертах историческая фонология химачали изложена в книге Hendriksen 1986, там же перечислены и основные изоглоссы, разделяющие эти идиомы, однако работа Хендриксена базируется главным образом на идиомах котгархи и кочи, а при рассмотрении изоглосс Хендриксен включает в химачали также идиомы киунтхали, бахати, джаунсары, сирмаури, куллуи и мандеали. В процессе работы над этимологизацией лексики куллуи неизбежно появляются и новые данные.

Проблема рефлексов общеиндоарийского **a* в ударной позиции в языке куллуи

Вокализм куллуи, подобно вокализму русского языка, наиболее удобно описывать с точки зрения позиционной фонологии. Противопоставления гласных фонем в сильной, средней и слабой позициях имеют разное количество элементов (от семи в сильной / ударной позиции до отсутствия противопоставлений в сверхслабой). Выделение словесного ударения в куллуи на акустическом уровне, как и во многих индоарийских языках, сопряжено с большими сложностями. Так, ответ на вопрос о наличии словесного ударения хинди не является очевидным (Зограф, Оранская 2011: 58; Cardona, Jain 2003: 261).

¹ Статья написана при поддержке РФФИ, проект 19-012-00355 «Исследование языка куллуи на основе корпуса устных текстов»

С. К. Чаттерджи (1977: 133–134) полагал, что в бенгальском словесное ударение полностью преодолевается фразовым. Аналогично, Б. Тивари (Tiwāgī 1966: 279) считал, что в хинди существует только фразовое ударение. В работе Р. Келкара (Kelkar 1968: 26) представлено положение о том, что словесное ударение в хинди существует только как часть фразового. Эта же идея позже была развита на основе акустического экспериментального анализа в работе Румянцева 1985. Однако мы попробуем описать просодические свойства словоформ в куллуи по результатам работы в системе Praat и последующих подсчётов в рабочей среде Lingvodoc.

Ударение в куллуи, по-видимому, имеет прежде всего качественный характер (то есть проявляется качественной редукцией безударных гласных), а длительность и интенсивность могут считаться его признаками лишь отчасти. С историко-фонетической точки зрения ударение в куллуи в целом наследует среднеиндийской системе ударения, в которой ударным становился первый по счёту от предпоследнего и до четвертого от конца тяжёлый слог (Masica 1993: 183). В случае отсутствия тяжёлых слогов в слове ударным становился первый слог. В современном куллуи в словах, подвергшихся падению конечных кратких гласных, это правило претерпело изменение: таким образом, отсчёт идёт от предпоследнего слога в словах, оканчивающихся на гласный, и от последнего – в словах, оканчивающихся на согласный; при этом, если слово оканчивается на одиночный согласный, последний слог рассматривается как открытый. Как правило, в заимствованиях, появившихся в куллуи в новоиндийский период, ударный слог выделяется по тому же принципу (хотя имеется и ряд отклонений). Средней позицией в слове является открытый слог конца слова, а слабой – все прочие позиции, среди которых можно выделить слабые и сверхслабые.

В сильной позиции имеется противопоставление следующих гласных: *i, e, ε, a, ə, o, u*. Как правило, такая позиция в слове одна, за рядом исключений. Таких позиций может быть более одной в следующих случаях:

- 1) некоторые композиты, сохранившие ударение в каждой из частей (как заимствованные, так и исконные): *indrə-dʰanuf* ‘радуга’ (санскритизм, буквально ‘лук Индры’), *kukər-tsoq* ‘амарант’ (буквально ‘петушиный гребень’), *tir-kəmaŋ* ‘лук’ (персизм, буквально ‘стрела и лук’), *mεgər-metʃʰ* ‘крокодил’ (частично адаптированный санскритизм, буквально ‘морское чудовище – рыба’);
- 2) некоторые суффиксальные словообразования: а) с персидским суффиксом *-dar*: *dzəmindar* ‘землевладелец’, *kardar* ‘управляющий’, б) с суффиксом лексического происхождения *-a|a*²: *kismata|a* ‘удачливый’, *bʰupʰia|a* ‘плесневый’, *tsiŋnua|a* ‘каменщик’, *budzʰiua|a* ‘заботливый’;
- 3) некоторые неадаптированные заимствования: *impatəns* ‘важность’, *kompjutar* ‘компьютер’;

В слове может не быть ни одной сильной позиции в следующих случаях:

- 1) некоторые неадаптированные заимствования: *kəbi* ‘поэт’ (скр. *kavi*), *dərzi* ‘портной’ (перс. *darzi*);
- 2) некоторые служебные слова: *matlab* ‘то есть’ (перс. *matlab*), *nə* ‘локативный послелог’;

В средней позиции различаются следующие гласные фонемы: *i, e, a, u*. Фонемы *ə, o* также возможны в абсолютном конце слова, но мало употребительны. Следует отме-

² Суффикс *-a|a* в куллуи имеет функции, сходные с его когнатом *-vāla* в хинди и урду. Этимологически является потомком др.-инд. *pālaka* ‘хранитель, защитник’. В урду *vāla* трактуется как самостоятельное слово и записывается отдельно. Таким образом, вполне типично, что в куллуи *-a|a* представляет собой самостоятельную просодическую единицу.

тить, что, кроме позиции абсолютного конца слова, как средние рассматривается ещё ряд позиций: а) позиция предпоследнего слога в перфективных причастиях с суффиксом фразеологического происхождения *-udal-ada/-idil-ede*³ (*tsoluda* ‘пошедший’); б) позиция третьего от конца слога в уже упоминавшихся отглагольных образованиях с суффиксом *-a|a*, присоединяющимся к глагольному суффиксу *ηи* (*budz^hηua|a* ‘заботливый’); в) позиция предпоследнего слога в глаголах с суффиксом фразеологического происхождения *-i/u-*⁴ (первоначально имевшим пассивную, инволитивную и отыменную функцию, но в современном куллуи часто плеонастическим), о вероятной этимологии суффикса в случае пассивных глаголов см. Ренковская 2020: 115–116, Hendriksen 1986: 141–143, 146–149: *fuk^hiηa* ‘хотеть пить’, *dahiηa* ‘болеть’, *b^huk^hiηa* ‘голодать’ и др. Все эти позиции в недавнем прошлом языка, по-видимому, были конечными, и на просодическом уровне их следует рассматривать именно таким образом. Средняя позиция в куллуи, в силу обширного списка различных фонем, может рассматриваться как несущая дополнительное ударение.

Слабые позиции в куллуи можно разделить на просто слабые и сверхслабые. Просто слабой является неконечная безударная позиция гласного, если он находится в непосредственном соседстве с ударным гласным или отделён от него только фонемами *h, w, j* (как правило, в беглом произношении выпадающими). В такой позиции различимы фонемы *ə, i, u* (при этом *ə* замещает различимые в ударной позиции, исторически зафиксированные или реконструированные *ɔ, ɛ, a; i* замещает *i, e; u* замещает *u, o*): *mətiən* ‘экзамен’ (из араб. *imtihān*), *dzeħər* ‘яд’ (из перс. *zahr*), *luħər* ‘кузнец’, *riħaηa* ‘показывать’, *suaηa* ‘усыплять’, *piŋ|a* ‘жёлтый’, *sūħər* ‘понедельник’ и др. Сверхслабыми являются иные позиции безударных, в которых противопоставление отсутствует, и возможен только гласный *ə*: *bəfaηa* ‘сажать’ (каузатив от *befiηa* ‘сидеться’), *fəra|* ‘волосы’ (из др.-инд. *śirovāla*), *nəre|i* ‘кальян’ (из др.-инд. *nārikēla*), *kukəŋ* ‘петух’ (из др.-инд. *kukkuṭā*), *bəfəħ* ‘надежда’ (из др.-инд. *viśvāsa*), *mūg^hə|i* ‘арахис’ (из др.-инд. *mudgá + phāla*), *nəfaŋ* ‘черта, знак’ (из перс. *nišān*), *gəlab* ‘роза’ (из перс. *gulāb*), *məsənt* ‘последний день месяца’ (из скр. *māsānta*), *səmaŋ* ‘вещи’ (из перс. *sāmān*), *pinsəl* ‘карандаш’ (из англ. *pencil*) и др. Однако в случае неадаптированных заимствований качество гласного может в той или иной степени сохраняться и в сверхслабой позиции: *dubara* ‘дважды’ (из кл.-перс. *dubāra*), *tʃabuk* ‘плеть’ (из перс. *čābok*, при наличии, однако, дублетной адаптированной формы *tsabək* с тем же значением). Предположительно, утрата противопоставления *i/u/ə* в сверхслабых позициях произошла относительно недавно.

В слабых и сверхслабых позициях *ə* может также подвергаться полной редукции, при этом основным принципом этой редукции является запрет на образование кластеров из трёх и более согласных. Практически всегда редуцируется *ə*, являющийся единственным гласным, разделяющим сильную и среднюю позицию: *bol(ə)ηa* ‘говорить’, *gel(ə)ti* ‘ошибка’.

Древнеиндийский гласный **a* на конце слова в куллуи полностью редуцировался, в прочих безударных слогах дал *ə* (также в ряде случаев подвергающееся полной редук-

³ Причастия на *-udal-ada/-idil-ede* включают в себя двойную флексию и очевидно происходят от сочетания старого перфективного причастия на *-u/-a/-i/-e* с имперфективным причастием вспомогательного глагола, по-видимому, *hona* ‘быть’: **tsolu hunda* etc. Таким образом, просодически в них сохраняются две конечные позиции гласных.

⁴ При том, что связь *-i/u-* с пассивным и деноминативным др.-инд. показателем *-īya-* несомненна, он, в отличие от каузативного показателя *-a-*, не влияет на место основного ударения (ср. *k^helηa, k^heliηa* ‘играть’, но *befiηa* ‘сидеться’ *bəfaηa* ‘сажать’). По-видимому, это может объясняться контаминацией старого глагольного показателя со старым деепричастным показателем в сочетаниях типа **keri hona*, послуживших источником употребляемого в перфективных формах алломорфа *u* суффикса *i/u*.

ции), а в ударных слогах дал ə. Рефлекс ə характерен и для ряда других диалектов химачали (Hendriksen 1986: 192). Однако в куллуи имеется и значительный пласт лексики, в которой древнеиндийскому *a в сильной позиции соответствует ε – гласный, по формантному составу близкий к безударному ə, но значительно отличающийся от него по длительности. Несомненно, с синхронной точки зрения ε и ə противопоставлены на фонологическом уровне, среди наших данных имеются минимальные пары: *betsu* ‘спасся’ – *bətsu* ‘телёнок’, *kət* ‘работа’ – *kət* ‘мало’, *bəŋa* (косв.форма от *bəŋ* ‘лес’) – *bεŋa* (герундий от *bεŋ-* ‘становиться’), *ŋi* ‘река’ – *ŋεi* ‘нет’, *pər* ‘в прошлом году’ – *pεr* ‘но’. Однако с точки зрения исторической фонологии разграничение этих фонем представляет собой научную проблему, так как в словах индоарийского происхождения они соответствуют одной и той же древнеиндийской фонеме, и очевидного распределения найти не удаётся.

В словаре куллуи, составляемом в рамках проекта «Исследование языка куллуи на основе корпуса устных текстов», включающем на данный момент около 3000 словарных статей, соотношение встречаемости фонем почти равное: ε встречается 239 раз, ə – 276. Однако в случае выделения пласта лексики неиндоарийского происхождения (главным образом, в арабизмах, персизмах и англицизмах) ε встречается в 51 слове, а ə – в двух (*bəkt* ‘время’ < перс.-араб. *vaqt* (Steingass 1892: 1475) и *bədzə* ‘повозка’ < перс.-араб. *wazā'i* ‘обоз’ (Steingass 1892: 1472)). Напрашивается предположение, что ε – фонема, соответствующая a в заимствованиях в ударной позиции, и лексика индоарийского происхождения, содержащая фонему ε, также является заимствованной (прежде всего из хинди, санскрита, панджаби и мандеали). В таких малочисленных и не имеющих высокого статуса идиомах, как куллуи, количество заимствований вполне может сравниться по численности с исконной лексикой.

При рассмотрении ə в заимствованиях оказывается, что главным образом эта фонема представлена в словах, с точки зрения исторической фонетики соответствующих другим «окающим» идиомам химачали, таким как котгархи или кочи. Важным исключением, помимо уже упомянутых *bəkt* ‘время’ и *bədzə* ‘повозка’, является санскритизм *pəwitr* ‘ритуально чистый’ < *pəwitra* ‘средство ритуального очищения’. Во всех этих словах ə предшествует губной согласный, что может объяснять характер гласного, хотя такая передача a и не является регулярной в заимствованиях. Ещё одним аргументом в пользу заимствованного происхождения ε является результат фонетической разметки лексики куллуи и последующего анализа в среде Lingvodoc (lingvodoc.ispras.ru). С точки зрения формантных показателей ε находится максимально близко к ə (соответствующему др.-инд. *a в безударной позиции), и отличается от него большей длительностью. Более того, в беглом произношении, особенно в служебных словах, ε становится и по длительности сопоставимым с ə. Наша интерпретация этих данных предполагает, что ε – результат вторичного ударения, появившегося после перехода a > ə в ударных слогах куллуи. Поскольку это ударение появилось относительно недавно, во многих случаях оно нестабильно. К данному моменту в словаре этимологизировано 167 слов с ε и 167 слов с ə. Большинство этимологизированных слов с ε (105 из 167)⁵, действительно, составляют явные заимствования, прежде всего, из арабского, персидского, английского и хинди, в то время как большая часть этимологизированных на данный момент слов с ə (121 из 167) полностью соответствуют исторической фонологии куллуи. Однако с целью более детальной интерпретации необходимо исследовать отдельные группы лексики. Одной из таких групп в куллуи являются глаголы.

⁵ Здесь и далее мы обсуждаем лишь ту часть лексики куллуи, которая подтвердилась полевой работой на материале деревни Наггар, но не лексику, полученную исключительно по данным других источников (Ranganatha 1980, Thakur 1975).

Рефлексы древнеиндийского *а в ударной позиции в глагольных основах

В отличие от имён, глаголы заимствуются достаточно редко, а в индоарийских языках заимствованные глаголы обычно оформляются как составные с использованием заимствованного имени и исконного вспомогательного глагола (куллуи *bend ker-* ‘закрывать’ из перс. *band* ‘закрой’). При этом соотношение в словаре куллуи простых глаголов с *ε* и *ɔ* составляет приблизительно 1 к 2.7 (18 к 48), что больше согласуется с гипотезой о заимствованиях, чем соотношение этих фонем в лексике куллуи в целом, однако всё же достаточно много. Те из глаголов с *ε*, которые не являются очевидными заимствованиями, в целом соответствуют не только исторической фонетике куллуи, но и исторической фонетике мандеали, территориально близкого идиома химачали, в котором, однако, не произошёл переход *a > ɔ* под ударением. Поэтому заранее оговоримся, что более внимательное рассмотрение этимологии этих глаголов может послужить лишь альтернативной гипотезе, тогда как самым простым будет предположение о заимствованиях из мандеали.

Глаголы с *ε*: *bedal-* ‘менять, меняться’, *bεŋ-* ‘становиться’, *bets-* ‘спасаться’, *bheg-* ‘бежать, уходить’, *des-* ‘рассказывать’, *ker-* ‘делать’, *k^herts-* ‘тратить’, *mεŋ-* ‘полагать’, *nei* ‘отрицательная глагольная связка’⁶, *p^hεŋt-* ‘трясти, взбивать’, *rek^h-* ‘держат’, *sek-* ‘мочь’, *semadz-* ‘понимать’, *ferm-* ‘стыдиться’, *set-* ‘бросать’, *t^hek-* ‘уоставать’, *tsek-* ‘поднимать’, *tsemak-* ‘сверкать’.

Глаголы с *ɔ*: *b^hɔn-* ‘ломать’, *b^hɔr-* ‘наполнять’, *b^hɔrak-* ‘болтать’, *bɔh-* ‘нести’, *bɔk-* ‘жечь’, *bɔn^h-* ‘связывать’, *bɔŋd-* ‘делить’, *dɔl-* ‘молоть, обмолачивать’, *dz^hɔl-* ‘падать’, *dzɔl-* ‘завидовать’, *dzɔm-* ‘рождаться’, *dzɔŋ-* ‘рождать’, *dzɔr-* ‘быть в жару, температурить’, *dzɔt-* ‘ссориться’, *q^hɔk-* ‘ловить’, *q^hɔl-* ‘течь’, *qɔr-* ‘бояться’, *hɔl-* ‘пахать’, *hɔs-* ‘смеяться’, *k^hɔl-* ‘промыть шерсть’, *kɔl^h-* ‘вытаскивать’, *kɔt-* ‘прясть’, *kɔm^h-* ‘дрожать в трансе’, *lɔdz-* ‘стыдиться’, *mɔl-* ‘тереть’, *mɔr-* ‘умирать’, *mɔg-* ‘просить’, *nɔf-* ‘уходить’, *nɔts-* ‘танцевать’, *ɔl-* ‘капать’, *ɔt^hi* ‘глагольная связка’, *ɔd-* ‘гулять’, *p^hɔl-* ‘впервые плодоносить’, *p^hɔŋ-* ‘трепать шерсть’, *pɔ-* ‘приходиться’, *pɔd-* ‘выпускать газы’, *pɔl^h-* ‘учиться’, *pɔt-* ‘закапывать’, *rɔdz-* ‘быть довольным, сытым’, *rɔh-* ‘оставаться, жить’, *rɔŋg-* ‘красить’, *ɔɔl-* ‘гнить’, *tɔr-* ‘плавать’, *tɔt^h-* ‘рубить’, *tsɔl-* ‘ходить’, *tsɔr-* ‘пасться’, *tɔk^h-* ‘гнить’, *tɔŋ-* ‘стонать’⁷.

