

Российский государственный гуманитарный университет
Russian State University for the Humanities



RSUH/RGGU BULLETIN

№ 4 (14)

Academic Journal

Series:

Philology. Journal of Language Relationship

Moscow 2016

ВЕСТНИК РГГУ

№ 4 (14)

Научный журнал

Серия

«Филология. Вопросы языкового родства»

Москва 2016

Редакционный совет серий «Вестника РГГУ»

Е.И. Пивовар, чл.-кор. РАН, д-р ист. н., проф. (председатель)

Н.И. Архипова, д-р экон. н., проф. (РГГУ), А.Б. Безбородов, д-р ист. н., проф. (РГГУ), Х. Варгас (Ун-т Кали, Колумбия), А.Д. Воскресенский, д-р полит. н., проф. (МГИМО (У) МИД России), Е. Вятр (Варшавский ун-т, Польша), Дж. Дебарделебен (Карлтонский ун-т, Канада), В.А. Дыбо, акад. РАН, д-р филол. н. (РГГУ), В.И. Заботкина, д-р филол. н., проф. (РГГУ), В.В. Иванов, акад. РАН, д-р филол. н., проф. (РГГУ; Калифорнийский ун-т Лос-Анджелеса, США), Э. Камия (Ун-т Тачибана г. Киото, Япония), Ш. Карнер (Ин-т по изучению последствий войн им. Л. Больцмана, Австрия), С.М. Каштанов, чл.-кор. РАН, д-р ист. н., проф. (ИВИ РАН), В. Кейдан (Ун-т Карло Бо, Италия), Ш. Кечкемети (Национальная Школа Хартий, Сорбонна, Франция), И. Клюканов (Восточно-Вашингтонский ун-т, США), В.П. Козлов, чл.-кор. РАН, д-р ист. н., проф. (ВНИИДАД), М. Коул (Калифорнийский ун-т Сан-Диего, США), Е.Е. Кравцова, д-р психол. н., проф. (РГГУ), М. Крэмер (Гарвардский ун-т, США), А.П. Логунов, д-р ист. н., проф. (РГГУ), Д. Ломар (Ун-т Кельна, Германия), Б. Луайер (Ин-т геополитики, Париж-VIII, Франция), С. Масамичи (Ун-т Чуо, Япония), В.И. Молчанов, д-р филос. н., проф. (РГГУ), В.Н. Незамайкин, д-р экон. н., проф. (Финансовый ун-т при Правительстве РФ), П. Новак (Ун-т Белостока, Польша), Ю.С. Пивоваров, акад. РАН, д-р полит. н., проф. (ИНИОН РАН), Е. ван Поведская (Ун-т Сантьяго-де-Компостела, Испания), С. Рапич (Ун-т Вупперталь, Германия), М. Сакаи (Ун-т Чуо, Япония), И.С. Смирнов, канд. филол. н. (РГГУ), В.А. Тишков, акад. РАН, д-р ист. н., проф. (ИЭА РАН), Ж.Т. Тощенко, чл.-кор. РАН, д-р филос. н., проф. (РГГУ), Д. Фоглесонг (Ун-т Ратгерс, США), И. Фолтыс (Политехнический ин-т г. Ополе, Польша), Т.И. Хорхордина, д-р ист. н., проф. (РГГУ), А.О. Чубарьян, акад. РАН, д-р ист. н., проф. (ИВИ РАН), Т.А. Шаклеина, д-р полит. н., проф. (МГИМО (У) МИД России), П.П. Шкаренков, д-р ист. н., проф. (РГГУ)

Серия «Филология. Вопросы языкового родства»

Редакционная коллегия серии

В.А. Дыбо, гл. ред., акад. РАН, д-р филол. н. (ИнСлав РАН / РГГУ), Г.С. Старостин, зам. гл. ред., канд. филол. н., доц. (РГГУ), Т.А. Михайлова, отв. сек., д-р филол. н., проф. (МГУ им. М.В. Ломоносова), А.В. Дыбо, д-р филол. н., член-корр. РАН (ИЯз РАН), С.В. Кулланда, канд. ист. н. (ИВ РАН), М.А. Молина (ИЯз РАН), И.С. Якубович, д-р филол. н. (Марбургский университет имени Филиппа)

Ответственные за выпуск: Г.С. Старостин, М.А. Молина

УДК 800(05)
ББК 80/84я5

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2016. — № 4(14). — x + 76 с. — (Вестник РГГУ. Серия «Филология. Вопросы языкового родства»: Научный журнал).

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, 2016. — No. 4(14). — x + 76 p. — (RSUH/RGGU Bulletin. Series: Philology. Journal of Language Relationship: Academic Journal).

ISSN 2073-6320

<http://www.jolr.ru/>
journal@jolr.ru

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

Подписано в печать 01.12.2016. Формат 60×90/8.

Бум. офсетная.

Печать офсетная. Тираж 1050 экз.