Перед тем, как проводить этимологический анализ этих глаголов, следует сделать небольшое отступление о ведийском ударении. До того, как в среднеиндийский период в индоарийских языках установилась упомянутая в разделе 1.0 система ударения, в древнеиндийский период в ведийском языке существовала другая система. Достаточно подробно эта система изложена в работе Зализняк 1996; здесь я приведу только самые необходимые сведения. Ударения отображались при записи ведийских текстов, но не текстов более позднего периода. Кроме того, во многих случаях ведийское ударение можно предсказать из известной нам грамматической информации. Наиболее общий принцип ударения в ведийской словоформе требует постановки ударения на первый по счёту

⁶ Морфологически и исторически данная лексема не является глагольной формой, однако мы включаем её в данный список, считая, что особое поведение ударения на глаголах по сравнению с другой лексикой может быть связано с их особой синтаксической ролью и фразовым ударением.

⁷ В данном списке приводятся только корни глаголов; в ряде случаев глагольные формы, полученные от информантов, включали в себя упоминавшийся выше суффикс *i/u*. Некоторые из этих форм были вариативны. Суффикс, изначально связанный с пассивным, инволитивным и деноминативным употреблением, в современном куллуи у некоторых информантов может носить откровенно плеонастический характер. В силу его явной вторичности и фразеологического происхождения, он никак не влияет на этимологический анализ. Однако в целях большей точности передачи полевого материала, приведём список корней обеих групп, зафиксированных с ним: *bedal-* ‘менять, меняться’, *ferm-* ‘стыдиться’, *dzɔr-* ‘быть в жару, температурить’, *dzɔt-* ‘ссориться’, *lɔdz-* ‘стыдиться’, *tɔk^h-* ‘гнить’.

слог, для которого морфонологически требуется полная или продлённая ступень слогаобразующего элемента, а в случае отсутствия такового – на последний слог.

Личные формы глагола без преверба в ведийском языке были безударны, за исключением случаев: (а) личной формы в начале фразы или стиха, или непосредственно после вокатива в начале фразы или стиха и (б) личной формы, стоящей в придаточном предложении или непосредственно после другой личной формы.

Таким образом, даже в случае, когда личная глагольная форма зафиксирована в ведийских текстах, ударение не всегда нам известно. Однако мы знаем, что презенс 1-го (основа: корень в ступени *giṇa* + суффикс *-a-*) и 4-го (корень в нулевой ступени + суффикс *-ya-*) класса имеют ударение на корне, презенс 6-го класса (корень в нулевой ступени + суффикс *-a-*), пассивный презенс (корень в нулевой ступени + суффикс *-ya-*), презенс 10-го класса и каузативный презенс (корень в ступени *giṇa* + суффикс *-aua-*) – на основообразующем элементе. В презенсе 5 класса (корень в нулевой ступени + *-ni-* или *-no-*) или 8 класса (корень в нулевой ступени + *-i-* или в ступени *giṇa* + *-o-*) ударение падает, в зависимости от формы, на основообразующий элемент или на окончание, но не на корень. При этом у существенной части корней (т. н. неполноизменяемых) ударение является единственным способом отличить презенс 1-го класса от презенса 6-го класса. Поэтому в случаях, когда ударные формы не зафиксированы в текстах, различие между 1-м (наиболее многочисленным) и 6-м классом, принятое в словарях и грамматиках, является в значительной мере условным. Также чрезвычайно близки презенс 4-го класса при присоединении медиальных окончаний и пассивный презенс, которые, кроме ударения, можно отличить друг от друга семантически, но это различие довольно расплывчато. Классы презенса различных глагольных корней определяются, во-первых, по зафиксированным в ведийских и санскритских в широком смысле текстах формам, а во-вторых, по предписаниям «Дхатупатхи» – санскритского списка глагольных корней, приложения к «Аштадхьяи» Панини. Кроме того, в книге Kulikov 2012 проведены реконструкции ударения ряда презенсов. С учётом этой информации, рассмотрим гипотезу о возможной связи двух рефлексов *a* в куллуи с ударением ведийского презенса.

Этимологический анализ глаголов с ϵ

Глаголы с ϵ с точки зрения этимологии можно разделить на пять групп⁸.

1. Основы, зафиксированные в санскрите, и имевшие в ведийском безударное *a*:

Куллуи	Этимология
<i>bep-</i> ‘становиться’	др.-инд. <i>vanōti</i> ‘превозмогает’
<i>bets-</i> ‘спасаться’	др.-инд. <i>vacyātē</i> ‘бегает, движется волнообразно’
<i>ker-</i> ‘делать’	др.-инд. <i>karōti</i> ‘делает’
<i>teṇ-</i> ‘полагать’	др.-инд. <i>manutē</i> ‘думает’
<i>neī</i> ‘отрицательная глагольная связка’	др.-инд. <i>nahī</i> , отрицательная частица
<i>semādz-</i> ‘понимать’	др.-инд. <i>sambudhyatē</i>

Таблица 1. Основы, зафиксированные в санскрите и имевшие в ведийском безударное *a*.

⁸ Этимологии приведены, как правило, согласно Turner 1931, Turner 1962–1985, переводы – согласно Бархударов 1972, Singh 1895, Steingass 1892, Monier-Williams 1899.

Формы презенса *karóti*, *vanóti*, *vasyátē*, *manuté* отличаются тем, что во всех них безударный корневой гласный *a*. Конечно, следует учитывать то, что в ведийском языке глагольный корень зачастую мог образовывать презенсы нескольких классов, таким образом, возможны и параллельные формы презенса с ударным корневым гласным, поэтому проверим возможные варианты по конкордансу санскритских глагольных корней (Whitney 1885):

vanóti ‘превозмогает’, 8 кл. (Веды, Брахманы, Сутры). Также имеются формы презенса 1 (*vánati*, Веды) и 6 (*vanāti*, Веды, Брахманы) класса. Таким образом, формы 8 и 6 класса с безударным корнем распространены обширнее;

vasyátē ‘бегает, движется волнообразно’, акцентуация пассива. В случае происхождения от активного презенса *vāñcati*, в куллуи ожидалось бы ***bāñdz-/beñdz-* с конечным сочетанием носового и звонкой аффрикаты, поэтому мы предполагаем связь формы куллуи с презенсом *vasyátē*;

karóti ‘делает’, 8 кл. (Ригведа 3 раза, Атхарваведа и далее), также имеются формы презенса 5 (*kṛṇóti* Веды, Брахманы, Сутры), 2 (*kárṣi*, Атхарваведа 1 раз) и 1 (*karanti*, Ригведа) классов. Наиболее обширно распространены формы 8 класса;

manuté ‘думает’, 8 кл. Веды и позже, также встречаются формы 4 (*mányate* Веды и позже, *mányati* Упанишады и позже) и 3 (*mamandhi* Ригведа) классов. Однако в случае происхождения от основ *manú-* и *maman-* в куллуи ожидалось бы соответственно конечный согласный *n* вместо *ṅ* (*n* происходит из среднеиндийской геминаты *nn* на месте древнеиндийского кластера, *ṅ* – из интервокального *n*) и следы удвоения корня.

Таким образом, во всех четырёх случаях наиболее вероятно связь формы куллуи с ведийскими презенсами, в которых зафиксирован безударный корень.

Глагол *sambudhyatē* с превербом не зафиксирован в ведийском языке. Однако нам известна форма *búdhyatē* (4 кл.) без преверба, а по ведийским правилам ударения, в случае ударности глагольной формы преверб оставался безударным: **sam-búdhyatē*. Поскольку в случае прочих глаголов мы рассматривали ситуацию, когда глагольная форма несёт на себе ударение, в данном случае мы также считаем гласный преверба безударным.

Последний случай (*nei < nahí*) не является глаголом в морфологическом смысле, но может примыкать к этому списку, так как выполняет в куллуи роль отрицательной глагольной связки, а также отрицательной частицы при глаголе. В случаях, когда он оказывается в непредикативной роли, он может не нести на себе фразового ударения и произноситься как *nai* или *ni*. Как можно заметить, в ведийском когнате *a* также безударно. Впрочем, пока не исследовано поведение частиц и других служебных элементов в куллуи в отношении рефлексов *a*, неясно, можно ли связывать эту форму с ведийским ударением. Однако в целом данные примеры иллюстрируют соответствие ε в куллуи безударному *a* в ведийском. При этом среди глаголов нет случаев соответствия ε в куллуи ударному *a* в ведийском, которые бы в остальном соответствовали исторической фонетике.

2. Заимствования неиндоарийского происхождения:

Куллуи	Этимология
<i>bedal-</i> ‘менять, меняться’	перс. араб. <i>badal</i> ‘замена, обмен’
<i>k^herts-</i> ‘тратить’	перс. араб. <i>kharj</i> ‘трата’
<i>ferm-</i> ‘стыдиться’	перс. <i>sharm</i> ‘стыд’

Таблица 2. Заимствования неиндоарийского происхождения.

3. Заимствования индоарийского происхождения:

Куллуи	Этимология
<i>dēs-</i> ‘рассказывать’	пандж. <i>das-</i> ‘рассказывать’ < др.-инд. <i>darśáyati</i> ‘показывает’ (возможно, через мандеали, территориально близкий к куллуи идиом, больше подвергшийся влиянию панджаби, в котором также представлен этот глагол (Ranganatha 1980: 125), хотя он и не соответствует исторической фонетике мандеали). Историческая фонетика куллуи предписывает $\acute{s} > f$, но не $\acute{s} > s$
<i>p^hεηt-</i> ‘трясти, взбивать’	др.-инд. <i>*phāñt-</i> ‘взбивать’ или <i>*phañt-</i> ‘резкое движение’. Обе праформы не соответствуют в полной мере исторической фонетике куллуи, где $\bar{a} > a$, $\eta t > \eta d$, $t t > t$. Вероятно заимствование из другого индоарийского идиома.
<i>rεk^h-</i> ‘держат’	хинди <i>rakh-</i> или пандж. <i>rakkh-</i> ‘держат’ < др.-инд. <i>rākṣati</i> , Историческая фонетика куллуи предписывает переход $kṣ > tʃ^h$, но не $kṣ > k^h$
<i>sεk-</i> ‘мочь’	хинди <i>sak-</i> или пандж. <i>sakk-</i> ‘мочь’ < др.-инд. <i>śakñóti</i> ‘мочь’. Историческая фонетика куллуи предписывает $\acute{s} > f$, но не $\acute{s} > s$

Таблица 3. Заимствования индоарийского происхождения.

4. Вторичные отричастные образования среднеиндийского периода:

Куллуи	Этимология
<i>bheg-</i> ‘бежать, уходить’	др.-инд. <i>bhagna</i> ‘сломанный, поражённый’
<i>fεt-</i> ‘бросать’	др.-инд. <i>sṛṣṭá</i> ‘брошенный’ (переход $s > f$, предположительно, в результате дистантной ассимиляции перед \acute{s})

Таблица 4. Вторичные отричастные образования среднеиндийского периода.

5. Вторичные образования, оканчивающиеся на *k* (возможно, со среднеиндийским суффиксом *-akka-* или глагольные композиты с *kṛ*):

Куллуи	Этимология
<i>t^hεk-</i> ‘уставать’	<i>*sthakk</i> ‘останавливаться’ (вероятно, связано с др.-инд. <i>sthā-</i> ‘стоять’, возможно суффиксальное расширение + композит с <i>kṛ</i>)
<i>tsek-</i> ‘поднимать’	др.-инд. <i>ucca</i> ‘высокий, возвышенный’ + <i>akka</i> (или другое расширение + композит с <i>kṛ</i>)
<i>tsemək-</i> ‘сверкать’	<i>*cammakka</i> ‘внезапное движение’ (видимо, связано с скр. <i>camat</i> ‘междометие удивления’ и композитом <i>camat-kṛ</i> ‘удивлять, удивляться’)

Таблица 5. Вторичные образования, оканчивающиеся на *k*.Этимологический анализ глаголов с *э*

Глаголы с *э* в куллуи также можно разделить на несколько этимологических групп.

1. Зафиксирована полностью соответствующая ведийская ударная форма:

Куллуи	Этимология
<i>b^hɔr-</i> ‘наполнять’	<i>bhárati</i> ‘несёт, наполняет’
<i>bɔh-</i> ‘нести’	<i>váhati</i> ‘везёт’
<i>dzo[-]</i> ‘завидовать’	<i>jvátati</i> ‘горит, пылает’
<i>dzoη-</i> ‘рождать’	<i>jánati</i> ‘рождает’

Куллуи	Этимология
<i>həs-</i> ‘смеяться’	<i>hásati</i> ‘смеётся’
<i>mɔr-</i> ‘умирать’	<i>máratē</i> ‘умирает’
<i>nɔf-</i> ‘уходить’	<i>násyati</i> ‘пропадает’
<i>nɔts-</i> ‘танцевать’	<i>nṛtyati</i> ‘танцует’
<i>ɔtʰi</i> ‘глагольная связка’	<i>ásti</i> ‘глагольная связка’
<i>pɔ-</i> ‘приходиться’	<i>pátati</i> ‘падает’
<i>tɔr-</i> ‘плавать’	<i>táratī</i> ‘пересекает’
<i>tɔtʰ-</i> ‘рубить’	<i>tákṣati</i> ‘обрабатывает’
<i>tsɔr-</i> ‘пасться’	<i>cáratī</i> ‘ходит’

Таблица 6. Глаголы куллуи, для которых зафиксирована полностью соответствующая ведийская ударная форма.

Как мы видим, все эти формы имеют накоренное ударение. Рассмотрим также другие зафиксированные презенсы от этих корней по Whitney 1885:

bhárati ‘несёт, наполняет’ (Веды и позже). Также встречается презенс 3 класса *bíbharti*, *bíbhrati* (Веды, Брахманы), *bibhárti* (Ригведа 1 раз), 2 класса *bharti* (Ригведа 2 раза). Дхатупатха предписывает 1 класс,⁹ возведение основы *bʰɔr-* к атематическому типу и тем более презенсу невозможно по историко-фонетическим соображениям;

váhati ‘везёт’ (Веды и позже). Также имеется презенс 2 класса *vakṣi* (Веды, Брахманы). Дхатупатха предписывает 1 класс, а возведение основы *bɔh-* к атематическому презенсу невозможно по историко-фонетическим соображениям;

jválati ‘горит, пылает’; формы других классов не обнаружены;

jánati ‘рождает’; также зафиксирован презенс 4 класса *jáyate*. Дхатупатха относит глагол к 3 классу, однако это, по-видимому, ошибка. По данным исторической фонетики, глагол куллуи соответствует здесь презенсу 1 класса;

hásati ‘смеётся’; формы других классов не зафиксированы. Дхатупатха также относит глагол к 1 классу;

máratē ‘умирает’; Дхатупатха относит глагол к 6 классу, однако наличие ведийских форм показывает, что в Дхатупатхе в данном случае ошибка. Возведение основы *mɔr-* к *mriyáte* невозможно по историко-фонетическим соображениям;

násyati ‘пропадает’; также 1 класс *násati*. Дхатупатха относит глагол к 4 классу, накоренное ударение в любом случае несомненно;

nṛtyati ‘танцует’ (переход *ɾ > a*, типичный в ср.-инд. период); другие презенсы не зафиксированы. Дхатупатха также предписывает 4 класс;

ásti ‘глагольная связка’; другие варианты презенса отсутствуют;

pátati ‘падает’; другие варианты презенса отсутствуют. Дхатупатха также относит глагол к 1 классу;

táratī ‘пересекает’; имеются также презенсы 6 класса *tiráti*, *turáti*, 3 класса *titartī*. Возведение основы *tɔr-* к какой-либо из этих форм невозможно по историко-фонетическим соображениям. Также встречается единичный случай 8 класса *tarute* в Ригведе, но 1 класс распространён гораздо шире;

tákṣati ‘обрабатывает’; Дхатупатха также относит глагол к 1 классу. Имеются презенсы 2 класса *tāṣṭi* и 5 класса *takṣṇuvánti*, однако 1 класс распространён наиболее широко;

cáratī ‘ходит’; Дхатупатха также относит глагол к 1 классу, формы других презенсов отсутствуют.

⁹ Здесь и далее сведения из Дхатупатхи приводятся по Westergaard 1841.

2. Глаголы зафиксированы в санскрите, но ударные формы отсутствуют.

Куллуи	Этимология
<i>bṃh^h</i> - ‘связывать’	<i>bandhati</i> ‘связывает’
<i>bṃṇḍ</i> - ‘делить’	<i>vaṇṭati</i> ‘делит’
<i>dṃl</i> - ‘молоть, обмолачивать’	<i>dalati</i> ‘ломает’
<i>dzṃr</i> - ‘быть в жару, температурить’	<i>jvarati</i> ‘пребывает в жару’
<i>dṃr</i> - ‘бояться’	<i>darati</i> ‘уважает’
<i>kṃmh</i> - ‘дрожать в трансе’	<i>kampatē</i> ‘дрожит’
<i>lṃdz</i> - ‘стыдиться’	<i>lajjate</i> ‘стыдится’
<i>mṃg</i> - ‘просить’	<i>mārgati</i> ‘ищет’
<i>pṃd</i> - ‘выпускать газы’	<i>pardatē</i> ‘выпускает газы’
<i>pṃṛ^h</i> - ‘учиться’	<i>paṛhati</i> ‘учится, читает’
<i>rṃdz</i> - ‘быть довольным, сытым’	<i>rajyate</i> ‘окрашивается, возбуждается’
<i>rṃh</i> - ‘оставаться, жить’	<i>rahatī</i> ‘отделяется, покидает’
<i>ṣṃr</i> - ‘гнить’	<i>śaṭati</i> ‘болеет’

Таблица 7. Глаголы, соответствия которых зафиксированы в санскрите, но ударные формы отсутствуют.

Некоторые дополнительные аргументы в пользу накоренного ударения этих глаголов:¹⁰

bandhati ‘связывает’; также распространён более ранний презенс 9 класса *badhnāti* (Ве-ды и позже), Дхатупатха также предписывает 9 класс, а форма *bandhati* встречается лишь в эпическом и классическом санскрите. Здесь мы вынуждены (вслед за Turner 1962–1985: 516) отойти от ранней фиксации и предписаний Дхатупатхи, так как ведийский презенс *badhnāti* должен был бы иметь своей параллелью в куллуи ***bṃd^h- / bṃd^h-*, *n* во второй части кластера не сохранилось бы (Masika 1991: 178–179). В работе Дыбо 2011: 149 презенс *bandhati*, который соответствует основе *bṃh^h*- в куллуи, также отнесён к 1 классу в соответствии с данными языка шина;

vaṇṭati- ‘делит’; Дхатупатха относит глагол к 1 и 10 классам;

dalati ‘ломает’; глагол традиционно относится к 1 классу. По-видимому, представляет собой вторичный вариант от корня *ḍr* ‘треснуть, расколоть’ 2 и 9 классов, субъюнктив аориста которого (**dārati / dālati*) был переосмыслен как презенс 1 класса;

jvarati ‘пребывает в жару’; форм с ударением не зафиксировано, глагол условно относят к 1 классу, но возможен и 6 класс. Однако, по всей вероятности, корень представляет собой фонетический вариант корня *jval-*, несомненно относящегося к 1 классу (Mayrhofer 1989: 607; Gotō 1987: 78, 153);

darati ‘уважает’; Дхатупатха предписывает 6 класс. В действительности в системе презенса глагол встречается только в традиционном пассиве (*IV в нотации Kulikov 2012) и с превербом *ā*. По-видимому, презенс *darati* представляет собой переосмысленный субъюнктив аориста (**dārati*), и в этом случае также несёт ударение на корне;

kampatē ‘дрожит’; Дхатупатха относит глагол к 1 классу;

lajjate ‘стыдится’; Дхатупатха относит глагол к 6 классу, однако форм, свидетельствующих об этом, не зафиксировано. При этом см. Kulikov 2012: 497 о связи этого глагола с *rajyate*. Принимая положение о том, что *lajja-* произошло от *rajya-* ‘окрашивается, возбуждается’ в результате фонетических преобразований в пракритах, и что *rajya-* отно-

¹⁰ Выражаю глубокую признательность Л. И. Куликову за ценные этимологические соображения.

сится к 4 классу с накоренным ударением (Kulikov 2012: 721), логично рассматривать *lajja-* как основу 1 класса с накоренным ударением;

mārgati ‘ищет’ (в ср-инд. период зафиксировано сокращение $\bar{a} > a$); Дхатупатха относит глагол к 10 классу. Поскольку ударные формы не зафиксированы, глагол можно отнести как к 1, так и к 6 классу;

pardatē ‘выпускает газы’; Дхатупатха относит глагол к 1 классу, полная ступень корня свидетельствует о том же;

paṭhati ‘учится, читает’; Дхатупатха предписывает 1 класс, ударные формы не зафиксированы. Однако исследователи (Maughofer 1989: 68; Gotō 1987: 86) считают этот глагол производным от *práthate* ‘распространяется’, что позволяет также отнести его к 1 классу;

rajyate ‘окрашивается, возбуждается’; Дхатупатха также предписывает 4 класс. Также имеется презенс *rañjati*, однако в этом случае в куллуи бы ожидался носовой согласный. В ведийском языке ударные формы не зафиксированы, однако в Kulikov 2012: 721 реконструируется корневое ударение;

rahati ‘отделяется, покидает’; Дхатупатха относит к 1 классу;

śatati ‘болеет’; Дхатупатха относит к 1 классу.

1. Фонетический облик когнатов позволяет восстановить презенс глагола, не зафиксированный в древнеиндийский период:

Куллуи	Этимология
<i>dzʰɔɽ-</i> ‘падать’	* <i>jhaṭati</i> ‘падает’
<i>dzɔɽ-</i> ‘ссориться’	* <i>jhaṭṭ</i> ‘внезапное движение’
<i>ɖʰɔɽ-</i> ‘течь’	* <i>ḍhalati</i> ‘падает’
<i>kɔɽʰ-</i> ‘выгаскивать’	* <i>kaḍḍhati</i> ‘тащит’
<i>kɔɽ-</i> ‘прясть’	* <i>kartati</i> ‘прядёт’
<i>mɔɽ-</i> ‘тереть’	* <i>malati</i> ‘разрушает’
<i>ṣḍ-</i> ‘гулять’	* <i>haṇṭ</i> ‘двигаться’
<i>pɔɽ-</i> ‘закапывать’	* <i>paṭyati</i> ‘раскалывает’
<i>tsɔɽ-</i> ‘ходить’	* <i>calyati</i> ‘идёт’

Таблица 8. Глаголы, для которых восстановлен презенс, не зафиксированный в древнеиндийский период.

Праформы **jhaṭati* ‘падает’ и **jhaṭṭ* ‘внезапное движение’ Тёрнер связывает с развитием зафиксированного в ведийском *kṣárati* ‘течёт, падает’, что позволяет нам предположить их принадлежность к 1 классу. Аналогично, для **ḍhalati* ‘падает’ Тёрнер предполагает происхождение от ведийского *dhvárati* ‘гнёт, роняет’. Кроме того, для корня **haṇṭ* ‘двигаться’ он находит неточные соответствия *áṇṭh* ‘движется’ и *hiṇḍ* ‘бродит’ в Дхатупатхе, где этим корням предписывается 1 класс. **kaḍḍhati* ‘тащит’, по-видимому, происходит из *kárṣati* ‘тащит’. Связь этих корней с ведийскими корнями 1 класса – аргумент в пользу накоренного ударения. Однако есть и случаи, когда восстановленная форма презенса не соответствует возможным зафиксированным параллелям. Так, восстановленным формам 4 класса (не пассивам, так как значение активно) **paṭyati* ‘раскалывает’ и **calyati* ‘идёт’ соответствуют зафиксированные презенсы 1 класса *paṭati* ‘открывается’ и *calati* ‘колеблется’. При этом в куллуи древнеиндийским *paṭati* и *calati*, согласно требованиям исторической фонетики к интервокальным древнеиндийским согласным, соответствовали бы незафиксированные ***pɔɽ-* и ***tsɔɽ-* вместо *pɔɽ-* и *tsɔɽ-*. Корень **mal-* ‘разрушать’, видимо, является вариантом **mɽ-* / **mɽn-* ‘ломать’, для которого зафиксированы формы 9 и 6 класса,

но только от варианта **ṛṇ-* с нулевой ступенью основы, поэтому на предположение о 1 классе **mal-* ‘разрушать’ они не влияют. Для праформы **kartati* ‘прядёт’ в ведийском языке встречается параллель 7 класса *kṛṇátti* ‘прядёт’, однако 1 класс предлагается в соответствии с данными языка шина в работе Дыбо 2011: 150. Таким образом, во всех этих формах вполне вероятно накоренное ударение.

2. Деноминативы:

Куллуи	Этимология
<i>dzṃt-</i> ‘рождаться’	<i>jánman</i> ‘рождение’
<i>hṣ-</i> ‘пахать’	<i>halá</i> ‘плуг’
<i>p^hṣ-</i> ‘впервые плодоносить’	<i>phalati</i> ‘плодоносит’ < <i>phála</i> ‘плод’
<i>rṣṅ-</i> ‘красить’	<i>raṅga</i> ‘цвет, краска’
<i>tṣk^h-</i> ‘гнить’	<i>takrá</i> ‘простокваша’

Таблица 9. Деноминативы.

Тёрнер восстанавливает деноминатив 10 класса **raṅgayati* из *raṅga* ‘краска’, однако из материала куллуи именно такая форма не следует, и мы вправе предположить деноминатив 1 класса **raṅgati* ‘красит’, аналогично зафиксированному *phalati* ‘плодоносит’ из *phála* ‘плод’ (1 класс по Дхатупатхе), и прочие деноминативы также отнести к 1 классу. Деноминатив *tṣk^h-* ‘гнить’ от *takrá* ‘простокваша’ образовался с участием характерной для куллуи и многих других индоарийских идиомов Западных Гималаев метатезы *r*, а затем регулярного перехода *tr > tṣ*.

3. Вторичное среднеиндийское отпричастное образование:

Куллуи	Этимология
<i>b^hṣṇ-</i> ‘ломать’	<i>*bhanna</i> ‘сломанный’

Таблица 10. Вторичное среднеиндийское отпричастное образование.

4. Для восстановленного древнеиндийского корня не зафиксировано никаких соответствий в древнеиндийский период, и нет возможности предположить класс презенса:

Куллуи	Этимология
<i>b^hṣṛak-</i> ‘болтать’	<i>*bhaṭ</i> ‘внезапное движение или шум’
<i>q^hṣk-</i> ‘ловить’	<i>*ḍhakk</i> ‘накрывать’

Таблица 11. Глаголы, для которых не восстанавливается класс презенса.

5. Заимствование индоарийского происхождения:

Куллуи	Этимология
<i>k^hṣ-</i> ‘промывать шерсть’	<i>k^hṣ-</i> < др.-инд. <i>kṣálati</i> ‘моет’ в каком-либо из диалектов химачали, которому форма соответствует по исторической фонетике (<i>kṣ > k^h, á > ṣ, -l -> ḷ</i>); к таким относятся (согласно Hendriksen 1986: 192) котгархи, кочи, кьонтхли, багхати, джаунсари, сирмаури

Таблица 12. Заимствование индоарийского происхождения:

6. *ɔ* происходит из других гласных в результате нерегулярного изменения не вполне понятной природы:

Куллуи	Этимология
<i>ɔ</i> - 'капать'	* <i>avalalati</i> , 'спускается' др.-инд. <i>ava</i> > <i>ɔ</i> (ср. куллуи <i>ɔɽ^h</i> < др. инд. * <i>avaḍḍhāla</i> 'наводнение')
<i>p^hɔɽ</i> - 'трепать шерсть'	<i>phāṇayati</i> 'стаскивает' др.-инд. <i>ā</i> > <i>ɔ</i> возможно, лабиализация после губного

Таблица 13. Глаголы, в которых *ɔ* является результатом нерегулярного изменения.

7. Этимология неясна: *bɔk*- 'жечь', *tʃɔɽ*- 'стонать'.

Выводы

Полученные данные, хотя и на небольшом материале, но подтверждают гипотезу, что в глагольной системе пракуллуи, в отличие от подавляющего большинства индоарийских языков, в среднеиндийский период сохранялась система ведийского ударения, и ударное **a* глагольных корней дало *ɔ* (как и в именах, где система акцентуации сменилась в среднеиндийский период), а безударное корневое **a* дало *a*, которое впоследствии, когда (относительно недавно) появилось требование обязательного ударения на глагольных формах, перешло в *ε*. Совпадение *i* и *u* в *a* в сверхслабой позиции, как и предполагалось, произошло достаточно поздно, позже этого вторичного ударения, судя по тому, что *ε* отражает только *a*, но не *i*, *u* (ср. *ʃuɽ*- 'слышать' < *śṛṇōti* 'слышит', *tsiɽ*- 'класть кладку, строить' < *cinōti* 'собирает'). Возможна и другая, упомянутая ранее интерпретация, согласно которой все перечисленные лексемы с *ε*, в остальном соответствующие исторической фонетике куллуи, являются заимствованиями из близкого идиома мандеали, в котором не произошёл переход *á* > *ɔ*, и *ε* в куллуи появилось в этих словах так же, как и в именных заимствованиях.

Литература

- Бархударов, А. С., В. М. Бескровный, Г. А. Зограф, В. П. Липеровский (под редакцией В. М. Бескровного). 1972. *Хинди-русский словарь в двух томах*. Москва: Советская энциклопедия.
- Дыбо, В. А. 2011. Древнеиндийский акцент в дардском языке шина как проблема индоевропейской акцентологии. В сб.: Е. К. Молчанова (ред.). *Лексика. Этимология. Языковые контакты. К юбилею доктора филологических наук профессора Джой Исифовны Эдельман*: 92–167. Москва: Тезаурус / Институт языкознания РАН.
- Зализняк, А. А. 1996. Грамматический очерк санскрита. В кн.: Кочергина, В. А. *Санскритско-русский словарь*. С прил. "Граммат. очерка санскрита" А. А. Зализняка: 785–895. Москва: Филология.
- Зограф, Г. А., Т. И. Оранская. 2011. Хинди язык. В кн.: Т. И. Оранская, Ю. В. Мазурова, А. А. Кибрик, Л. И. Куликов, А. Ю. Русаков (ред.). *Языки мира: Новые индоарийские языки*: 47–105. Москва: Academia / Институт языкознания РАН.
- Крылова, А. С., Е. А. Ренковская. 2020. Контактные явления в языке куллуи. В сб.: А. И. Коган (ред.). *Труды Института востоковедения РАН. Вып. 29: Проблемы общей и востоковедной лингвистики. Язык меняющийся: семантические и грамматические изменения в языках Азии и Африки*: 90–102. Москва: Институт востоковедения РАН.
- Ренковская, Е. А. 2020. *Ареально-генетические и типологические факторы формирования послеложно-надежной системы в языке кулаони. Диссертация на соискание учёной степени кандидата филологических наук*. Москва: Институт языкознания РАН.
- Румянцева, И. М. 1985. *Ритмическая структура слова в хинди. Анализ*. Москва: Наука.
- Чаттерджи, С. К. 1977. *Введение в индоарийское языкознание*. Москва: Наука.

References

- Barhudarov, A. S., V. M. Beskrovnyj, G. A. Zograf, V. P. Liperovskij. 1972. *Hindi-russkij slovar' v dvou tomax*. Moskva: Sovetskaja enciklopedija.
- Chatterdzhii, S. K. 1977. *Vvedenie v indoarijskoe jazykoznanie*. Moskva: Nauka.
- Cardona, George, Dinesh Jain (eds). 2003. *The Indo-Aryan languages (Routledge Language Family Series 2)*. London / New York: Routledge.
- Dybo, V. A. 2011. Drevneindijskij akcent v dardskom jazyke shina kak problema indoevropskoj akcentologii. In: Ye. K. Molchanova (ed.). *Leksika. Etimologija. Jazykovye kontakty. K jubileju doktora filologičeskix nauk professora Dzhoj Iosifovny Edel'man*: 92–167. Moskva: Tezaurus / Institut jazykoznanija RAN.
- Gotō, Toshifumi. 1987. *Die "I. Präsensklasse" im Vedischen: Untersuchungen der vollstufigen thematischen Wurzelpräsentia*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Hendriksen, Hans. 1986. *Himachali studies. III. Grammar*. København: Munksgaard.
- Kelkar, Ashok Ramchandra. 1968. *Studies in Hindi-Urdu. 1. Introduction and Word Phonology*. Poona: Postgraduate and Research Institute, Deccan College.
- Krylova, A. S., E. A. Renkovskaya. 2020. Kontaktnye javlenija v jazyke kullui. In: A. I. Kogan (ed.). *Trudy Instituta vostokovedenija RAN. Vyp. 29: Problemy obščej i vostokovednoj lingvistiky. Jazyk men'ajushchijs'a: semantičeskije i grammatičeskije izmenenija v jazykax Azii i Afriki*: 90–102. Moskva: Institut vostokovedenija RAN.
- Kulikov, Leonid. 2012. *The Vedic -ya-presents; Passives and intransitivity in Old Indo-Aryan*. Amsterdam: Rodopi Publishers.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge University Press.
- Mayrhofer, Manfred. 1989. *Etymologisches Wörterbuch des Altindoarischen*. Heidelberg: Carl Winter Universitätsverlag.
- Monier-Williams, Monier. 1899. *A Sanskrit-English dictionary: Etymologically and philologically arranged with special reference to Cognate Indo-European languages*. Oxford: The Clarendon Press.
- Ranganatha, M. R. 1980. *Survey of Mandeali and Kullui in Himachal Pradesh. Census of India. 1971. Language monograph No. 7*. New Delhi: Office of the Registrar General, India, Language Division.
- Renkovskaya, E. A. 2020. *Areal'no-genetičeskije i tipologičeskije faktory formirovanija posleložno-padežnoj sistemy v jazyke kumaoni. Dissertacija na soiskanie učenoj stepeni kandidata filologičeskix nauk*. Moskva: Institut jazykoznanija RAN.
- Rumjanceva, I. M. 1985. *Ritmičeskaja struktura slova v hindi. Analiz*. Moskva: Nauka.
- Singh, Maya. 1895. *The Panjabi dictionary*. Lahore: Munshi Gulab Singh & Sons.
- Tiwāri, Bholānāth. 1966. *Bhaṣa-Vigyān*. Ilāhābād.
- Turner, Ralph Lilley. 1962–1985. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.
- Turner, Ralph Lilley. 1931. *A comparative and etymological dictionary of the Nepali language*. London: K. Paul, Trench, Trübner.
- Steingass, Francis Joseph. 1892. *A Comprehensive Persian-English dictionary, including the Arabic words and phrases to be met with in Persian literature*. London: Routledge & K. Paul.
- Westergaard, Niels Ludvig. 1841. *Radices linguae sanscritae ad decreta grammaticorum definitivae atque copia exemplorum exquisitiorum illustravit*. Bonnae: H. B. König.
- Whitney, William Dwight. 1885. *The Roots, Verb-forms and primary derivatives of the Sanskrit language. A supplement to his Sanskrit grammar*. Leipzig: Breitkopf and Härtel.
- Zaloznjak, A. A. 1996. Grammatičeskij očerok sanskrita. In: V. A. Kochergina. *Sanskritsko-russkij slovar'. S prilozheniem "Grammatičeskogo očerka sanskrita" A. A. Zaloznjaka*: 785–895. Moskva: Filologija.
- Zograf, G.A., T. I. Oranskaja. 2011. Hindi jazyk. In: T. I. Oranskaja, Ju. V. Mazurova, A. A. Kibrik, L. I. Kulikov, A. Ju. Rusakov (eds.). *Jazyki mira: Novye indoarijskie jazyki*: 47–105. Moskva: Academia / Institut jazykoznanija RAN.

Anastasiya Krylova. On the possible preservation of Vedic verbal stress in the Naggar subdialect of the Kullui language

In this paper, I discuss the reflexes of Common Middle Indo-Aryan stressed vowel *a in verbal stems of the Kullui language. The accent system of Kullui generally reflects the corre-

sponding system of Middle Indo-Aryan. Common Middle Indo-Aryan stressed **a* generally yields *ɔ* in Kullui, but in some cases it also corresponds to Kullui *ε*. Etymological analysis of Kullui verbal roots containing *ɔ* and *ε* demonstrates one correlation between Kullui *ɔ* and stressed *a* and another of Kullui *ε* to non-stressed *a* in the Vedic language. Consequently, we can suppose that certain traces of Vedic accent have been preserved in Kullui verbs.