Заказ №143

Издательский центр
Российского государственного гуманитарного университета
125993, Москва, Миусская пл., 6
www.rgggu.ru
www.knigirgggu.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 I. Kogan. Genealogical classification of New Indo-Aryan languages and lexicostatistics</i>	227
[<i>A. И. Коган. Генеалогическая классификация индо-арийских языков и лексикостатистика</i>]	
<i>М. Е. Васильев, М. Н. Саенко. К вопросу о точности глоттохронологии: датирование процесса лексических замен по данным романских языков</i>	259
[<i>Mikhail E. Vasilyev, Mikhail N. Saenko. How accurate glottochronology can be? Dating the lexical replacement process in the Romance languages</i>]	
<i>И. С. Якубович. Славянский чертежник: этимология слав. *čьrtь ‘черт’</i>	279
[<i>Ilya S. Yakubovich. The Slavic draughtsman: etymology of Slav. *čьrtь ‘devil’</i>]	
<i>Mikhail A. Zhirov. The origin of Khanty retroflex nasal</i>	293
[<i>М. А. Живлов. Происхождение хантыйского ретрофлексного носового</i>]	

Contributors

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

Mikhail N. Saenko — candidate of sciences (Philology), junior researcher, Institute of Slavic studies of the Russian Academy of Sciences, Moscow, veraetatis@yandex.ru

Mikhail E. Vasilyev — researcher, School for Advanced Studies in the Humanities, Russian Presidential Academy of National Economy and Public Administration (Moscow), mvhumanity@gmail.com

Ilya S. Yakubovich — doctor of sciences (Philology), Philipps-Universität Marburg; Russian State University for the Humanities; researcher, sogdiana783@gmail.com

Mikhail A. Zhivlov — candidate of sciences (Philology), senior lecturer, Institute for Oriental and Classical Studies, Russian State University for the Humanities; researcher, School for Advanced Studies in the Humanities, RANEPА (Moscow), zhivlov@gmail.com

Сведения об авторах

Васильев, Михаил Евгеньевич — науч. сотрудник ШАГИ Российской академии народного хозяйства (Москва), mvhumanity@gmail.com

Живлов, Михаил Александрович — канд. филол. наук, ст. преп. Центра компаративистики ИВКА РГГУ, науч. сотрудник ШАГИ РАНХиГС (Москва), zhivlov@gmail.com

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

Саенко, Михаил Николаевич — канд. филол. наук, мл. науч. сотрудник Института славяноведения РАН (Москва), veraetatis@yandex.ru

Якубович, Илья Сергеевич — доктор филол. наук, Марбургский университет имени Филиппа; Российский государственный гуманитарный университет (Москва), sogdiana783@gmail.com

Note for Contributors

Journal of Language Relationship welcomes submissions from everyone specializing in comparative-historical linguistics and related disciplines, in the form of original articles as well as reviews of recent publications. All such submissions should be sent to the managing editor:

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

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

For more detailed guidelines on article submission and editorial policies, please see our website at: <http://www.jolr.ru> or address the editorial staff directly at journal@nostratic.ru.

Будущим авторам

Журнал *Вопросы языкового родства* принимает заявки на публикацию оригинальных научных статей, а также рецензий от всех, кто специализируется в области сравнительно-исторического языкознания и смежных дисциплин. Рукописи можно высылать непосредственно заместителю главного редактора по адресу:

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

Предпочтительные языки публикации — английский или русский, хотя возможна также публикация статей на других европейских языках (французский, немецкий и т. п.). К каждой статье обязательно прикладывается резюме (не более 300 слов) и список ключевых слов.

Подробнее о требованиях к оформлению рукописи, редакционной политике журнала и т. п. вы можете узнать на нашем сайте по адресу: <http://www.jolr.ru> или же непосредственно, обратившись к редакции по электронной почте (journal@nostratic.ru).

Genealogical classification of New Indo-Aryan languages and lexicostatistics

Genetic relations among Indo-Aryan languages are still unclear. Existing classifications are often intuitive and do not rest upon rigorous criteria. In the present article an attempt is made to create a classification of New Indo-Aryan languages, based on up-to-date lexicostatistical data. The comparative analysis of the resulting genealogical tree and traditional classifications allows the author to draw conclusions about the most probable genealogy of the Indo-Aryan languages.

Keywords: Indo-Aryan languages, language classification, lexicostatistics, glottochronology.

The Indo-Aryan group is one of the few groups of Indo-European languages, if not the only one, for which no classification based on rigorous genetic criteria has been suggested thus far. The cause of such a situation is neither lack of data, nor even the low level of its historical interpretation, but rather the existence of certain prejudices which are widespread among Indologists. One of them is the belief that real genetic relations between the Indo-Aryan languages cannot be clarified because these languages form a dialect continuum. Such an argument can hardly seem convincing to a comparative linguist, since dialect continuum is by no means a unique phenomenon: it is characteristic of many regions, including those where Indo-European languages are spoken, e.g. the Slavic and Romance-speaking areas. Since genealogical classifications of Slavic and Romance languages do exist, there is no reason to believe that the taxonomy of Indo-Aryan languages cannot be established.