Keywords: Kullui language; accent systems; Indo-Aryan languages; Vedic language; verbal accentuation.

Ударение множественного числа имен с показателем *-ān* в парачи и его связь с праиранской акцентуацией

Парачи является одним из иранских языков, которые сохраняют рефлекс праиранского ударения, в целом наследующего индоевропейскому. В. А. Ефимов заметил для парачи акцентуационное распределение при образовании форм множественного числа существительных с формантом *-ān*: некоторые имена во множественном числе имеют ударение на основе, а некоторые – на окончании, причем такая ситуация характерна и для унаследованных, и для заимствованных основ. В статье в ходе анализа всего доступного материала делается предположение о том, что данное распределение подчиняется следующим правилам: унаследованные основы в зависимости от того, относились ли они к баритонезе или окситонезе в праязыке, соответственно имеют вариант с ударением на основе или на окончании *-ān*. В случае с заимствованиями установить точное правило не так легко, но к ударению на основе тяготеют слова с исходом на согласный вне зависимости от места ударения, а ударение на окончании *-ān* обычно имеют заимствованные существительные, заканчивающиеся ударным гласным. В результате в языке парачи выявлено 9 унаследованных слов с ударением на основе во множественном числе, а также 10 слов с ударением на окончании в аналогичных формах. Таким образом, несмотря на утрату следов праиранского ударения в формах единственного числа, во множественном числе в парачи исторически баритонированные существительные имеют ударение на основе, а окситонированные – на окончании.

Ключевые слова: язык парачи, праиранское ударение, окситонеза, баритонеза, акцентуационное распределение.

Как было установлено благодаря многочисленным исследованиям, наиболее важными из которых являются работы Г. Моргенштерне (Morgenstierne 1942; 1973; 1983), В. А. Дыбо (Дыбо 1972; 1974; 1989) и В. А. Ефимова (Ефимов 1979; 1985; 1986; 1997), рефлекс праиранского ударения лучше всего сохраняются в ряде восточноиранских языков, а именно в афганском, йидга и мунджанском, в ормури и парачи.

Наиболее последовательно сохраняется система ударения в пушту: исторически двусложные и трехсложные словоформы делятся на *oxytona* и *barytona* в соответствии с делением, наблюдаемым в ведийском и других индоевропейских языках, а производные слова, прежде всего на **-ka*, имеют систему акцентовки «по контрасту»: произошедшие от баритонированных основ становятся окситонированными и наоборот. Подробно эта система описана для именных форм в статье Дыбо 1974, для глагольных – во второй ее части (Дыбо 1989).

Несмотря на сохранение места ударения в целом в ряде именных и глагольных основ и производных именах, базовые двусложные окситонированные основы в ормури и особенно парачи обычно утрачивают первоначальную окситонезу, ср. Ефимов 1985: 49–50; Ефимов 1986: 127.

Но при этом в языке парачи есть любопытная особенность в образовании множественного числа имен, которая, как представляется, может указывать на более древнее распределение. Различие в ударении основ парачи, от которых образуется множественное число с формантом *-ān*, было отмечено и описано В. А. Ефимовым в работах Ефимов

1997; Ефимов 2009. Здесь уместно привести полностью его рассуждение относительно причин наблюдаемого распределения:

«Показатель множественного числа *-ān* восходит к окончанию род. п. мн. числа у основ на **-a*, **-ā*, т. е. к ир. **-ānām*. Происхождение двух контрастных вариантов *-ān* не вполне ясно. Сам факт существования этих вариантов, казалось бы, служит доказательством того, что в архаичном состоянии языка парачи (как и ормури, см. Ефимов 1985; 1986, 107–111, 125–138) наличествовало разноместное ударение типа ведического и что разная акцентировка данного показателя является отражением ударного и неударного прототипа (т. е. соответственно **-ānām* и **'-ānām*) наподобие того, который засвидетельствован в древнеиндийском у *a*- и *ā*-основ (Whitney 1879, 110–111, 121, 128–129; Елизаренкова 1982, 225, 227–228). Возможно, что именно к этим истокам и восходят акцентные варианты *-ān* в парачи (вопреки акцентологическим представлениям Г. Моргенштерне о количественной по своему характеру природе ударения в диалекте – предке парачи (IFL I, 30–32), которые в данном случае не могут быть применены). Однако дистрибуция этих акцентных вариантов в современных именах существительных – причем не только исконных, но и заимствованных, в том числе и довольно поздних, – не поддается объяснению и, во всяком случае, не связана с конкретными лексемами-этимонами праязыка» (Ефимов 1997: 479–480).

Тем не менее, как представляется, изначальное распределение основ с ударным и безударным формантом *-ān* в парачи, затрагивающее унаследованные лексемы, можно свести именно к древней дистрибуции баритонированных и окситонированных основ. Соответственно, множественное число окситонированных основ ожидаемо должно давать ударный формант *-ān* < **-ānām*, а множественное число баритонированных основ – безударный вариант *-ān* < **'-ānām*. В представленном далее списке будут разобраны все лексемы с ударным и безударным формантом множественного числа *-ān*, приведенные В. А. Ефимовым в работах Ефимов 1997: 479; Ефимов 2009: 38, а также случаи с колебанием в ударении. Все рассматриваемые слова будут поделены на имеющие надежную иранскую и индоевропейскую этимологию, на слова с неясной этимологией и явные заимствования.

Ударный формант *-ān*

Собственные слова парачи

1. *xi*, pl. *xiyān* ‘сестра’ < праир. **hwāhī-* ← **hwāhar-*

Очевидно, что данное слово так или иначе восходит к архаической баритонированной основе **hwāhar-*, но детали не до конца ясны. Г. Моргенштерне под вопросом дает реконструкцию **hwāhī-* (заместившую **hwahā-*) (IFL-1: 298), она же приводится в «Этимологическом словаре иранских языков» (ЭСИЯ-3: 434). Другой реконструкции придерживается В. А. Ефимов, возводящий пар. *xi* напрямую к **hwāhā*, им. п. ед. ч. от **hwāhar-* через промежуточный этап **wāhā* или **wāwā* (Ефимов 1997: 441).

Если говорить про засвидетельствованную в парачи форму, *x-* является закономерным развитием ир. **hw-* и **x^v-*: в качестве примера можно привести такие слова, как **xar-* ‘есть, пить, принимать пищу’ < **x^vāra-* ‘есть, принимать пищу’ и *xō / xo* ‘сам; себя’ < **x^vātah* (Ефимов 1997: 457). Соответственно, начальный согласный не сопряжен с проблемами.

При этом развитие гласного требует более развернутого комментария. В. А. Ефимов считает, что реконструируемая им праформа **hwāhā* трансформировалась в **wāhā* или

*wáwǎ- (Ефимов 1997: 441). Первая форма представляется маловероятной, поскольку *ǎhǎ, по формулировке самого В. А. Ефимова, сохраняется в таких случаях, как mahók ‘луна’ < *mahákā- < *mahakā- с контрастным ударением по отношению к баритонированной тематической основе *máha-, ср. др.-инд. mása- (RV+) ‘луна’ наряду с máś- [maas] ‘луна’ (существование именно этой основы подтверждается для иранских благодаря бактр. мао, маво ‘месяц (период времени)’ < *máha- и, возможно, вах. тыу ‘месяц’ < *máh(a)-, хотя значение не совпадает) и sahók ‘заяц’ < *śāhāka- < *śāhakā-, предполагается непосредственное происхождение последней основы от *śāha- ‘заячий’, ср. др.-инд. śāśa- ‘заячий’ при śāśá- ‘заяц’ (Ефимов 1997: 437); закономерная баритонеза для обозначения ‘зайца’ наблюдается в производных женского рода афг. soy, f. sóya, диал. sówa ‘зайчиха’ < *śāhakā- ← окситонеза *śahá- (NEVP: 77). Еще один вариант трактовки форм наподобие sahók в отношении суффикса – предположение о вторичном характере -k в словоформе (в разных случаях в суффиксах *k отражается в парачи по-разному: как k, как γ или выпадает; все детали распределения не ясны).

Также можно привести пример M¹ bhâγ ‘зола; земля’ < *b^hahākā-, ср. др.-инд. bhás- ‘пепел, зола’ и bhásman- ‘пепел, зола’ (отсутствует в формулировке В. А. Ефимова, ср. Ефимов 1997: 435–437, у него это слово засвидетельствовано в виде bāγ со значением ‘прах, останки’ в Ефимов 2009: 177) и отметить такую очевидную форму, как čōr ‘четыре’ < *čahāra- < *čaθwāra- (Ефимов 1997: 442).

Поэтому следует усомниться в возможности перехода áhǎ- > пар. i, реконструируемого В. А. Ефимовым для объяснения вокализма основы xi ‘сестра’.

Сочетание *-ǎvǎ-, судя по наличию таких примеров, как nō ‘девять’ < *náwa- (Ефимов 1997: 442) и агентив tō ‘ты’ < *tawa- (IFL I: 62), обычно развивается в парачи в ō. Переход *ǎvǎ > i, постулируемый В. А. Ефимовым для объяснения презентной основы bi- ‘быть’ < *báva- (Ефимов 1997: 440), сомнителен: во-первых, форма bō(n) ‘был’ может восходить к *báva-, ср. авест. 3 sg. pret. bauuaṭ (IFL I: 84), а во-вторых, -i в форме bi(n) лучше возводить к причастию *būta-, с переходом, аналогичным звуковому развитию в слове dī ‘дым’ < *dūta-, nī ‘сегодня’ < *nū(n/ram) и т. д. (Ефимов 1997: 440; IFL I: 26).

Таким образом, предпочтительнее реконструировать именно *hwāhī-, как предлагал Г. Моргенштерне, особенно если учесть, что есть определенные параллели в согдийском языке: ср. согд. будд. γw’rh, ман. хр. xw’r [x^oār] ‘сестра’ < *hwahrī- (ЭСИЯ-3: 434). В данной праформе можно ожидать окситонезы по контрасту с производным *hwáhar-; на справедливость такой реконструкции предположительно может косвенно указывать прагерманское производное *swiz-já- m. ‘сын сестры’, ср. др.-англ. swiria, -an ‘сын сестры’, др.-швед. swiri, др.-сакс. swiri ‘сын тетки по матери’. Еще один возможный источник конечного ударения – влияние таких основ, как jīnj ~ jīnč, pl. jīnján ~ jīnčán ‘женщина’ < праир. *janičī- и net, pl. netán ‘внучка’ < праир. naptī-, о которых пойдет речь в данном разделе (случаи 2 и 3).

Соответственно, в данном случае может наблюдаться историческая окситонеза или окситонеза, возникшая под влиянием сходных по значению и форме основ.

Форму множественного числа xiγán следует с осторожностью возводить к праформе *hwāhijánām (возможно, что < *hwāhijánām).

2. jīnj ~ jīnč, pl. jīnján ~ jīnčán ‘женщина’ < праир. *janičī- ← *jáni-.

Праформа *janičī с окситонезой получена путем применения правила контрастного ударения от баритонированной основы *jáni-, ср. др.-инд. jāni-, jānī f. ‘женщина; жена’, авест. jāni-, jāni- ‘женщина’ (ЭСИЯ-4: 141–143). По всей видимости, гласный второго слога

¹ Здесь и далее под M подразумевается, что данная форма засвидетельствована Г. Моргенштерне.

был рано утрачен, а при переходе ударения в промежуточной праформе **jinčī* на первый слог было получено засвидетельствованное *jinj ~ jinč*. Форма множественного числа *jinjān ~ jinčān* должна возводиться к **janič(ī)yānām*.

3. *net*, pl. *netān* ‘внучка’ < праир. *naptī-*

Праформа реконструируется на основании сравнения с др.-инд. *naptṛī, naptī-* f. ‘внучка’, авест. *naptī-* ‘внучка’. Древнеиндийские данные уверенно указывают на окситонезу. Интересно, что среди иранских языков рефлекс древней основы в таком виде есть только в авестийском и парачи (ЭСИЯ-5: 470–475).

Форма множественного числа *netān* должна возводиться к **napt(ī)yānām*.

4. *poš*, pl. *pošān* ‘сын’ < праир. **putrā-*

Баритонированная основа, непосредственное продолжение которой утрачено в пушту (ЭСИЯ-6: 318), но на место первоначального ударения указывает др.-инд. *putrā-* ‘сын; ребенок’.

Соответственно, pl. *pošān* < **putrānām*.

5. *dot*, pl. *dotān* ‘дочь’ < праир. **duxtār-*

Несмотря на то, что *dot* указывает на **duxtṛ-*, ударное окончание во множественном числе *dotān* может сохранять следы окситонезы и восходить к той стадии развития парачи, когда существовала форма **duxtār-(am)* или, скорее, **duxtṛ-ām*.

Возводить pl. *dotān* непосредственно к **duxt(a)rānām* несколько сомнительно, поскольку форма могла подвергнуться перестройке, но такое объяснение в свете остальных данных выглядит предпочтительным.

6. *boj*, pl. *bojān* ‘коза’ < праир. **būzā-*

Место ударения в праиранском и праиндоиранском устанавливается благодаря афганской форме *wuzá ~ wza*, также *uzá, bza* и *bezá* f. ‘коза’ при *wuz ~ waz* m. ‘козел’, ср. авест. *būza-* ‘козел’ (NEVP: 94; ЭСИЯ-2: 191–192). Дж. Ченг полагает, что афганское ударение вторично: согласно его анализу, поскольку данные современных иранских языков, в частности, совр. перс. *buz ~ boz* ‘козел, коза’, указывают на краткость гласного, изначально ударение в праформе было баритонированным, а затем передвинулось на последний слог с долгим гласным. Долгота в авестийском исследователем рассматривается как вторичная (Cheung 2010: 119). Тем не менее, нельзя согласиться с формулировкой правил ударения, предложенной Дж. Ченгом: согласно его правилу для основ, заканчивающихся долгими гласными, ударение должно перетягиваться на последний слог, если первый краток (Cheung 2010: 117). Многочисленные отступления от этого правила не позволяют принять его, ср. *róṇā* f. ‘полная’ < **r̥ṇā-*; *cármā* f. ‘сторона, граница’ < **čárma-* (NEVP: 18); *γóna* ‘паук, тарантул’ < **γárnyā* < **gárθn(i)yā-* ? (NEVP: 32); *stárga* ‘глаз’ < **stṛ̥-kā-* (NEVP: 77), которые, при всей неясности их индоиранских праформ, очевидно нарушают правило Дж. Ченга для слов вроде *mā* ‘поясница’ (< **melā*) < **mádyā-*; ср. др.-инд. *mádhyā-* ‘середина’.

Соответственно, в случае существования праформы **būzā-* в пушту сохранилось бы начальное ударение *#wáza*.

Итак, множественное число *bojān* ‘козы’ возводится к праформе **būzānām* с дальнейшим сокращением **ū > *u > o*.

7. *γos*, pl. *γosān ~ γosānān* ‘дом’ < праир. **wís-*

Данная корневая основа в ведийском в слабых падежах имеет ударение на окончании, ср. InstrSg *viś-ā*, DatSg *viś-é* и т. д. Соответственно, переход в окситонезу может быть вызван выравниванием по косвенным падежам. Подобное же развитие предполагалось В. А. Дыбо для пшт. *zɾə* ‘сердце’, ведийское соответствие которого, *hṛd-*, в косвенных па-

дежах имеет окситонезу, ср. InstrSg *hṛd-ā*, DatSg *hṛd-é*, GenSg *hṛd-ás*, LocSg *hṛd-í* и т. д. (Дыбо 1974: 83)².

Соответственно, форма множественного числа *γosān* восходит к **wiśānām*, ср. др.-инд. GenPl. *viś-ām* с ударением на окончании. Вариант *γosānān* мог возникнуть под влиянием форм множественного числа слов, которые заканчиваются в единственном на *-(n)ān*: *danān*, pl. *danānān* ‘зуб’; *panān*, pl. *panānān* ‘дорога, путь’ и т. п.

8. *bīn*, pl. *bīnān* ~ *bīnānān* ‘дерево’ < праир. **br̥zn(y)á-*

Праформу **br̥zná-* для слова ‘дерево’ предложил реконструировать Г. Моргенштерне: по его мнению, последовательность фонетических изменений выглядела как *bhīn* < **būhn* < **br̥zná-* (IFL I: 41). Далее основа сравнивается с праиндоевропейским названием для березы, ср. др.-инд. *bhūrjā-* ‘вид березы’. Поскольку **r̥* действительно дает *i*³, а появление придыхательного согласного может быть результатом наличия **s* или **z* в праформе, данная этимология выглядит надежно. Более того, праиранская основа **wānā-* ‘дерево’ в парачи меняет значение, ср. пар. *γan* ‘дуб’, а приобретение родового значения основой со значением ‘береза’ с учетом неизвестности березы в ареале обитания носителей парачи становится особенно вероятным (IFL I: 240). Вообще наиболее близкой к пар. *bīn* формой любопытным образом оказывается лат. *fārnus* ‘ясень’ < **b^hr̥Hǵ-nó-* или **b^hr̥Hǵ-snó-* с сокращением по закону Дыбо (Дыбо 2007: 23).

Менее удачная этимология предложена В. А. Ефимовым, который сравнивал *bīn* с авест. *barazan-* ‘высота, вышина’ (Ефимов 1997: 464).

Множественное число *bīnān* ~ *bīnānān* может быть возведено к **br̥znānām*.

9. *merzá*, pl. *merzān* ‘борона; лопатка, мастерок’ < праир. **mr̥zákā-* < **mr̥zákā-*

Сравнивается с рядом родственных слов иранских языков, образованных от **marz-* ‘прикасаться, гладить, тереть, вытирать; чистить’: белудж. *marz* ‘бревно для дробления комьев земли; борона’; семн., ласг., шам., санг. *marze*, сурх. *morze* ‘скалка’ (ЭСИЯ-5: 254–255). По всей видимости, *merzá* не является непосредственным рефлексом праир. **mr̥zákā-*, поскольку ударный **á* в парачи обычно отражается как *ō*, что было показано на примерах выше, а *-er-* не может быть рефлексом **r̥* перед слогом, содержащим *ā*. К тому же есть примеры с сохранением **k* в виде *k* или *γ* (*mahók* ‘луна’ < **māhákā-* < **māhakā-* и М *bhāy* ‘зола; земля’; Е *bāy* ‘прах, останки’ < **b^hahākā-*). Поэтому данное слово могло в парачи когда-то иметь форму без конечного гласного *#merz* < **mr̥zī-*, а затем испытать влияние форм северо-западных иранских языков.

Множественное число *merzān* в данном случае имеет окситонезу, как кажется, по той причине, что *merzá* представляет собой форму с ударением на окончании в открытом слоге.

Слова с неясной этимологией

10. *lēṛé*, pl. *lēṛé(y)ān* ‘мальчик, малыш’ (также используется супплетивная форма множественного числа *bālā*, Ефимов 2009: 212).

Этимология данного слова неясна; наиболее вероятные версии – заимствование из индоарийского или иранского источника или звукоподражательный характер лексемы. Для сравнения можно привести прежде всего такую индоарийскую праформу, как

² Более вероятно все же возведение *zr̥* к **zr̥díya-*, ср. греч. *καρδία* ‘сердце’, такое решение предлагается В. А. Ефимовым в монографии про язык ормури (Efimov 1986: 100, 132)

³ Хотя в корне на праиндоевропейском уровне реконструируется **r̥* (**r̥H*), в праиранском следовало бы ожидать **ar*, но ср. авест. *pərəna-* ‘полный’ < **p̥r̥H-no-* < *p̥r̥r̥H-no-*; другие случаи с попыткой фонетического объяснения – в работе (Cantera 2001).

T 11116 **lēlla* ‘ягненок; ребенок’ с такими рефлексами, как синдхи *lelo* ‘ребенок’; лахнда *lēlā* m., *lēlī* f. ‘ягненок, ребенок’; панджаби *lelā* m., *lelī* f. ‘ягненок’ и т. д. (Turner 1966: 647). Было отмечено возможное заимствование данной индоиранской основы в пушту в виде *lére* pl. ‘стадо ягнят, сакмал’ (NEVP: 44); в словаре М. Г. Асланова слово засвидетельствовано с другим окончанием множественного числа и ударением: *лері* мн. ч. ‘сакмал, стадо ягнят’ (Асланов 1985: 773). Интересно и производное *lerúnay* (*лерунай*) ‘барашек-сосунок (без матки)’ (NEVP: 44; Асланов 1985: 773). Препятствием для безусловного признания сопоставления формы *lērē* с пушту, лахнда или панджаби является церебрализация *r*. Еще один возможный источник – слово одного из индоарийских или дардских языков, продолжающее праформу T 10896 **laghu-da-* ‘маленький, младший’ (← **laghu-* ‘легкий’), ср. панджаби диал. *lauhṛa* ‘маленький; молодой’, сави *layéṛo* ‘младший’ и т. д. (Turner 1966: 630). В данном случае получает объяснение церебрализация, но проблему представляет вокализм.

Наконец, можно уподобить таким образованиям, как рус. *ляля*⁴ с дальнейшей частотной в языках мира диссимиляцией **l – l > l – r* и возможным влиянием индоарийских, дардских или афганских основ, рассмотренных выше. Ударение на форманте обусловлено конечным ударением данной основы с гласным исходом.

11. *wāšená*, pl. *wāšenán* ‘петух’

Г. Моргенштерне регистрирует три формы этого слова: в его записи это *bāšaná*, *wāšīnā* и *wāšenā* ‘петух’ (IFL I: 243). Он с сомнением сравнивает его с осет. диг. *wasængæ* (ирон. *wasæg*) ‘петух’, отмечая при этом, что формы парачи скорее представляют собой заимствование из неизвестного источника. В. И. Абаев считает, что осетинское слово является образованием от осет. *wasyn* ‘петь’ и представляет собой имя с внутренней формой ‘певун’, как рус. *петух*, *петел* и т. п. (Абаев-4: 54–55). Подобная возможность объяснения осетинского слова, безусловно, ослабляет предложенное Г. Моргенштерне сопоставление.

Возможно, слово как-то связано с перс. *bāša* ‘ястреб’ (также заимствовано в парачи в виде *bāša* ‘ястреб’, IFL I: 243). Скорее всего, форма *bāšaná* могла испытать его влияние, хотя, если вариант с начальным *w-* является исходным, персидское слово не может быть непосредственным источником заимствования (но может в свою очередь восходить к праформе вида **wāša-*, если само унаследовано).

Ударение множественного числа, по-видимому, обусловлено конечным ударением трехсложной формы *wāšená* вкупе с его вокальным исходом.

12. *bān*, pl. *bānán* ‘язык’ < праир. **zbān-* или персидское заимствование (IFL I: 241).

Если слово унаследованное, то оно может представлять собой основу с окситонезой *(*ži*)*žwān*- < **žízwā-*, ср. пшт. *žába* ‘язык’, но др.-инд. *jihvá-* m., *jihvā-* f., *juhí-* f. ‘язык’ с акцентовкой по контрасту или в конкретном суффиксе **-āna*. Во множественном числе в парачи установилось соответствующее конечное ударение, обусловленное окситонезой основы. Правда, в данном случае необходимо рассматривать слово в пушту как отражающее ударение, не соответствующее ведийскому и парачи, что в принципе возможно, но ослабляет приводимые доводы.

13. *č(‘)ač(‘)á*, pl. *č(‘)ač(‘)án* ‘вид воробья’

Данная основа выглядит как индоарийское заимствование, подобно пар. *č(‘)áč(‘)ó* ‘белый’. Возможно, заимствовано из одного из языков, продолжающих индоарийскую праформу T 4571 **caṭaka-* ‘воробей’ (Turner 1966: 248).

⁴ Д. И. Эдельман реконструирует для праиранского достаточно условное «детское» слово **lala-*, потому что которого имеют в иранских языках разнообразные значения: кл. перс. *lālā* ‘слуга, приставленный смотреть за мальчиками’; ягн. *lūlák* ‘младенец’; пшт. *lāla-* ‘дядя (обращение)’; тадж. диал. детск. *lolo*, *lulu* ‘хороший, милый, красивый’ и т. д. (ЭСИЯ-5: 62–63).

Прозрачные заимствования

15. *tapá*, pl. *tapáñ* ‘холм’ – заимствование из персидского, ср. перс. *tepe*, дари *tepa* ‘холм, пригорок; курган’ (Киселева-Миколайчик 1978: 197).

16. *taxtá*, pl. *taxtáñ* ‘доска’ – заимствование из персидского, ср. дари *taxta* ‘доска; щит; плита’ (Киселева-Миколайчик 1978: 202).

17. *wē*, pl. *wēáñ* ‘главная потолочная балка’ – согласно Г. Моргенштерне, заимствование из дардских или индоарийских языков Р. Л. Тернер реконструирует праформу Т 11182 **wañsiya-* ~ **wañsiya-* с такими рефлексам, как пашаи диал. лауровани *waš* ‘главная потолочная балка’; дамели *bāš*, *bāš* ‘стропило; балка’; ория *bāūsā* ‘стропило; балка’ и т. д. (Turner 1966: 652). Продолжения праир. **wa(m)ś(y)a-* или адаптированные дардские заимствования широко распространены в памирских языках, ср. вах. *was* ‘матица; основная балка потолка’; шугн. *wūs*, сар. *wys* ‘главная потолочная балка’; пшт. ваз. *wešá* ‘балка’, *baš* ‘главная балка’ (Стеблин-Каменский 1999: 387–388).

18. *sīnó*, pl. *sīnó(w)áñ* ‘грудь (женская)’ – заимствование из персидского, ср. дари *sina* ‘грудь’ (Киселева-Миколайчик 1978: 469). В данном виде представляет собой адаптированное заимствование, также Г. Моргенштерне записал форму *síná*, непосредственно соответствующую персидскому (IFFL I: 287).

Таким образом, все слова, которые в парачи исторически восходят к окситонированным основам или к корневым именам с ударением на окончании в косвенных падежах, имеют ударное окончание *-áñ*, всего это 9 основ.

Из заимствованных слов к данной категории относятся те, которые заканчиваются гласным и имеют ударение на последнем слоге. Это *tapá*, pl. *tapáñ* ‘холм’; *taxtá*, pl. *taxtáñ* ‘доска’; *wē*, pl. *wēáñ* ‘потолочная балка’ и *sīnó*, pl. *sīnó(w)áñ* ‘грудь (женская)’.

Три основы с неясной этимологией примыкают по своему внешнему облику к заимствованным и могут быть объяснены сходным образом: *lērē*, pl. *lērē(y)áñ* ‘мальчик, малыш’; *wāšená*, pl. *wāšenáñ* ‘петух’ и *č(‘)ač(‘)á*, pl. *č(‘)ač(‘)áñ* ‘вид воробья’.

Наконец, только одна основа с неясной этимологией заканчивается на согласный, *bāñ*, pl. *bāñáñ* ‘язык’, причем в том случае, если это унаследованное слово с праформой **zbāñ-* < **(ži)zwan-*, оно подчиняется общему правилу для первой части основ.

Таким образом, проведенный анализ показывает, что к словам с ударным окончанием *-áñ* в парачи относятся основы, имеющие историческую, «скрытую» окситонезу и те, которые обнаруживают окситонезу синхронную, «поверхностную», реализуемую стандартно как ударный последний открытый слог преимущественно в двусложных формах. Примечательно, что в списке из 18 основ, который практически идентичен составленному В. А. Ефимовым, нет исключений.

Неударное окончание *-áñ*

Собственные слова

1. *dōst*, pl. *dōstáñ* ‘рука’ < праир. **dásta-* / **zásta-*

Праформа **zásta-* реконструируется на основании др.-инд. *hásta-* ‘рука’, авест. *zasta-* ‘рука’. В подавляющем большинстве уже древних иранских языков праформа **zásta-* заменена на **dásta-*, при этом регулярный фонетический переход **ž- > d-* характерен только для юго-западных иранских языков (ЭСИЯ-2: 371). Предлагались различные объяснения

замены **ž* на *d*-, *δ*-, *l*- в иранских языках разных эпох, но важнее, что такие формы, как афг. *lās* ‘рука’, также указывают на баритонезу.

Фонетический переход **á* > *ō* (в записи Г. Моргенштерне *ō*) характерен для парачи и является еще одним указанием на баритонезу, ср. *dōs* ‘десять’ < **dása*-; *pōnč* ‘пять’ < **pánča*-; *sōr* ‘голова’ < **sāra*- и т. д. (IFL I: 22–23).

Поэтому форма множественного числа в парачи *dōstān* закономерно восходит к **dástānām*.

2. *bōr*, pl. *bōrān* ‘дверь’ < праир. **dwár*- / **dwār*-

В языке корневое имя, судя по таким формам, как вед. AccPl. *dúras*, *dvāras* = греч. *θύραζε*, и тематизированному имени *dvāra*- ‘дверь’ (EWAia I: 764–765), относилось к типу с накоренным ударением и, соответственно, переходило в баритонированное тематическое существительное **dwāra*- или **dwāra*-. Для праиндоевропейского состояния на основании привлечения данных в том числе балтославянских и германских языков баритонеза реконструируется для этого корневого имени в работе (Дыбо 2014: 33). Впрочем, нельзя не отметить, что накоренное ударение в подобных формах может быть характерным и для односложных имен с подвижным ударением, ср. вед. асс. *púras* (Дыбо 2014: 32). Также на баритонезу указывает пшт. *war* < **dwāra*-, хотя данные **a*-основ наименее показательны (NEVP: 89).

Соответственно, pl. *bōrān* < **dwārānām*.

3. *giš*, pl. *gúšān* ‘ухо’ < праир. **gáwša*-

Ударение в праформе устанавливается благодаря др.-инд. *ghóṣa*- ‘шум, крик, рев; шум, молва’, пшт. *γwaž* (*γwaḥ*) ‘ухо’, в кандагарском диалекте также ‘шум’ (Дыбо 1974: 78).

Здесь необходимо отметить, что *giš*, pl. *gúšān* – несомненно, заимствованная или деформированная под давлением персидского форма данного слова, поскольку в парачи интервокальный *š* выпадает, ср. *espó* ~ *espóh* ‘вошь’ < праир. **šwiš*- ‘вошь’, авест. *spiš*- ‘вошь’. Но в шутульском диалекте сохраняется унаследованная форма *gi* при соответствующих пачаганской *giš* и гочуланской *gōš*; при этом шутульская форма имеет множественное число *gúšān* (Ефимов 2009: 196), которая может быть перестроена из **gúān*.

Препятствием для такой трактовки может служить однажды засвидетельствованная Г. Моргенштерне форма множественного числа в тексте, а именно *gūān*: *gūān tar-au jímki sī* ‘in thine ears are ear-rings’ (IFL I: 219). В данном случае определено унаследованное множественное число имеет ударный вариант окончания. Если предполагать, что он не является исконным, можно думать о переходе ударения на окончание из-за того, что форма *gi* заканчивается гласным, хотя, естественно, такое предположение нельзя назвать достаточно убедительным.

Аналогичное ударение имеет множественное число от слова *gi* ‘корова’ – *gūān* (Ефимов 1997: 480), но этот вариант не является единственным; более стандартный – *gi-γān* ~ *gi-hān* с суффиксом множественности *-γān*, который является ударным. Соответственно, могло быть и давление похожего множественного *gūān* ‘коровы’, хотя это и достаточно маловероятный сценарий.

Таким образом, только в том случае, если форма pl. *gúšān* является перестройкой более старой формы **gúān*, она восходит к **gáwšānām*. Данный пример сомнителен.

4. *tox*, pl. *tóxān* ‘лицо; щека’ < праир. **túxa*-

Ср. др.-инд. *túkha*- ‘лицо, рот; устье, передняя часть’; пшт. *tāx* ‘лицо’ с баритонезой (Дыбо 1974: 77). Форма парачи не является заимствованием (IFL I: 274), на это также указывает полисемия в парачи, отличающаяся от той, что засвидетельствована в афганском соответствии. М. Г. Асланов перечисляет следующие значения слова *tāx*: «1) лицо; 2) сторона, край; 3) верх, поверхность; 4) мат. грань; 5) сиденье (напр. стула); 6) страница; 7) лицо, лицевая сторона (ткани); 8) заготовка (для обуви); 9) лезвие (сабли, ножа); 10) пе-

рен. честь; слава; репутация, авторитет, влияние» (Асланов 1985: 801–802). Дополнительно баритонезу доказывает пшт. *táxa* ‘сторона, направление’ < **tíxā-* (Асланов 1985: 805), не отмеченное в NEVP.

Форма множественного числа *tóxān* восходит к **tíxānām*.

5. *naṓn*, pl. *naṓnān* ‘хлеб’ < праир. **nikāna-*

Праформа данного слова обычно понимается как приставочное образование от глагола **kan-* ‘копать’, т. е. ‘закопанный в / под горячую золу’, ‘положенный в/под горячую золу’ либо ‘засыпанный горячей золой’ (ЭСИЯ-5: 418–419). Несмотря на то, что для большинства иранских языков предпочтительнее реконструировать промежуточную праформу **naḡana-* или **naḡ(a)na-* (ЭСИЯ-5: 418), парачи *naṓn* может возводиться к **nikāna-* напрямую: праир. **á* под ударением переходит в *ō*, **k* может в интервокальной позиции давать *γ*, а **i* в безударной позиции иногда переходит в *a*, что особенно важно – в приставочных образованиях с **ni-*: ср. *naḡō* ‘вышел’ < **nirgata-* (PIFL I: 25). Из других иранских языков прежде всего интересны бактр. *vaḡavo* ‘хлеб’ и пшт. *n(a)ḡān* f., pl. *naḡni* ‘пшеничный хлеб; хлеб, печенный на сковороде’; последнее дополнительно указывает на место праиранского ударения.

Множественное число *naṓnān* возводится к **nikānām*.

6. *giwān*, pl. *giwānān* ‘вымя’ < праир. **gaw-dāna-* или **gaw-rāna-*

Основа **gau-dāna-* прослеживается в ряде иранских языков, из этих форм наиболее важное продолжение – пшт. *γulānza*, *γwālānza* < **gu-dāna-čī* с более поздним суффиксом *-a* < **-ā* (ЭСИЯ-3: 215). Наиболее вероятно, что праформа **gu-dāna-* имела именно такое ударение, а дополнительные суффиксы не изменили его место (также ср. др.-инд. *dāna-* ‘дар, дарение’, пшт. *prólāna* ~ др.-инд. *pradāna-* ‘дар, подарок, подношение’). Единственное препятствие для такой реконструкции – нерегулярный согласный *w* на месте **d* (в парачи интервокальный *w* мог бы быть получен только из **r* или **b*).

Если предположить наличие **r* в праформе и иранское происхождение основы, то сложное слово **gaw-rāna-* должно быть образовано от **rā-* ~ **rāy-* ‘пить’ (ЭСИЯ-6: 290–291), но в иранских языках надежно прослеживаются только глагольные основы и рефлексивы причастия **pita-*. Осет. ирон. *fædæg* ‘сосок’ возводился В. И. Абаевым к **pātaka-*, но *æ* соответствует праир. **a* (Абаев I: 427–428; ЭСИЯ-6: 291), поэтому Дж. Ченг предпочитает с сомнением реконструировать **pataka-*, производя данную основу от **pat-* ‘падать’ (Cheung 2002: 183), хотя предлагаемая им этимология слаба семантически. Более интересна этимология В. И. Абаева для осет. ирон. *fæzdon*, диг. *fæzdonæ* ‘вымя’ < **pat-dāna-* или **pāt-dāna-* с сокращением гласного первого слога (Абаев I 428; 467–468), но и она вряд ли верна из-за сомнительного морфологического образования (Cheung 2002: 188). Соответственно, реконструкция праир. **rāna-* сомнительна.