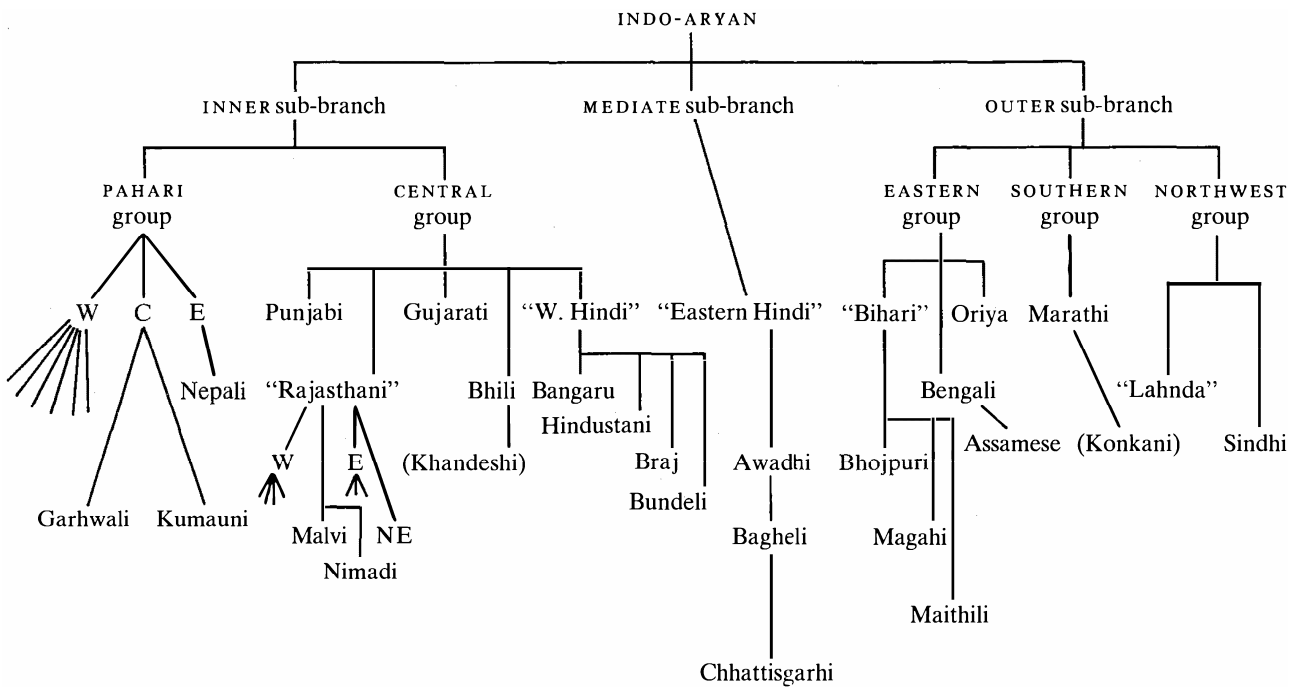
However, this scholarly pessimism does have some grounds. Certain approaches and methods used nowadays in Indological historical linguistics have proven to be inefficient, and without changes in research methodology, significant progress in the genetic classification of the Indo-Aryan languages is hardly possible. This issue will be discussed at some length below.

In the past Indologists had more than once attempted to classify the languages that they studied. In the late 19th and 20th centuries several alternative classifications were suggested. Since some of them still appear in modern Indological publications, I feel it necessary to dwell on them here. The scholar who seems to have been the first one to approach this problem is Rudolf Hoernle. He hypothesized that the Aryans migrated to the Indian subcontinent in two successive waves. The migrants belonging to these waves spoke two different dialects of Old Indo-Aryan, which he called Magadhi and Sauraseni (Hoernle 1880).¹ Magadhi, according to him, was the common ancestor of modern languages spoken in the South and East of the Indo-Aryan speaking area, i.e. of Marathi, Konkani, Bengali, Oriya, and Bihari dialects,² whereas Sauraseni was considered to constitute the protolanguage for the forms of speech current in

¹ These names should not be confused with the identical names of literary Prakrits (Masica 1991: 447).

² Hoernle preferred to group these dialects under the name of Eastern Hindi.

Figure 1. Genealogical classification of Indo-Aryan languages according to G. Grierson (adapted from Masica 1991: 449).



the North and West, i.e. for Nepali, Garhwali, Kumauni, Gujarati, Sindhi, Punjabi including Multani, and Western Hindi including Rajasthani.³ Hoernle was of the opinion that the speakers of “Magadhi” once occupied the entire North India but were later pushed back by “Sauraseni” speakers, and it is for this reason that the languages of the Northwest still possess certain vestigial features that are common with the southeastern languages.

Hoernle’s idea of a two-wave migration was taken over and further developed by George A. Grierson — a British scholar, the main author of the famous “Linguistic Survey of India”. His classification of Indo-Aryan languages included two main subbranches, which he called Inner and Outer, although they did not precisely correspond to Hoernle’s Sauraseni and Magadhi respectively. The main point of divergence between Grierson’s and Hoernle’s models was the position of the languages spoken in the Northwest of the subcontinent, namely Sindhi and the dialects of Western Punjab including Multani.⁴ Grierson preferred to include these forms of speech into the Outer subbranch, i.e. to group them together with Marathi, Konkani and the languages of Eastern India (Grierson 1927). Grierson’s arguments in favor of this point of view will be discussed separately below. Moreover, alongside the main subbranches embracing the bulk of the Indo-Aryan languages, Grierson’s classificatory scheme also featured a third one: it was called Mediate and included certain Hindi dialects, the most important of which is Awadhi.⁵ These dialects, according to Grierson, possessed both “Inner” and “Outer” features. Grierson’s classification is reproduced on Figure 1 in the form of a genealogical tree.

³ To this group Hoernle also added Pashto and Kashmiri, which he reckoned among the Indo-Aryan languages.

⁴ These dialects, spoken in the vast area covering northern and western parts of the present-day Pakistani province of Punjab, were grouped by Grierson under the name of Lahnda. This term is still rather popular among the Indologists.

⁵ The other two are Bagheli and Chhattisgarhi.

As we can see, both Inner and Outer languages are subdivided into further groups, i.e. Pahari,⁶ Central, Eastern, Southern and Northwestern. Grierson gave no clear-cut reasons for the postulation of these groups or the existence of such entities as “Lahnda”, “Western Hindi” or “Rajasthani”. The two main subbranches, however, were established by him on the basis of features which he thought to be diagnostic for classification. The most significant of them are:

- 1) retention of MIA *s* (< OIA *s*, *ś*, *ṣ*) in the Inner languages vs. change of this sibilant into other phonemes in the Outer subbranch;
- 2) loss of the final short vowels in the Inner subbranch vs. their preservation in the Outer languages;
- 3) the use of the suffix *-i-* to form verbal perfect stems in the Inner languages vs. the formation of such stems in the Outer languages with the suffix *-l-*;
- 4) the analytic typology of the Inner languages vs. the synthetic character of the Outer sub-branch.