Поэтому вторая часть композита может быть старым заимствованием или калькой из пашаи, ср. пашаи диал. вегали *rān*, диал. арети *rōn* ‘вымя’, которое Р. Л. Тернером реконструируется как Т 8081 **rāna-* ‘питье’, ср. др.-инд. *rāna-* ‘питье, напиток’ (Turner 1966: 456) или, что вероятнее, как Т 12763 / 14685 **śwāna-* ‘вымя’ < **śwā-* ‘набухать’ (Turner 1966: 740, 832). Возможно, основа **gaw-rāna-* появилась в результате контаминации с распространенным праиранским композитом **gau-dāna-*.

Таким образом, pl. *giwānān* может теоретически возводиться к pl. **gawdānām* или **gawrānām* (хотя праформа, естественно, условна).

7. *ātaḡ*, pl. *ātaḡān* ‘яблоко’ < праир. *(*ā*)*marnā-* < **malnā-*

Праиранская основа с колебанием вариантов с начальным гласным и без него: праформа с начальным гласным дала согд. **m’nk* [*atang*] ‘яблоко’; мундж. *āteīḡā* ~ *āteīḡa*, йидга *ātino* ~ *ātino* ~ *atino*; пшт. *taḡā* ‘яблоко’ (с отпадением начального гласного) (ЭСИЯ-5: 238–239; PFL II: 190; NEVP: 51). Предполагается, что это ареальная основа: с од-

ной стороны, ср. лат. *mālum* ‘древесный плод; яблоко’, греч. μῆλον, дор. μᾶλον ‘яблоко’, с другой – пратюрк. **alma* ‘яблоко’ (ЭСИЯ-5: 238–239).

Афганская форма указывает на окситонезу, этой акцентовке противоречат данные мунджанского и парачи. Форма *ámaṛ*, кроме того, не демонстрирует переход **á > ō*, что может говорить о позднем характере ударения в имеющейся форме. Ударение во множественном числе *ámaṛān* может быть обусловлено начальным ударением в единственном числе *ámaṛ*.

8. *a(y)ír*, pl. *a(y)írān* ‘облако’ < праир. **abryá-*

Окситонированная основа **abrá-* реконструируется на основании др.-инд. *abhrá-* ‘облако’. Пшт. *ōwrá*, также *orá* и *wará*, диал. *awár* ‘облако, тучка; бельмо’ наиболее вероятно восходит к **abryá-* (NEVP: 10), хотя производная основа с суффиксом могла бы иметь контрастную акцентовку.

В целом случай трудный для трактовки. Если для парачи *a(y)ír* реконструируется окситонированная праформа **abryá-*, соответствующая афганским данным, ударение в форме множественного числа *a(y)írān* незакономерно и представляет собой исключение, возникшее, вероятно, в результате перестройки основы. Это наиболее вероятная версия. Если же существовал не совпадавший с реконструируемым для пушту вариант ударения **ábrya-*, то ударение в форме множественного числа ожидаемо, но форма опять же была перестроена хотя бы потому, что ударение переместилось. Такому предположению противоречат вокализм (**á* не перешло в *ō*) и несовпадение места акцента с тем, которое засвидетельствовано для пушту.

Одно из исключений из правил.

Слова с неясной этимологией

9. *aštáw*, pl. *aštáwān* ‘живот’ < **stáfa-* ?

Г. Моргенштерне предположительно связывает данное слово с др.-инд. *stabhi-* ~ *stambhi-* ‘укреплять, защищать’ (IFL I: 237). Можно с осторожностью думать о связи с др.-инд. *sthávi-* ‘ткач’ и *sthavi-* lex. ‘мешок, сумка’ с соответствиями в среднеиндийских и новых индоарийских языках, ср. T 13746 (EWAia III: 524–525; Turner 1966: 793). Вокализм нерегулярен (нет перехода в *ō*).

Праиндоиранское ударение не прослеживается, поэтому баритонеза предположительная.

Множественное число должно возводиться к **stáfānām*, как бы эта форма ни объяснялась этимологически.

10. *tōx*, pl. *tōxān* ‘козел’ < праир. **táxa-/táka-* ?

Несмотря на то, что в очерке В. А. Ефимова данное слово рассматривается «в числе существительных, засвидетельствованных с обоими акцентными вариантами» (Ефимов 1997: 479), которые в целом будут рассмотрены ниже, в монографии «Язык парачи» формы множественного числа, образованные от *tōx*, были описаны следующим образом: «*tōx* ‘козел’ — мн.ч. *tax* (наряду с *tōxān*, *taxān*)» (Ефимов 2009: 40).

Более архаичным вариантом множественного числа из перечисленных представляется первый с перегласовкой; несколько позже или параллельно с ней была образована форма *tōxān*, а *taxān* с совмещением двух способов образования множественного числа кажется самой поздней. В таком случае форма *tōxān* (Ефимов 2009: 13), скорее всего, вообще имеет вторичное ударение по образцу *taxān*. Поэтому баритонеза доказывается самыми архаичными формами *tax* и *tōxān*, что хорошо согласуется с тем фактом, что в парачи *ō < *á* (IFL I: 22–23), ср. *dōst*, pl. *dōstān* ‘рука’.

Г. Моргенштерне записал форму *tōy* ‘горный козел-самец’ с другим конечным согласным от двух информантов (IFL I: 292). Похожие формы встречаются в других индоиранских языках, ср. вах. *tuṣ* ‘коза’; ягн. *taká* ‘домашний козел’; кхов. *toy* ‘нехолощенный козел’ (Стеблин-Каменский 1999: 360–361). Судя по перс. *taka* ‘козел’, заимствованному, в свою очередь, из тюркских, может быть очень древним заимствованием (там же), произошедшим до фонетического перехода **á > ō* в парачи и **a > u* в ваханском с не до конца понятными деталями.

В связи с этим можно условно реконструировать форму множественного числа как **táxānām* / **tākānām* с более старой *tax* < **taxās* ?

11. *γárāk*, pl. *γárākān* ‘ворона’

Этимология неясна. Напоминает др.-пруск. *warne* (вторичная *ē*-основа), лит. *várna*, сербохорв. *vàrna*, рус. *ворóна* (IEW: 1166), но установить непосредственную связь сложно.

Безударное множественное может быть обусловлено начальным ударением в двухсложном слове *γárāk*.

12. *tanók*, pl. *tanókān* ‘овца’ < праир. **mayšnákā-* < **mayšnakā-* ?

Баритонеза для **mayšnakā-* может быть незакономерной ввиду др.-инд. *meṣá-* m. ‘баран’, *meṣí-* ‘овца’ < **mayśá-*. На баритонезу, впрочем, указывают афганские формы, ср. пшт. *maž* ‘баран’, *mež* f. ‘овца’, которые В. А. Дыбо относятся к исключениям, но при этом предлагает реконструировать пару **moysós* ~ **móysuə*, «с дальнейшим выравниваем по языкам» (Дыбо 1974: 85–86). Впрочем, рассуждения об ударении в обсуждаемой праформе осложняются также наличием суффикса **-na*.

В любом случае pl. *tanókān* можно напрямую возвести к **mayšnákānām*. Другим объяснением может быть автоматическое сохранение ударения на *ō* по образцу других основ, содержащих этот гласный.

13. *teč* ~ *teč*, pl. *téčān* ~ *téčān* ‘глаз’

Г. Моргенштерне справедливо указал, что связь с перс. *čāšt* ‘глаз’ невероятна (IFL I: 292); попытка сравнения с перс. *dāda* ‘глаз; око’ и реконструкции **dhīā* < **dihā* там же выглядит столь же сомнительно.

На мой взгляд, может быть трансформированным заимствованием из дардских языков, где праформа Т 43 **akši* дает такие слова конкретных языков, как пашаи *ačhí*, *ač*; дамели *ēč*; кхов. *yeč(h)* и т. д. (Turner 1966: 2–3).

14. *mayás*, pl. *mayásān* ‘муха’ < **makaśā-* ?

Г. Моргенштерне полагал, что это унаследованное слово, для которого может быть реконструирована праформа **makaśā-* (Morgenstierne 1926: 20; IFL I: 23), однако ввиду вокализма (ударный **a* дал бы *ō*) может быть адаптированным заимствованием, ср. дари *magas* ‘муха’ (Киселева-Миколайчик 1978: 665–666).

Если слово унаследовано, необходимо реконструировать окситонезу **makaśā-* с относительно поздней оттяжкой ударения на второй слог, из-за чего **a* удерживается. Внешними данными однозначно подтвердить ударение не удастся, но ср. йидга *mojúso* ‘муха’ (ЭСИЯ-5: 174).

Если это так, то *mayás*, pl. *mayásān* – наиболее заметное исключение из правил. Оно может объясняться поздним появлением формы множественного числа и возводиться условно к **makásānām*.

Прозрачные заимствования

15. *bolból*, pl. *bolbólān* ‘соловей’ – заимствование из персидского, ср. дари *bolból* ‘соловей’ (Киселева-Миколайчик 1978: 120).

16. *kalá*, pl. *kaláân* ‘голова’ – заимствование из персидского, ср. дари *kalla* ‘череп; голова’ (Киселева-Миколайчик 1978: 589).

17. *mâṇḍ*, pl. *mâṇḍân* ‘гнездо’.

По форме это слово выглядит как индоарийское или афганское заимствование, но непосредственный источник не до конца ясен. Ср. пашаи *mān* ‘гнездо’ < Т 10042 **māna*- ‘дом, жилище’, ср. др.-инд. *māna*- ‘постройка; жилище’ (Turner 1966: 577). В пушту есть слово *мандзола* ‘гнездо’ (Асланов 1978: 855). Пшт. *манḍ*, наиболее напоминающее по форме парачи *mâṇḍ*, имеет следующий круг значений: «1) шаг; 2) след (ног); 3) пень; 4) дорога по насту; дорога (тропа) по снегу; 5) основа, опора, фундамент; 6) источник (зла и т. п.)» (Асланов 1978: 856). Можно с осторожностью предположить, что в отдельных диалектах пятое значение может развиваться в ‘гнездо’.

18. *ḡonḍî*, pl. *ḡonḍîyân* ‘горка, холм’ – по всей видимости, заимствование из пшт. *ḡonḍây* ‘холм; горка’, которое связывается с *ḡunḍ* ~ *ḡwânḍ* ~ *ḡwunḍ* ‘круглый, округлый; шарообразный, шаровидный’ (NEVP: 31).

Очевидно, что это слово выбивается из правила для окситонированных основ и заимствованных слов с гласным исходом и ударением на этом последнем гласном, которое было сформулировано при разборе слов с ударным окончанием *-ân*. Можно предположить, что старая форма заканчивалась, как и в пушту, на согласный, т. е. **ḡonḍîy*.

19. *ketâb*, pl. *ketâbân* ‘книга’ – прозрачное персидское или афгано-персидское заимствование арабского происхождения, дари *ketab* ‘книга’ (Киселева-Миколайчик 1978: 577). В отношении вокализма ср. пшт. *китâб* ‘книга’ (Асланов 1978: 662).

20. *nâk*, pl. *nâkân* ‘груша’ – заимствование из персидского, ср. дари *nâk* ‘груша (плод)’ (Киселева-Миколайчик 1978: 688).

21. *šâx*, pl. *šâxân* ‘ветвь’ – заимствование из персидского, ср. дари *šaxa* ‘ветвь, сук; ответвление, отрог; отрасль’ (Киселева-Миколайчик 1978: 470).

Можно видеть, что унаследованные основы регулярно сохраняют ударение на основе в том случае, если в праязыке они были баритонированными. Наиболее важные примеры в данном случае – те, что содержат ударный гласный *ô*, поскольку в парачи праиранский краткий гласный **a* под ударением дает как раз *ô*. Это 4 непроезвучившие основы с *ô*, *dôst*, pl. *dôstân* ‘рука’; *bôr*, pl. *bôrân* ‘дверь’; *naḡôn*, pl. *naḡônân* ‘хлеб’ и *tôx*, pl. *tôxân* ‘козел’, а также одна суффиксальная, *manôk*, pl. *manôkân* ‘овца’.

Что касается заимствованных основ, то не переносят ударение на окончание во множественном числе односложные и двусложные основы, заканчивающиеся согласным, вне зависимости от ударения. Особенно интересно отметить значительное количество основ с вокализмом *â*: *mâṇḍ*, pl. *mâṇḍân* ‘гнездо’; *ketâb*, pl. *ketâbân* ‘книга’; *nâk*, pl. *nâkân* ‘груша’ и *šâx*, pl. *šâxân* ‘ветвь’.

Среди рассмотренного 21 случая есть 2 исключения: . *a(y)îr*, pl. *a(y)îrân* ‘облако’ < праир. **abryá-* и *kalá*, pl. *kaláân* ‘голова’. Слово *ḡonḍî*, pl. *ḡonḍîyân* ‘горка, холм’ только кажется исключением, потому что более старый облик данного слова – **ḡonḍîy* (пшт. *ḡunḍây*).

Слова с колебанием в ударении

1. *dust*, pl. *dústân* и *dustân* ‘друг’ – заимствование из персидского, ср. дари *dôst* ‘друг, приятель’ (Киселева-Миколайчик 1978: 380). В материалах Г. Моргенштерне слово зафиксировано в виде *dúst* и *dôst* ‘друг’ (IIFL I: 251).

2. *siz*, pl. *sízân* и *sizân* ‘грудь’

Наиболее вероятно, что это основа собственного происхождения звукоподражательного характера, как осет. ирон. *zīzī*, диг. *zeze* ‘женская грудь’, арм. *sic*, ягн. *šič* ~ *šeč* ‘то же’ и т. д. (Абаев I: 397).

3. *nēšt* ~ *nēšt*, pl. *nēštān* и *nēštān* ‘нос’ < праир. **nāstī*-

Г. Моргенштерне реконструирует праформу **nāstī*- и указывает, что от такой же по виду основы образованы формы пашаи и других индоарийских и дардских языков (IFL I: 277), ср. Т 7031 пашаи диал. лауровани (Laurōwānī) *nāst*, *nās* < **nāsta*-; вотапури (Woṭapūrī) *nyet* f., пхалура *nāst* f. ‘нос’ (Turner 1966: 403–404). Также основы с суффиксом, содержащим **t*-, засвидетельствованы и в иранских языках: бахтиари *noft* ‘нос’ (Anonby-Asadi 2014: 157), менее вероятно йидга *nezyo*, мундж. *nazyā* ‘слизь из носа’ < **nēz(d)gā* < **nāstikā*- ? (IFL II: 235; ЭСИЯ-5: 425). Переход **st* > *št* перед *ī* характерен для парачи, ср. *yošt*, М *yušt* ‘брошенный’ < **wista*- (IFL I: 43).

Ударение в праформе установить на основании непосредственного сравнения с формами, которые бы соответствовали **nāstī*-, невозможно.

4. *mēz*, pl. *mēzān* и *mēzān* ‘стол’ – персидское заимствование, ср. дари *mēz* ‘стол; верстак’ (Киселева-Миколайчик 1978: 679).

5. *māwṛ*, pl. *māwṛān* и *māwṛān* ‘сад’

Может быть индоарийским или дардским заимствованием, особенно учитывая наличие *r* в словоформе. Непосредственный источник заимствования невозможно установить.

Из известных праформ, с которыми можно было бы предположительно связать парачи *māwṛ*, следует назвать праир. **marwā*- ~ **margā*- ‘луг’. Рефлексы звука *w* наличествуют в персидском, ср. ср.-перс. *marv*, *marv* ‘трава, (луговые) растения’, кл. перс. *marv* ‘название пахучей травы’; в остальных языках наблюдаются рефлексы **g*, ср. парф. *mary* [mrg] ‘лес; лужайка’; согд. *mry(h)* ‘луг’, ягн. *mary* ‘трава – лучший корм для скота’; пшт. *maryá* ‘пальчатник, свиной (злак); всякая трава, дающая дерновину; газон, лужайка’; мундж. *mírya* ~ *múrya* ‘луг’ и т. д. (ЭСИЯ-5: 231–233). Также привлекает внимание пшт. *mavṛí* pl. m. ‘белая пшеница’ (Асланов 1985: 865). Возможно, парачи *māwṛ* – старое заимствование из одной из похожих форм с метатезой **-rw-* > **wṛ* и вторичной церебрализацией *r*.

6. *pā*, pl. *pān* ~ *pānān* ‘нога’ < праир. **pād(a)*-

Данный случай должен рассматриваться скорее наряду с такими, как *bīn*, pl. *bīnān* ~ *bīnānān* ‘дерево’ и *yoś*, pl. *yośān* ~ *yośānān* ‘дом’.

Множественное число *pān* должно объясняться, скорее всего, как получившееся по правилу для форм, заканчивающихся гласным звуком, поскольку корневое имя **pād*- ‘нога’ имело именительный падеж **pād* и косвенные падежи с ударением на окончаниях, то есть подвижную акцентную парадигму, которая полностью реконструирована в докладе (Дыбо 2014: 34). Родительный падеж множественного числа **pādām*, с другой стороны, мог послужить непосредственным источником формы *pān*, поскольку интервокальный **-d-* в парачи выпадает, ср. парачи *namś* ‘девятнадцать’ < праир. **nāvadaśa*-, др.-инд. *nāvadaśa*- ‘девятнадцать’ (IFL I: 36).

Можно видеть, что шесть основ с колебанием ударения, приведенные В. А. Ефимовым, поровну делятся на унаследованные и заимствованные. Какие-то закономерности ввиду скудности материала сложно установить, к тому же из-за отсутствия надежных количественных данных нельзя выдвигать предположения о том, какие именно варианты – с ударным или безударным окончанием – являются изначальными в рассматриваемой группе случаев. Интересно, что три слова с колебанием ударения во множественном числе, *dust* ‘друг’, *nēšt* ~ *nēšt* ‘нос’ и *māwṛ* ‘сад’, имеют в коде два согласных, а *siz* ‘грудь’ и *mēz* ‘стол’ заканчиваются одним согласным, *-z*, но является ли это фактором, влияющим на ударение форм множественного числа, неясно.