Upon close examination, none of the above arguments can be considered valid. The loss of final short vowels in a number of Indo-Aryan languages took place in the New Indo-Aryan period and thus has nothing to do with dialectal differences in Old Indo-Aryan. The same holds true for the formation of perfect stems. The use of the suffix *-l-* for such a purpose is a comparatively recent phenomenon; moreover, it is not characteristic of all the Outer languages and is found in some Inner ones (e.g. in Gujarati). The development of Old Indo-Aryan sibilants was totally different in the East and Northwest of the subcontinent. In the East all the three sibilant phonemes have merged into one, i.e. *ś*. This reflex is already attested in the Magadhi Prakrit and found, e.g., in present day Bengali. In almost all the other Indo-Aryan languages, including Northwestern ones, the sibilants merged into *s*,⁷ sometimes with subsequent phonetic changes in certain positions. This means that reconstructing something like a “Common Outer” or “Proto-Outer” development of the Old Indian sibilant system is simply out of the question.

As for the typological argument, it is generally accepted among comparative linguists that such arguments are not relevant for genealogical classification. In addition, Grierson’s statement that all the Outer languages belong to the same synthetic morphological type is not fully correct. In reality, Eastern Indo-Aryan languages possess agglutinative morphology of secondary origin, which developed from the older analytic system as a result of the transformation of function words (e.g. postpositions) into affixes, whereas Northwestern languages are mainly analytic, although sometimes they preserve a few vestiges of old inflection.

Grierson’s arguments were analyzed by the Indian historical linguist and philologist Suniti Kumar Chatterji. In the introduction to his renowned work “Origin and development of the Bengali language” (Chatterji 1926), he managed to convincingly show their invalidity, as well as the incorrectness of the Inner-Outer model. As an alternative, he suggested his own classification, which is in certain respects similar to that of Grierson but without the Inner and Outer sub-branches as separate taxa.⁸ The reality of the Mediate subbranch was also denied. All the dialects that Grierson classified as Mediate were included by Chatterji into the Eastern

⁶ The Pahari group includes Indo-Aryan languages spoken in the sub-Himalayan region stretching from Nepal in the Southeast to the southern areas of the Indian state of Jammu and Kashmir in the Northwest.

⁷ Exceptions include several Pahari languages and Romany, which distinguish between two sibilants, i.e. *s* and *ś*. The latter reflects both *ś* and *ṣ* of Old Indo-Aryan.

⁸ Chatterji’s classificatory scheme was published in his abovementioned book as a part of the table illustrating the development of Aryan speech in India, see Chatterji 1926: 6.

group. Northwestern, Pahari,⁹ Southern, and Eastern groups were considered on the same taxonomic level. Moreover, Chatterji postulated two more subbranches, namely, Southwestern (including Gujarati and Rajasthani) and a subbranch consisting of Sinhalese and Maldivian (Dhivehi).¹⁰ The list of languages in some groups was somewhat different from the one offered by Grierson. Thus, both Punjabi and Romany were classified with the Northwestern group. The Central subbranch, called Midland by Chatterji, included only the forms of speech traditionally grouped together under the name of Western Hindi, i.e. standard Hindi and Urdu (with Khariboli as their common dialectal basis), Haryanvi, Braj, Kannauji and Bundeli. While working out this classification, Chatterji largely based himself on intuition. Only in relatively rare cases certain historical-phonological isoglosses were taken into account. Among the cited innovations one can mention the merger of the Old Indian sibilants into *ś* in the East or the simplification of the Middle Indo-Aryan geminates followed by compensatory lengthening of the preceding vowels in the majority of New Indo-Aryan languages vs. the absence of such a process in the Northwest, i.e. in Sindhi, Lahnda and Punjabi.

Both Grierson's and Chatterji's schemes can even nowadays be occasionally found in Indological linguistic literature. They coexist with several alternative classifications, which were suggested during the last five decades. The latter, however, differ only slightly from Chatterji's¹¹ and, likewise, remain mainly intuitive.

The 20th century saw great progress in the study of Indo-Aryan historical phonology. In addition to the already mentioned monograph on Bengali by Chatterji, Jules Bloch's book on Marathi (Bloch 1920) and R.L. Turner's works on Gujarati, Sindhi, Romany and Nepali (Turner 1921a; 1921b 1924; 1926; 1931) significantly extended our knowledge of sound change in a number of New Indo-Aryan languages. This research prompted some scholars to suggest a genealogical tree based on historical phonological isoglosses. Soon, however, it became clear that this task is extremely difficult, if not impossible. Isoglosses are sometimes easily detectable, but usually they cannot be brought together into bundles that would be peculiar for a particular language or language group. For example, the above-mentioned compensatory lengthening of short vowels before simplified geminates is characteristic of such languages as Hindi-Urdu, Gujarati, Marathi, Bengali and many others, but not of Punjabi, Lahnda or Sindhi. This fact makes it tempting to postulate two groups, one of which would include the three last idioms and the other would include all the rest. According to such a classification, Hindi-Urdu must be declared a language more closely related to Bengali and Gujarati than to Punjabi. But if we take another feature, e.g. the development of the Old Indic sibilants, the picture will be quite different. Bengali, where, as noted above, these phonemes have merged into *ś*, would form a separate group together with Assamese, for which "*ś*-reflexation" can be traced historically, whereas Hindi-Urdu will find itself closer to Punjabi, Lahnda, Sindhi and Gujarati, where the reflex is *s*. Gujarati, however, is not affected by another isogloss, which is common for many Indo-Aryan languages, namely the lenition of intervocalic *-m-* into *-v-* (sometimes with a further change into *ũ* and nasalization of the preceding vowel). In this respect, Hindi-Urdu shows close affinity to Marathi, Sindhi, Punjabi, Nepali, and Romany but not to Gujarati (cf. Hindi-Urdu *gāv*, Marathi *gāv*, Sindhi *gāu*, Punjabi *girāū*, Nepali *gāū*, Romany *gav*, but Gujarati *gām* 'village' < OIA *grāma-*).

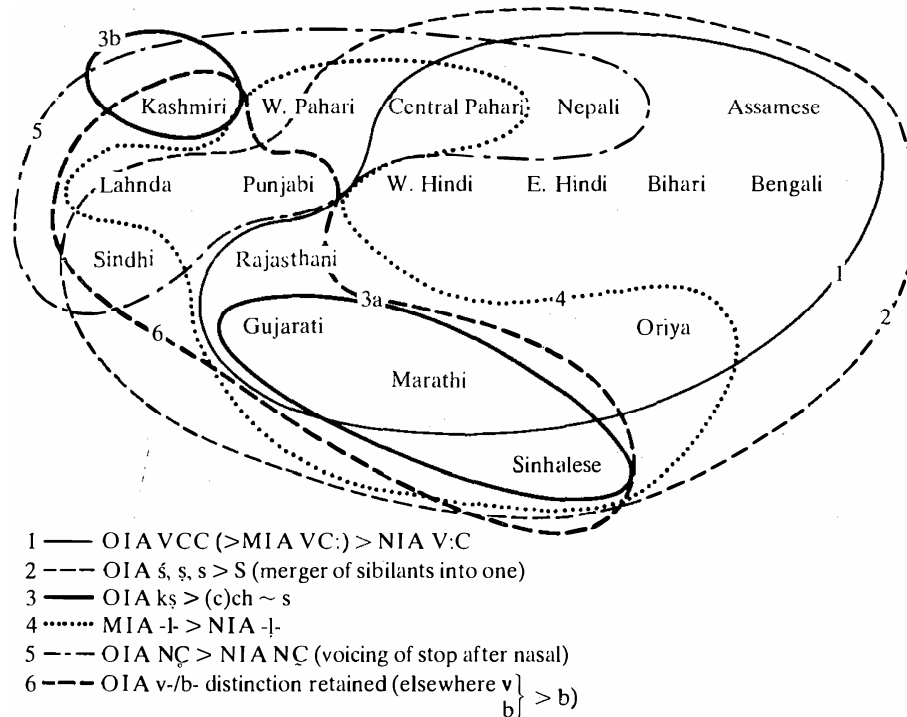
This complex situation, where different historical-phonological isoglosses are in "conflict" with each other, was tackled by Colin Masica in his book (Masica 1991). He analyzed the geographical distribution of six historical phonological features, i.e. compensatory lengthening,

⁹ For Grierson's Pahari languages, Chatterji prefers the term "North" or "Himalayan".

¹⁰ The latter two languages were ignored in Grierson's scheme.

¹¹ These classifications are cited, e.g. by Colin Masica in Masica 1991: 454–456.

Figure 2. The distribution of some basic New Indo-Aryan phonological isoglosses (adapted from Masica 1991: 459).



merger of the three OIA sibilants, OIA $k\acute{s} > (c)ch (> s)$, cerebralization of intervocalic MIA $-l-$, voicing of voiceless stops after nasals, retention of OIA initial $v-$ (elsewhere changed to b). The results of his analysis were presented in the form of the scheme reproduced on Figure 2.

As can be plainly seen, the above data make it hardly possible to postulate a group of languages (i.e. a sub-group of Indo-Aryan) on the basis of more than one isogloss that would be peculiar only for this group, as opposed to all the rest.

There can be no doubt that such a state of things is indeed the result of intensive and long-lasting language contact, which in South Asia was and is still often facilitated by absence of natural barriers. The gradual spread of certain contact-induced phonological features can be traced back to Middle Indo-Aryan. For example, OIA intervocalic stops (other than cerebrals) were dropped in the Maharashtri Prakrit, while still being preserved in Śauraseni. But later on the disappearance of stops in the intervocalic position also affected the Śauraseni area, i.e. the central part of the Indo-Gangetic Plain. It is attested already in Apabhraṃśa, a late Middle Indo-Aryan literary language formed in this very region, and is characteristic of almost all the local New Indo-Aryan forms of speech. The merger of OIA sibilants in many areas of the Indian subcontinent likewise dates back to the Middle Indo-Aryan period. Thus, in the language of Aśoka's inscriptions from most parts of India no distinction was made between the reflexes of OIA s, \acute{s} and \grave{s} . This threefold contrast was fully retained only in the inscriptions from the Northwest, which means that in Aśoka's time, i.e. in the 3rd century BC, the isogloss in question did not affect this area. Later, however, the sibilants did merge in the Northwest as well. In the New Indo-Aryan dialects spoken there nowadays the results of this process can be clearly seen (cf. Lahnda *sap(p)* 'snake' < OIA *sarpa-*; *sad(d)* 'call, shout' < OIA *śabda-* 'articulate sound, noise'; *solā* '16' < OIA *ṣoḍaśa*; *ghāh* 'grass' < OIA *ghāsa-* 'food, pasture grass'; *dāh* '10' < OIA *daśa*; *nōh* 'son's wife' < OIA *suṣā-*).