После рассмотрения всех слов, отмеченных В. А. Ефимовым в его очерке и грамматике, необходимо с известной долей скептицизма рассмотреть все случаи, отмеченные в его парачи-русском словаре, который прилагается к грамматике (Ефимов 2009: 170–254). Дело в том, что словарь опубликован посмертно и некоторые ударения в нем приводятся ошибочно, например, для слова *bôr* ‘дверь’ приводится форма множественного числа *bôrân*, противоречащая другим материалам автора и маловероятная ввиду вокализма данного слова (Ефимов 2009: 179). Кроме того, система обозначения ударения форм множественного числа в словаре непоследовательна, для некоторых основ место ударения во множественном числе вообще не указано, поэтому я и решил не включать материалы из словаря в основной список, где приводятся случаи, внимательно отобранные автором. Примеры даются в точности как у автора в словаре во избежание путаницы.

Слова, имеющие окончание *-ân* в словаре Ефимова

1. *ōsp*, мн. ч. *ōspân/ asp, aspân* ‘лошадь’ (Ефимов 2009: 224).

Судя по распределению, изначальная пара *ōsp* ~ pl. *asp* была видоизменена с добавлением окончания *-ân* как к форме единственного, так и к форме множественного числа. Ударение в формах *ōspân* и *aspân* противоречит баритонезе *ōsp* < **áswa-*, ср. др.-инд. *ásva-* ‘конь’. Причины, по которым окончание ударно, неясны. Быть может, на передвижение ударения мог влиять тот факт, что формы *ōsp* и *asp* заканчиваются сочетанием согласных.

2. *bahr*, *-ân* ‘море’ (Ефимов 2009: 175).

Арабо-персидское заимствование, ср. дари *bahr* ‘море’ с ломаным множественным *behâr* и *abhâr* (Киселева-Миколайчик 1978: 100).

Место ударения в словарной статье не указано. Наиболее вероятно, что слово имеет безударное окончание *-ân* или колебание в ударении, что не противоречит правилам для заимствований (ударность окончания в словаре обозначается в том числе знаком ударения над *-ân*).

3. *bâl*, мн. ч. *-ân* ‘крыло; перо’ (Ефимов 2009: 176).

Прозрачное персидское заимствование, ср. дари *bâl* ‘крыло; перо, плавник (рыбы)’ (Киселева-Миколайчик 1978: 95).

Удостовериться в том, что окончание в этом слове безударно, помогает пример из текста: *xokân bâlân-tar-a páda-padá bâlân déra* г. ‘у нее на крыльях черные-пречерные перья’ (Ефимов 2009: 164). В этом же примере, записанном во вхождении словаря, допущена ошибка в форме *bâlân-tar-a*, которая представлена в виде *bâlân-tar-a* (Ефимов 2009: 176).

Слово подчиняется правилам для заимствованных односложных основ с согласным исходом и вокализмом *â*.

4. *bâló*, мн. ч. *bâlá, bâlân* ‘мальчик, ребенок (от года до 10 лет)’ (Ефимов 2009: 176).

Пример из текста: *bâlân če žujâysák nântân* ‘усадили мальчишек в каком-нибудь месте’ (Ефимов 2009: 147). Г. Моргенстьерне наряду с формой множественного числа *bâlân* отмечает также *bâlú* (IFL I: 240).

Дардское заимствование, ср., например, пашаи *bārā* ‘мальчик, юноша’ (IFL I: 240; Turner 1966: 514).

Соответствует правилу для заимствованных основ с гласным исходом.

5. *bâw*, мн. ч. *-á, -ân, -yâr* ‘отец’ (Ефимов 2009: 176).

Дардское заимствование, ср. пашаи *bâw* ‘отец’ (IFL I: 243).

В данном случае ударение, очевидно, под влиянием других вариантов образования множественного числа: так, суффикс *-yâr* всегда является ударным.

6. *beâ, byâ, beâr* ш., мн. ч. *beârâ* и *beârân* ‘брат’ (Ефимов 2009: 177).

Сложный случай. Г. Моргенштерне справедливо замечает, что эту слово проблематично возводить напрямую к **brâtar-* (IFL I: 42), поэтому речь может идти о перестроенной форме, в которой конечное ударение стало восприниматься как окситонеза.

7. *b(o)ôta* п., *buța* ш., мн. ч. *-ân* ‘куст, кустарник; сорняки; хворост’ (Ефимов 2009: 178).

Дардское или индоарийское заимствование, ср. Т 9297 **bûta-* ‘кустарник, растение’ (Turner 1966: 527).

Г. Моргенштерне отмечает ударение на первом слоге для варианта М *bûta* ‘plant, shrub’ (IFL I: 243), которое является персидским заимствованием, ср. дари *bota* ‘куст, кустарник’ (Киселева-Миколайчик 1978: 99).

Подчиняется правилу для заимствованных основ с гласным исходом.

8. *čiz*, мн. ч. *-ân* ‘предмет, вещь’ (Ефимов 2009: 184).

Персидское заимствование, ср. дари *čiz* ‘вещь, предмет’ (Киселева-Миколайчик 1978: 296).

Ударное окончание не соответствует правилу для односложных словоформ с одним согласным в коде; но ср. *mēz*, pl. *mēzân* и *mēzân* ‘стол’ с ударным вариантом окончания.

9. *čur* п., ш., *čur* п., ш., мн. ч. *-ân* ‘вор’ (Ефимов 2009: 184).

Ударение подтверждается примерами из текстов, ср. *če čurân ná-bara* ‘чтобы не унесли воры’ (Ефимов 2009: 140). Дардское заимствование, ср. пашаи *čūr* ‘вор’ (IFL I: 246; Turner 1966: 269).

Ударение закономерно, одно из исключений.

10. *danân*, мн. ч. *-ân* ‘зуб’ (Ефимов 2009: 185).

Ошибка в словарной статье: судя по имеющимся примерам, на самом деле имеет безударное окончание *-ân*. Ср. *bân wa danânân mēn šoŋd-tar sēn* ‘язык и зубы находятся во рту’ (Ефимов 2009: 162). Также именно такая форма множественного числа отмечена у Г. Моргенштерне в виде *danânân* (IFL I: 162, 212, 219)

Продолжает праир. **dantâna-*, pl. **dantânânâm*; такая основа достаточно широко представлена в иранских языках (ЭСИЯ-2: 329–330)⁵. Тем не менее, это однозначно унаследованная основа, а не персидское заимствование, потому что в ней происходит характерный для парачи фонетический переход **-nt- > n* (IFL I: 37). В некоторых дардских языках тоже засвидетельствован такой переход, но заимствование из них слова со значением ‘зуб’ маловероятно.

Подчиняется правилу для баритонированных имен.

11. *dar(r)á*, мн. ч. *dar(r)ân* ‘ущелье, долина’ (Ефимов 2009: 185).

В парачи заимствование из пушту или дари, ср. пшт. *dar(r)á* ‘долина; ущелье, горный проход’ (Асланов 1985: 407), дари *dara* ‘ущелье, каньон; горный проход; долина’ (Киселева-Миколайчик 1978: 356).

12. *dâr, dâr* ш., мн. ч. *-ân* ‘гора, холм’ (Ефимов 2009: 187).

Дардское заимствование, ср. пашаи *dhâr* ‘холм’ (IFL I: 248; Turner 1966: 388).

Ударение на окончании подтверждается примерами из текстов, ср. *hendukošeká dârân-tar* ‘через горы Гиндукуша’ (Ефимов 2009: 137).

Данная основа выпадает из установленного распределения.

13. *espâr*, мн. ч. *-ân* ‘деревянная соха с железным лезвием’ (Ефимов 2009: 193).

Заимствование из дари, ср. дари *espâr* ‘деревянная соха с железным лезвием’ (Киселева-Миколайчик 1978: 40).

Ударение на основе в форме множественного числа нельзя доказать примерами из текстов, но правилам она не противоречит.

⁵ Формы парачи в словаре не отмечены.

14. *éstēč*, мн. ч. *-ān* ‘звезда’ (Ефимов 2009: 193).

Ср. *mahók estēcān-tar gāṇḍ-ā* ‘луна больше звезд’ (Ефимов 2009: 132), но у другого информанта *mahók az éstēcān gāṇḍ-a* ‘луна больше звезд’ (Ефимов 2009: 154).

Надо отметить, что сложно установить, характеризовалось ли имя **h₂ster-* подвижностью или неподвижностью ударения в праязыке: у корневого имени *star-* ‘звезда’ в Ригведе ударение на основе, InstrPl *stṛbhis*, но более поздняя основа *tārakā-* ‘звезда’ имеет неожиданную баритонезу (а не окситонезу по контрасту). Соответственно, ударение в иранской праформе **stārčī-* ‘звезда’ нельзя проследить с уверенностью.

Все же видно, что форма pl. *éstēcān* образуется от варианта *éstēč* с начальным ударением.

15. *eštén*, мн. ч. *-ān* ‘козленок’, точнее – ‘козочка возрастом до одного года’ (Ефимов 2009: 194; IFL I: 237).

Продолжает праформу **fštānyā-* или **fštānī-*, окситонеза по контрасту с **fštāna-* ‘женская грудь’ (ЭСИЯ-3: 88–89). На окситонезу в имени женского рода от этой основы также указывают др.-инд. *viśvá-psnya* ‘alle Nahrung enthaltend, allnährend’ с древней нулевой ступенью, заменившейся в рассматриваемом иранском производном на продленную.

16. *g(‘)inō*, мн. ч. *g(‘)inōān* п., *g(‘)inān* ш., г. ‘волос, волосик’ (Ефимов 2009: 195).

Для основы **gawna-* может быть установлена окситонеза на основании др.-инд. *guṇā-* ‘нить; волокно’ и пшт. *ḡūṇay* ‘волос (на теле); кожа; пора’ < **gāwnaka-* с ударением по контрасту (ЭСИЯ-3: 240–245).

Возможно, форма *g(‘)inō* с ударением на первый слог является как раз продолжением производной баритонированной основы **gāwnaka-*. Соответственно, множественное число с ударением на основе может быть обусловлено изначальной баритонезой.

17. *gol*, мн. ч. *-ān* ‘цветок’ (Ефимов 2009: 196)

Пример из текста: *bedun-e ō golān-i, če ma yosānān bikāla hēn* ‘за исключением тех цветов, которые находятся в домах’ (Ефимов 2009: 139–140).

Персидское заимствование, ср. дари *gol* ‘цветок’ (Киселева-Миколайчик 1978: 608).

Исключение из правил для заимствованных односложных основ с исходом на согласный.

18. *gu*, мн. ч. *guān* п. ‘корова’, также *guḡān* (Ефимов 2009: 39, 196).

По всей видимости, данная основа имеет ударение на окончании под влиянием характерного для названий животных суффикса множественного числа *-ḡān*. Кроме того, можно думать об ударении на окончании основ с гласным исходом.

19. *yasó*, мн. ч. *yasá* п., *yasān* п. ‘теленки’ (Ефимов 2009: 197).

Этимологически связано с др.-инд. *vatsá-* ‘теленки, детеныш животного’. Если указывает на изначальную окситонезу, имеет закономерное ударение на окончании во множественном числе.

20. *haḍ*, мн. ч. *-ān* ‘кость’ (Ефимов 2009: 199).

Г. Моргенштерне зарегистрировал формы *haḍḍ*, *haṭ* и *hāṭ* (IFL I: 258).

Заимствование из индоарийских языков, ср. лахнда *haḍḍ* ‘кость’ (IFL I: 258; Turner 1966: 807).

Ударение незакономерно. Может быть связано с исходом на геминату.

21. *ǰā*, *ǰāy*, мн. ч. *-ān* ‘место, местность’ (Ефимов 2009: 205).

Заимствование из персидского, ср. дари *ǰāy* ‘место, местность’ (Киселева-Миколайчик 1978: 242).

Подчиняется правилу для заимствованных основ, заканчивающихся гласным.

22. *kala* (пачаганский диалект), *kal* (гочуланский диалект), *kalla*, мн. ч. *-ān* ‘голова’ (Ефимов 2009: 206).

Словарная статья противоречит данным в очерке и грамматике, т. е. случаю 16. *kalá*, pl. *kaláân* ‘голова’ из списка слов с безударным окончанием множественного числа *-ân*. Если именно такое место ударения верно, то оно закономерно для заимствованных основ с гласным исходом. Впрочем, имеющихся материалов недостаточно для выводов.

23. *kâla*, *kâlâ*, мн. ч. *-ân* ‘вещи, утварь; одежда, белье’ (Ефимов 2009: 208).

Персидское заимствование, ср. дари *kâlâ* ‘вещи, утварь; одежда, белье’ (Киселева-Миколайчик 1978: 575).

Подчиняется правилу для заимствованных основ, заканчивающихся гласным.

24. *kôr*, ш. *kôr*, мн. ч. *-ân* ‘осел’ (Ефимов 2009: 209).

Вокализм указывает на баритонезу **xâra*- ‘осел’, хотя пшт. *xra* ‘ослица’ (Асланов 1985: 375) косвенно указывает на окситонезу.

Наиболее значимое исключение, хотя его значение уменьшается из-за того, что невозможно определенно установить ударение в форме множественного числа.

25. *kôdak*, мн. ч. *-ân* ‘дитя, ребенок’ (Ефимов 2009: 210).

Персидское заимствование, ср. дари *kudak* ‘ребенок; дитя, младенец’ (Киселева-Миколайчик 1978: 575).

Безударное окончание для подобной двусложной основы характерно.

26. *kor*, мн. ч. *-ân* ‘пещера’ (Ефимов 2009: 210).

Этимология неясна. Возможно, видоизмененное дари *γâr* ‘пещера, грот’ (Киселева-Миколайчик 1978: 527).

Ударение незакономерно, одно из исключений.

27. *kory*, мн. ч. *-ân* ‘курица’ (Ефимов 2009: 210).

Унаследованное слово с праформой **kr̥ka-*, **kr̥kâ-* f., *kr̥kĩ-* f., для которого невозможно однозначно установить ударение, но орм. лог. *kerži*, *kirži*, орм. кан. *kirzi* ‘курица; куры’ (ЭСИЯ-4: 398–401) могут указывать на окситонезу.

28. *kuka* ш. ‘сапожный гвоздь’ (Ефимов 2009: 211), pl. М *kūkân* ‘железные гвозди’ (IFL I: 265).

Индоарийское заимствование, ср. лахнда *kōkā* ‘гвоздь, гвоздик’ (IFL I: 265; Turner 1966: 181). У Г. Моргенштерне ошибочно реконструирована форма единственного числа *kūk* (IFL I: 265).

Рассматриваемое слово закономерно имеет ударение на форманте в форме множественного числа.

29. *maktabi*, мн. ч. *-ân* ‘школьник’ (Ефимов 2009: 213).

Пример из текста: *maktabiyân-astan xô-pen bârta* ‘вели с собой детей-школьников’ (Ефимов 2009: 120).

Подчиняется правилу для многосложных заимствованных основ, заканчивающихся гласным.

30. *mardom*, мн. ч. *-ân* ‘люди, народ’ (Ефимов 2009: 214).

Персидское заимствование, ср. дари *mardom* ‘люди; народ’ (Киселева-Миколайчик 1978: 647). Пример из текста: *mardomân jârî* ‘люди сказали’ (Ефимов 2009: 141).

Вероятно, конечное ударение обусловлено конечным ударением самого слова; более того, множественное число в самом дари – *mardomân*.

31. *mâl* ‘имущество, добро; скот’ (мн. ч. *-ân*) (Ефимов 2009: 215).

Пример из текста: *mâlân-ê γory je bēn* ‘волк зарезал (/ убил) его скот’ (Ефимов 2009: 152). Персидское заимствование, ср. дари *mâl* ‘имущество, добро; товар; скот’ (Киселева-Миколайчик 1978: 627).

Слово подчиняется правилам для заимствованных односложных основ с согласным исходом и вокализмом *â*.

32. *mâlu*, мн. ч. *mâluγâr*, *mâwol*, мн. ч. *-ân* ‘дядя (со стороны матери)’ (Ефимов 2009: 215).

Пример из текста: *manân mâwolânâ byâ* ‘брат моих дядей’ (Ефимов 2009: 144).

Конечное ударение ожидаемо, особенно ввиду формы *mâluγâr*.

33. *mēr*, мн. ч. *-ân* (а также *mērîna*) ‘человек; мужчина’ (Ефимов 2009: 216).

Продолжает баритонированную праформу **mârtya-*, ср. др.-инд. *mârtya-*, *mârtya-* ‘смертный; человек’ (ЭСИЯ-5: 215–216). Но может отражать праиранский окситонированный вариант, ср. пшт. *mērô* ‘муж, супруг; храбрый, мужественный человек’ (Дыбо 1974: 87). Также не исключено влияние ударного варианта множественного числа *-îna* < пшт. *-îna* ?

34. *morčē*, мн. ч. *-ân* ‘воробей’ (Ефимов 2009: 218).

Должно восходить к **mṛga-čī-ká-* с окситонезой, но в таком случае ударение должно было быть на окончании. Исключение из правил.

35. *nērōkōsp*, мн. ч. *-ân* (Ефимов 2009: 220).

Ударение *-ân* вслед за *ōsprân* ‘кони’.

36. *panân*, мн. ч. *-ân* ‘дорога, путь’ (Ефимов 2009: 225).

Праир. **pantānam*. Закономерное ударение на основе при исходной баритонезе.

37. *pošta*, мн. ч. *-(w)ân* ‘холм’ (Ефимов 2009: 230).

Персидское заимствование, ср. дари *pošta* ‘торка, холм, курган; куча, груда, насыпь, вал, бугор’ (Киселева-Миколайчик 1978: 166).

Подчиняется правилу для двусложных заимствованных основ, заканчивающихся гласным.

38. *rafiq*, мн. ч. *-ân* ‘товарищ, приятель’ (Ефимов 2009: 231).

Арабо-персидское заимствование, ср. дари *rafiq*, мн. ч. *-rofaqâ* ‘товарищ, друг; спутник, попутчик’ (Киселева-Миколайчик 1978: 405).

Незакономерное ударение на окончании, исключение из правил.

39. *setâra*, мн. ч. *-ân* ‘звезда’ (Ефимов 2009: 235).

Персидское заимствование, ср. дари *setâra* ‘звезда, планета’ (Киселева-Миколайчик 1978: 438).

Подчиняется правилу для многосложных заимствованных основ, заканчивающихся гласным.

40. *šahr*, *šâr*, мн. ч. *-ân* ‘город; страна’ (Ефимов 2009: 237).

Персидское заимствование, ср. дари *šahr* ‘город’ (Киселева-Миколайчик 1978: 489).

Ударение на окончании неожиданно. Не исключено, что можно уподобить словам с колебанием в ударении, которые заканчиваются сочетанием согласных.

41. *waxt*, *waqt*, мн. ч. *-ân* ‘время, пора, промежуток времени’ (Ефимов 2009: 245).

Пример из текста: *howî waxtân-tar* ‘в те самые (былые) времена’ (Ефимов 2009: 150). Арабо-персидское заимствование, ср. дари *waqt* ‘время, промежуток времени; рано’ (Киселева-Миколайчик 1978: 724–725).

Ударение на окончании незакономерно. Не исключено, что можно уподобить словам с колебанием в ударении, которые заканчиваются сочетанием согласных, ср. 40. *šahr*, *šâr*.

42. *zâyu*, мн. ч. *zâγân*, *zâhâyu* п. ‘сын, дитя’ (Ефимов 2009: 252).

Унаследованная основа, праформа вроде **zâka-* реконструируется для согд. *z’k* ‘ребенок’, пшт. *zoγ* ‘сын; детеныш’, белуджск. *zâg* ‘сын’ (NEVP: 103). Нельзя установить, была ли данная праформа баритонированной или окситонированной. Возможно, праир. **zâkâ-*.

Рассмотрение менее надежных случаев показывает, что и они в большинстве своем подчиняются установленному распределению. Следует перечислить исключения для унаследованных и заимствованных основ.

Исключения среди унаследованных основ: *ōsp*, мн. ч. *ōspān/ asp, aspān* ‘лошадь’; *beā, byā, beār* ш., мн. ч. *beārā* и *beārān* ‘брат’; *kōr*, ш. *kōr*, мн. ч. *-ān* ‘осел’; *mēr*, мн. ч. *-ān*, а также *mērīna* (незакономерное ударное окончание множественного числа); *morčē*, мн. ч. *-ān* ‘воробей’ (незакономерная ударность основы во множественном числе).

Как минимум в трех случаях незакономерное ударение может объясняться перестройкой формы слова. Кроме того, не исключены ошибки в передаче места ударения в словарных статьях, не все случаи удается проверить.

Исключения среди заимствованных основ: *čiz*, мн. ч. *-ān* ‘предмет, вещь’; *čur* п., ш., *čur* п., ш., мн. ч. *-ān* ‘вор’; *dār, dār* ш., мн. ч. *-ān* ‘гора, холм’; *gol*, мн. ч. *-ān* ‘цветок’; *hađ*, мн. ч. *-ān* ‘кость’; *kor*, мн. ч. *-ān* ‘пещера’; *rafiq*, мн. ч. *-ān* ‘товарищ, приятель’; *šahr, šār*, мн. ч. *-ān* ‘город; страна’; *waxt, waqt*, мн. ч. *-ān* ‘время, пора, промежуток времени’ (незакономерное ударное окончание множественного числа).

Обращает на себя внимание, что исключения среди заимствованных слов заканчиваются в единственном числе на *-z, -r, -l*, нехарактерные для собственных слов парачи *-đ* и *-q* и на группы согласных. Непонятно, почему множественное число именно таких слов могло бы тяготеть к ударности окончания, но определенная закономерность, пусть и труднообъяснимая, прослеживается.

В заключение я повторяю формулировки правил ударения в формах множественном числе с окончанием *-ān*:

1. В унаследованных основах в зависимости от того, относилась ли она к баритонезе или окситонезе в праязыке, окончание *-ān* соответственно имеет безударный или ударный вариант. Это правило можно продемонстрировать следующей таблицей (некоторые сомнительные в отношении праязыкового ударения случаи не включены)⁶.

Баритонеза	Окситонеза
<i>dōst</i> , pl. <i>dōstān</i> ‘рука’ < * <i>dāsta-</i> / * <i>zāsta-</i>	<i>xi</i> , pl. <i>xiyān</i> ‘сестра’ < * <i>hwāhī-</i>
<i>bōr</i> , pl. <i>bōrān</i> ‘дверь’ < * <i>dwār-</i> / * <i>dwār-</i>	<i>jinj</i> ~ <i>jinč</i> , pl. <i>jinjān</i> ~ <i>jinčān</i> ‘женщина’ < * <i>janičī-</i>
<i>gu</i> , pl. <i>gūšān</i> ‘ухо’ < * <i>gāwša-</i>	<i>net</i> , pl. <i>netān</i> ‘внучка’ < * <i>naptī-</i>
<i>mox</i> , pl. <i>mōxān</i> ‘лицо; щека’ < * <i>mīxa-</i>	<i>poš</i> , pl. <i>pošān</i> ‘сын’ < * <i>putrá-</i>
<i>naḡōn</i> , pl. <i>naḡōnān</i> ‘хлеб’ < * <i>nikāna-</i>	<i>dot</i> , pl. <i>dotān</i> ‘дочь’ < * <i>duxtār-</i>
<i>guwān</i> , pl. <i>guwānān</i> ‘вымя’ < * <i>gaw-dāna-</i> или * <i>gaw-pāna-</i> ?	<i>boj</i> , pl. <i>bojān</i> ‘коза’ < * <i>būzā-</i>
<i>tōx</i> , pl. <i>tōxān</i> ‘козел’ < * <i>tāxa-</i> / * <i>tāka-</i> ?	<i>γos</i> , pl. <i>γosān</i> ~ <i>γosānān</i> ‘дом’ < * <i>wīś-</i> , pl. * <i>wīśānām</i>
<i>danān</i> , pl. <i>-ān</i> ‘зуб’ < * <i>dantāna-</i>	<i>bīn</i> , pl. <i>bīnān</i> ~ <i>bīnānān</i> ‘дерево’ < * <i>bṛzn(y)á-</i>
<i>panān</i> , pl. <i>-ān</i> ‘дорога, путь’ < * <i>pantānam</i>	<i>ešén</i> , pl. <i>-ān</i> ‘козочка возрастом до одного года’ < * <i>fštānyā-</i> или * <i>fštānī-</i>
	<i>γasó</i> , pl. <i>γasá</i> ~ <i>γasān</i> ‘теленочек’ < * <i>watsá-</i> ?

2. Во множественном числе вида *-ān* на основе имеют окончания одно- и двусложные заимствованные основы с исходом на согласный вне зависимости от места ударения.

3. Ударение на окончании *-ān* имеют одно-, дву- и трехсложные заимствованные и некоторые унаследованные основы, заканчивающиеся ударным гласным.

4. Колебания в месте ударения или ударение на окончании во множественном числе с формантом *-ān* имеют слова, которые заканчиваются в единственном числе на *-z, -r, -l*, нехарактерные для собственных слов парачи *-đ* и *-q* и на группы согласных.

⁶ Это основы *merzā*, pl. *merzān* ‘борона; лопатка, мастерок’, *bān*, pl. *bānān* ‘язык’, *a(y)ír*, pl. *a(y)írān* ‘облако’, *pā*, pl. *pān* ~ *pānān* ‘нога’ и т. д.

Эти правила показывают, что парачи, хотя и утратил различия баритонезы и окситонезы в большинстве непроизводных и производных имен в единственном числе, во множественном с формантом *-ān* последовательно сохраняет древнее распределение.

Литература

- Абаев, В. И. 1958–1995. *Историко-этимологический словарь осетинского языка. В 4-х томах.* Москва / Ленинград: Издательство АН СССР.
- Асланов, М. Г. 1985. *Афганско-русский словарь (пушту).* Москва: Советская энциклопедия.
- Дыбо, В. А. 1972. О рефлексах индоевропейского ударения в индоиранских языках. В сб.: С. Б. Бернштейн (ред.). *Конференция по сравнительно-исторической грамматике индоевропейских языков (12–14 декабря): Предварительные материалы.* Москва: Наука.
- Дыбо, В. А. 1974. Афганское ударение и его значение для индоевропейской и балтославянской акцентологии. I. Именная акцентуация. В сб.: Т. М. Судник (ред.). *Балтославянские исследования: 67–105.* Москва: Наука.
- Дыбо, В. А. 1989. Афганское ударение и его значение для индоевропейской и балтославянской акцентологии. II. Глагольная акцентуация. В сб.: Н. И. Толстой (ред.). *Славянский и балканский фольклор. Просодия: Сборник статей: 106–147.* Москва: Наука.
- Дыбо, В. А. 2007. Рефлексы индоевропейских слоговых долгих плавных и носовых в кельто-италийских языках (в защиту концепции Х. Педерсена). В сб.: И. С. Смирнов (ред.). *Аспекты компаративистики 2: 11–27.* Москва: Издательство РГГУ.
- Дыбо, В. А. 2014. Балто-славянская акцентная система и итоги индоевропейской акцентологической реконструкции. В сб.: *IWoBA VIII. Реферати VIII међународног скупа о балтословенској акцентологији (Славистички зборник. Нова серија, књига I): 17–94.* Нови Сад.
- Елизаренкова, Т. Я. 1982. *Грамматика ведийского языка.* Москва: Наука.
- Ефимов, В. А. 1979. О некоторых архаических чертах морфологической структуры презенса в ормури. В сб.: *Совещание по общим вопросам диалектологии и истории языка. (Душанбе. 12–15 ноября 1979 г.). Тезисы докладов и сообщений.* Москва: Издательство АН СССР.
- Ефимов, В. А. 1985. Ударение в языке ормури. В сб.: *Иранское языкознание. Ежегодник 1981: 31–64.* Москва: Наука.
- Ефимов, В. А. 1986. *Язык ормури в синхронном и историческом освещении.* Москва: Наука.
- Ефимов, В. А. 1997. Язык парачи. В сб.: В. А. Ефимов (ред.). *Основы иранского языкознания. Новоиранские языки. Северо-западная группа II: 419–560.* Москва: Наука.
- Ефимов, В. А. 2009. *Язык парачи.* Москва: Восточная литература.
- Киселева, Л. Н., В. И. Миколайчик. 1978. *Дари-русский словарь.* Москва: Русский язык.
- Стеблин-Каменский, И. М. 1999. *Этимологический словарь ваханского языка.* Санкт-Петербург: Петербургское востоковедение.

References

- Abaev, V. I. 1958–1995. *Istoriko-ètimologičeskij slovar' osetinskogo jazyka. V 4-x tomah [Historical-Etymological Dictionary of Ossetic. In 4 volumes].* Moskva-Leningrad: Izdatel'stvo AN SSSR.
- Aslanov, M. G. 1985. *Afgansko-russkij slovar (pushtu) [Afghan-Russian Dictionary (Pashto)].* Moskva: Sovetskaja enciklopedija.
- Cantera, Alberto. 2001. Die Behandlung der indogermanischen Lautfolge *(C)RHC- im Iranischen. *Münchener Studien zur Sprachwissenschaft* 61: 1–27.
- CDIAL = Turner, Ralph L. 1962–1985. *A comparative dictionary of Indo-Aryan languages.* London: Oxford University Press.
- Cheung, Johnny. 2002. *Studies in the historical development of the Ossetic vocalism.* Wiesbaden: Reichert.
- Cheung, Johnny. 2010. Selected Pashto problems. I: The accent in Pashto. *Persica* 23: 109–121.
- Cheung, Johnny. 2011. Selected Pashto problems II. Historical phonology 1: On vocalism and etyma. *Iran and Caucasus* 15: 169–205.

- Dybo, V. A. 1972. O refleksax indoevropskogo udarenija v indoiranskix jazykax. In: S. B. Bernshtein (ed.). *Konferencija po sravnitel'no-istoricheskoj grammatike indoevropskix jazykov (12–14 dekabnja): Predvaritel'nye materialy*. Moskva: Nauka.
- Dybo, V. A. 1974. Afganskoe udarenie i ego znachenie dl'a indoevropskoj i baltoslavjanskoj akcentologii. I. Imennaja akcentuacija. In: T. M. Sudnik (ed.). *Baltoslavjanskije issledovanija*: 67–105. Moskva: Nauka.
- Dybo, V. A. 1989. Afganskoe udarenie i ego znachenie dl'a indoevropskoj i baltoslavjanskoj akcentologii. II. Glagol'naja akcentuacija. In: N. I. Tolstoj (ed.). *Slav'anskij i balkanskij fol'klor. Prosodija: Sbornik statej*: 106–147. Moskva: Nauka.
- Dybo, V. A. 2007. Refleksy indoevropskix slogovyx dolgix plavnyx i nosovyx v kel'to-italijskix jazykax (V zaščitu koncepcii H. Pedersena). In: I. S. Smirnov (ed.). *Aspekty komparativistiki 2*: 11–27. Moskva: RSUH Publishers.
- Dybo, V. A. 2014. Balto-slav'anskaja akcentnaja sistema i itogi indoevropskoj akcentologičeskoj rekonstrukcii. In: *IWoBA VIII. Referati VIII međunarodnog skupa o baltoslovenskoj akcentologiji (Slavistichki zbornik. Nova serija, kn'iga I)*: 17–94. Novi Sad.
- Elizarenkova, T. Ya. 1982. *Grammatika vedijskogo jazyka*. Moskva: Nauka.
- Efimov, V. A. 1979. O nekotoryx arxaicheskix chertax morfoložičeskoj struktury prezensa v ormuri. In: *Soveshčanie po obščim voprosam dialektologii i istorii jazyka (Dushanbe. 12–15 nojabr'a 1979 g.)*. Tezisy dokladov i soobščenenij. Moskva: Izdatel'stvo AN SSSR.
- Efimov, V. A. 1985. Udarenie v jazyke ormuri. In: *Iranskoe jazykoznanie. Ezhegodnik-1981*: 31–64. Moskva: Nauka.
- Efimov, V. A. 1986. *Jazyk ormuri v sinxronnom i istoričeskom osvješčenii*. Moskva: Nauka.
- Efimov, V. A. 1997. Jazyk parachi. In: V. A. Efimov (ed.). *Osnovy iranskogo jazykoznanija. Novoiranskije jazyki. Severo-zapadnaja grupa II*: 419–560. Moskva: Nauka.
- Efimov, V. A. 2009. *Jazyk parachi*. Moskva: Vostochnaja literatura.
- ESIJ = Rastorgueva, V. S., D. I. Edel'man. 2000–. *Ètimologičeskij slovar' iranskix jazykov* [Etymological dictionary of the Iranian languages]. Moskva: Vostochnaja literatura RAN.
- EWAia = Mayrhofer, Manfred. 1992–2001. *Etymologisches Wörterbuch des Altindoarischen. I–III Bde*. Heidelberg: Universitätsverlag Winter.
- IEW = Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern: French & European Publications.
- IIFL I = Morgenstierne, Georg. 1929. *Indo-Iranian Frontier Languages. Vol. I: Parachi and Ormuri*. Oslo: Institut for Sammenlignende Kulturforskning.
- IIFL II = Morgenstierne, Georg. 1938. *Indo-Iranian Frontier Languages. Vol. II: Iranian Pamir languages (Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi)*. Oslo: Institut for Sammenlignende Kulturforskning.
- Kiseleva, L. N., V. I. Mikolajčik. 1978. *Dari-russkij slovar'*. Moskva: Russkij jazyk.
- Morgenstierne, Georg. 1942. Archaisms and innovations in Pashto Morphology. *Norsk Tidsskrift for Sprogvidenskap* 12: 88–114.
- Morgenstierne, Georg. 1973. Traces of Indo-European Accentuation in Pashto? *Norwegian Journal of Linguistics (formerly Norsk Tidsskrift for Sprogvidenskap)* 27: 61–65.
- Morgenstierne, Georg. 1983. Bemerkungen zum Wort-Akzent in den Gathas und im Paschto. *Münchener Studien zur Sprachwissenschaft* 42: 167–175.
- NEVP = Morgenstierne, Georg. 2003. *A new etymological vocabulary of Pashto. Compiled and edited by J. Elfenbein, D. N. MacKenzie and Nicholas Sims-Williams*. Wiesbaden: Reichert.
- Steblin-Kamenskij, I. M. 1999. *Ètimologičeskij slovar' vaxanskogo jazyka*. St.-Petersburg: Peterburgskoe vostokovedenie.
- Whitney, William Dwight. 1879. *Sanskrit Grammar Including Both the Classical Language and the Older Dialects of Veda and Brahmana*. Leipzig: Breitkopf und Härtel / Boston: Ginn, Heath & Company.

Artem Trofimov. Stress position of nominal plurals in *-ân* in the Parachi language and its connection with the accent in Proto-Iranian

It has been established that Parachi, a minor Iranian language, partially preserves reflexes of the Proto-Iranian accent. V. A. Efimov noticed a peculiar accentual distribution in Parachi for plural forms of nouns with the formant *-ân*: a number of such nouns show stress on the stem, while others prefer a stressed ending. This distribution is characteristic of both inherited and

borrowed nouns. In this paper, based on analysis of available Parachi data, I assume that this distribution corresponds to the features of Proto-Iranian: inherited stems reflecting Proto-Iranian barytona are root-stressed, while nouns continuing oxytona have plural forms with stress on the ending *-ān*, respectively. It is more difficult to establish rules that govern borrowings, although there is a tendency for words ending in a consonant to feature root stress in plural, while borrowed nouns ending in a stressed vowel obtain the stress in the plural ending *-ān*. My analysis allows to identify 9 inherited Parachi words with root stress on plural forms continuing old barytona, and 10 words with ending stress on plural forms continuing old oxytona. Thus, despite loss of any traces of Proto-Iranian stress in singular forms, Parachi nouns may still preserve them in corresponding plural formations.

Keywords: Parachi language, Proto-Iranian stress, oxytona, barytona, accentual distribution.