Such a situation can be best represented within the framework of Schmidt's wave model. For this reason, many scholars believe that the latter model is the most preferable, if not the

only possible, for the Indo-Aryan group, whereas the tree model is not applicable to it. This pessimism was perhaps most vividly expressed by C. Masica in his previously mentioned book: “We might therefore be well-advised to give up as vain the quest for a final and “correct” NIA historical taxonomy, which no amount of tinkering can achieve, and concentrate instead on working out the history of various features, letting such feature-specific historical groupings emerge as they may, with their overall non-coincidence as testimonial to the complexity of the situation” (Masica 1991: 460). Masica’s practical suggestion was that scholars should confine themselves to drawing isoglosses on the map and postulating zones that they demarcate. Since such zones usually overlap, the term “overlapping genetic zones” was accepted for them. Such a term, however, is patently self-contradictory. ‘Overlapping’ implies that there must be a language or languages belonging to more than one zone, which can never be the case with genetic subdivisions. Moreover, it is well known that areal and genealogical groupings do not always coincide, and for this reason, the use of the term ‘zone’ in the genetic classification does not seem to be correct. Nevertheless, the proposed term became very popular among Indologists, and has even penetrated into some encyclopaedic editions.

It is justified to state that genealogical classification of the Indo-Aryan languages has presently been substituted by the areal one, and that the main reason for this paradigm change is the failure to classify the languages in question on the basis of phonological innovations. It should be noted in this connection that phonological isoglosses, as they are traditionally postulated, are not always unquestionable. Sometimes certain sound changes, which are considered to be common for many languages, in reality coincide only partially. This seems to hold true, e.g., for the simplification of the MIA geminates with compensatory lengthening of the preceding short vowels. In quite a number of Indo-Aryan dialects this development takes place in all words and positions, whereas in standard Hindi-Urdu it apparently affects mainly monosyllabic words. If the number of syllables is more than one, geminates are often retained and the vowel remains short. This hypothetical rule is most likely to be valid when the MIA vowel is *a*: *pakkā* ‘ripe, mature’ < MIA *pakka-* < OIA *pakva-*, cf. Nepali *pāko*, Bengali *paka* (*a* < *ā*), Romany *pako* (*a* < *ā*), Gujarati *pākū* ‘ripe’; *makkhī* ‘fly’ < MIA *makkhiā-* < OIA *makṣikā-*, cf. Nepali, Kumauni *mākho*, Assamese *mākhi*, Gujarati *mākhi*, Romany *makh*, *maki*; *saccā* ‘true’ < MIA *sacca-* < OIA *satya-* ‘truth’, cf. Nepali, Kumauni *sāco*, Bengali *śāca*, Awadhi *sācu*, Marwari *sāco*, Gujarati *sācū*, Romany *čačo*; *acchā* ‘good’ < MIA *accha-* ‘clear, transparent, pure, clean’ < OIA *accha-* ‘clear, transparent’, cf. Oriya *āchā*, Kumauni *ācho* ‘good’, Gujarati *āchū* ‘thin, elegant’; *pattā* ‘leaf’ < OIA *pattra-*, MIA *patta-*, cf. Kumauni *pātī* ‘leaves, letter’, Nepali *pāto* ‘page, blade of a knife’, Bengali *pata* ‘leaf, blade’, Awadhi *pātā*, Gujarati *pātū* ‘leaf’; *patthar* ‘stone’ < MIA *patthara-* id. < OIA *prastara-* ‘anything strewn; flat surface; rock, stone’, cf. Awadhi, Kumauni *pāthar*, Bengali, Assamese *pathor*, Konkani *phāttaru* ‘stone’, Marathi *pāthar* ‘flat stone’; *apnā* ‘one’s own’ < MIA *appaṇaya-* < OIA **ātmanaka-*,¹² cf. Kumauni *āpṇo*, Nepali *āphnu*, Bengali *apṇo*, Gujarati *āpṇū*. Note that new Indo-Aryan disyllables ending in a vowel most probably reflect Old Indian trisyllabic bases enlarged with the suffix *-k-*,¹³ i.e. *pakvaka-*, *satyaka-*, *acchaka-*, *patraka-* etc.

Of particular interest are those cases where we find two cognates in Hindi-Urdu, one of which is monosyllabic and the other is di- or trisyllabic. Such cognates always show a divergent phonetic development: *sāt* ‘7’ < MIA *satta* < OIA *sapta* vs. *sattā* ‘aggregate of 7; seven in

¹² See Turner 1966: 51.

¹³ Cf. the above-cited words for ‘fly’, for which the prototype with the suffix *-k-* (*makṣikā-*) is attested already in Old Indian. On enlarged noun bases in *-k-* see in general (Bloch 1965: 111, 163–165).

cards' < MIA *sattaya-* < NIA *saptaka-*; *hāth* 'hand' < MIA *hattha-* < OIA *hasta-* vs. *hatthā* 'handle' < OIA *hastaka-*; *lāj* 'shame' < MIA, OIA *lajjā-* vs. *nilajā* 'shameless' < MIA *nilajja-* < OIA *nirlajja-*; *kām* 'work, act' < MIA *kamma-* < OIA *karman-* vs. *nikammā* 'idle; useless, good-for-nothing' < MIA **nikkamma-*¹⁴ < OIA *niṣkarman-* 'inactive'.

The phenomenon described here certainly needs further study, because there are some unexplained counterexamples (cf., e.g. *māthā* 'forehead' < MIA *mattha(ya)-* 'head' < OIA *masta(ka)-* 'head, skull'¹⁵), but what can be stated with certainty is that the issue of common historical phonological isoglosses is much more complicated than it might seem upon first sight. Establishing such isoglosses for the Indo-Aryan group is possible only after a detailed and in-depth analysis of all available data. Since such analysis has not always been properly conducted, it is still premature to say that we know the full picture.

This, however, can hardly affect the conclusion that intensive language contact has sometimes made it almost impossible to distinguish between phonological innovations common for a genetic subgroup and contact-driven sound changes. It means that Indo-Aryan languages should not be classified based only (or even mainly) on historical phonology, as was frequently done in the past. But what kind of linguistic data should we then use as criteria for classification? It is tempting to turn to morphology, but, as C. Masica points out in his book, "...morphological criteria conflict just as much as phonological criteria" (Masica 1991: 460). Since syntax is even more prone to radical restructuring due to foreign influence and is, moreover, very similar throughout Indo-Aryan, syntactic data can hardly help us to clarify genetic relations within the Indo-Aryan group. The only domain of language that can provide us with relevant information for genealogical classification appears to be the lexicon.

To the best of my knowledge, no scholar has so far seriously attempted to create a genealogical tree of Indo-Aryan languages based on lexical isoglosses. Such a state of affairs arguably results from insufficient attention paid by many Indologists to the lexical level in general and basic vocabulary in particular. This is rather unfortunate, because the in-depth study of this part of the lexicon actually helps to solve a variety of problems of Indological comparative linguistics, including those of historical phonology, because it is well known that in basic vocabulary the number of loanwords is always limited and genuine phonetic development always predominates. The latter fact also suggests the possibility of using the lexicostatistical method for classifying the Indo-Aryan languages.

The 100-item Swadesh wordlist has on many occasions been successfully used as a sample of basic vocabulary. The greater part of its items is cross-linguistically stable, and cases of phonological change that are characteristic of borrowings from closely related languages are therefore always in a minority within this set. Since the historical phonology of many Indo-Aryan languages has now been studied in sufficient detail, such cases must often be easily detectable, even if one factors in the limitations of our knowledge stated above. It means that the problem of unidentified loanwords in Indo-Aryan wordlists is hardly crucial, and the resulting genealogical tree is unlikely to differ to a great extent from the real picture.

Consequently, in the present article we present an attempt of genealogical classification based on lexicostatistics. The Indo-Aryan lexicostatistical database, prepared by myself,¹⁶

¹⁴ Cf. Pali *nikkamma-*, Prakrit *ṅikkamma-* 'unoccupied' (Turner 1966: 422).

¹⁵ It should also be noted that the compensatory lengthening of MIA *i* and *u* appears not to be confined to monosyllables: *sīdhā* 'straightforward' < MIA, OIA *siddha-* 'perfected'; *sūkhā* 'dry' < MIA *sukha-* < OIA *śuṣka-*.

¹⁶ I wish to thank Anastasiya Krylova and Eugenia Renkovskaya (Russian State University for the Humanities, Moscow) for their help in the preparation of the database and Ilya S. Yakubovich (Moscow State University) for providing me with dictionaries of several New Indo-Aryan languages.

consists of Swadesh lists for 35 languages, namely Hindi-Urdu, Dakhini,¹⁷ Punjabi, Pothohari, Hindko, Gojri, Dogri, Lahnda (Multani), Sindhi, Kutchi, Rajasthani (Marwari), Gujarati, Marathi, Konkani, Bengali, Assamese, Oriya, Nepali, Sinhalese, Maldivian (Dhivehi), Kotgarhi, Himachali, Kului, Mandeali,¹⁸ Kumauni, Garhwali, Awadhi, Braj, Mewati,¹⁹ Wagdi,²⁰ Banjari, Maithili, Parya, Domaaki (Dumaki), and Romany. Unfortunately, it turned out to be impossible to include Old and Middle Indo-Aryan wordlists in the database, because they are either not securely datable (e.g., Vedic and Pali wordlists) or contain too many lacunae (e.g., wordlists of Aśokan Prakrits). Moreover, certain Middle Indo-Aryan languages, such as literary Prakrits and literary Apabhraṃśas, are to a great extent artificial constructs and do not fully reflect spoken dialects of their time. The wordlists are given in a special appendix after the main text of the article. Before proceeding to the results of the lexicostatistical calculations, it seems necessary to make some remarks concerning synonyms and loanwords.

Since in many cases we know little or nothing either about semantic nuances or frequency of a particular Indo-Aryan word on the list, it is sometimes impossible to determine the main synonym. In such a situation, we suggest that the best solution to the problem of synonymy is apparently to include no more than two synonyms on the list in the case when each of them has cognates in other Indo-Aryan languages. If only one of the synonyms finds etymological parallels within the group, it is technically considered as the main one. Likewise, in those cases where both an inherited word and a loanword are attested for the same Swadesh meaning, only the former is included in the database (since addition or omission of the latter will be irrelevant for the lexicostatistical results anyway).

Loanwords on the lists are for the most part easily identifiable. Usually they are of either Persian or Sanskrit origin. The latter group embraces not only *tatsamas* (borrowings from Sanskrit, preserved more or less unchanged in modern languages), but also the so-called *semi-tatsama* or *ardhatatsama* words, i.e. early Sanskrit loans which have undergone certain phonetic changes (e.g., Punjabi *purakh* ‘man’ < Skr. *puruṣa-*; Dogri, Himachali, Mandeali, Awadhi *barkhā* ‘rain’ < Skr. *varṣā-*). Dravidian loanwords are found chiefly in Sinhalese and Konkani wordlists.²¹ The noun *poṭ* ‘belly’ seems to be a borrowing from Dravidian in Marathi and Konkani (cf. Proto-Dravidian **poṭ* and its reflexes in different Dravidian languages given in (Burrow, Emeneau 1961: 397–398)). Phonetically similar words for ‘belly’ in many other Indo-Aryan languages (Hindi-Urdu, Punjabi, Parya, Gujarati, Bengali *peṭ*, Assamese *pet*, Oriya *peṭa*, Garhwali *pyaṭ*, Romany *perr*) were connected by R.L.Turner with OIA *peṭa-* ‘basket’ (Turner 1966: 475). Although such a semantic development, typologically quite possible,²² could in principle take place spontaneously in Indo-Aryan, it cannot be ruled out either that this change was “catalyzed” by the influence of the abovementioned Dravidian noun.

¹⁷ Dakhini is frequently considered a regional form of Urdu. Actually, it is a group of closely related dialects spoken by the Muslim population of the Deccan plateau in Central and South India, chiefly in the Telangana, Karnataka and Maharashtra states. Its speakers are mostly descendants of immigrants from North India and the Middle East. The lexical material used in my database belongs to the dialect spoken in Northern Karnataka.

¹⁸ Kotgarhi, Himachali, Kului, and Mandeali are spoken in Western Himalaya, mainly in the present-day Himachal Pradesh and Uttarakhand states of India. They are traditionally included in the Pahari subgroup.

¹⁹ Traditionally classified as a dialect of Rajasthani.

²⁰ A dialect of Bhili.

²¹ Sinhalese seems to have been influenced by Dravidian since very early times. In Konkani the Dravidian lexical stratum is mainly the result of contact between this language and its Southern neighbor Kannada. The Swadesh list for Konkani contains 4 Kannada loanwords, viz. *moḍ* ‘cloud’, *tanthe* ‘egg’, *urūṭ* ‘round’ and *bāl* ‘tail’.

²² As an approximate parallel cf. English *chest*, meaning both ‘box’ and ‘thorax’.

The languages that show the largest number of loans are Domaaki and Romany. In Domaaki there are 27 loanwords belonging to the Swadesh list. This percentage is no doubt abnormally high, but nevertheless quite explicable for a language on the verge of extinction, whose 300 speakers are all bi- or trilingual. The donor language for 19 loans is Shina,²³ the majority language and lingua franca of the area where Domaaki is spoken. The remaining 8 borrowings are of Burushaski origin.²⁴ The Romany Swadesh list²⁵ contains 19 loanwords, which are borrowed from different sources, namely Dardic (*parno* ‘white’²⁶), Burushaski (*cigno* ‘small’²⁷), Iranian (*por* ‘feather’, *čehran* ‘star’), Armenian (*morči* ‘skin’), Kartvelian (*kišay* ‘sand’²⁸), Greek (*kokalo* ‘bone’, *drom* ‘road’), Slavic (*zeleno* ‘green’, *koreno* ‘root’, *pliv-* ‘swim’), Romanian (*skarča* ‘bark’, *unjiya* ‘nail’, *nuvero* ‘cloud’, *lungo* ‘long’, *mint’a* ‘mountain’, *rotato* ‘round’, *seminca* ‘seed’, *galbeno* ‘yellow’).

The main source of etymologies is R. L. Turner’s comparative dictionary (Turner 1966). Domaaki and Parya etymologies are also taken from Buddruss 1984 and Oranskiy 1977 respectively.

The results of lexicostatistical calculations are given in Table 1.

The genealogical tree constructed by the StarLing system on the basis of the above data is reproduced on Figure 3.

As one can see, the classification represented by this tree differs from earlier classificatory schemes in quite a number of points. Below I list those differences which are, in my opinion, the most important ones.

1. According to the above classification, Indo-Aryan is subdivided into two main sub-branches — one including Sinhalese and Dhivehi (Maldivian), and the other consisting of all the other New Indo-Aryan languages. The most proper names for these subgroups would be “Insular” and “Continental”. The split of Proto-Indo-Aryan dates back to the close of the 2nd millennium B.C.

2. The Continental subgroup includes an outlying branch embracing Marathi and Konkani.

3. Romany turns out to form a common subgroup with Hindi-Urdu, Punjabi and dialects of the sub-Himalayan region, traditionally classified as Pahari (Nepali, Garhwali, Kumauni, Himachali, Kului, Mandeali and Kotgarhi). The closest relative of Romany is Domaaki, as was first hypothesized by D. L. R. Lorimer (1939). The split of “Proto-Hindi-Pahari-Romany” dates

²³ In order to save space, I will not list all the Shina loanwords here, especially since their identification usually presents no difficulty. The only doubtful case is Domaaki *šuno* ‘dog’, which, according to Georg Buddruss, should not be considered a borrowing from Shina because of its irregular inflection (Buddruss 1984: 14). Buddruss’s argument, however, does not seem convincing. The reflexes of Proto-Indo-Iranian **šuan-/šun-*, being widespread in Dardic, are very rare in New Indo-Aryan and almost never used there as the main word for dog, except in a few West Pahari languages, spoken adjacently to the Dardic-speaking area (cf., e.g., Siraji *šunā* ‘dog’). This fact suggests a high probability of borrowing from Dardic into Indo-Aryan. The immediate source for the above-cited Domaaki word may have been some older form of Shina *šū* ‘dog’.

²⁴ These are: *burinč* ‘cloud’, *tigon* ‘egg’, *čhumo* ‘fish’, *duwal-* ‘fly’, *jušiqam* ‘green’, *čhišā* ‘mountain’, *thop* ‘night’, *yono* ‘seed’.

²⁵ The Romany material reflects the Kalderash dialect, spoken in Romania and Moldova.

²⁶ Cf. Tirahi *parana*, Maiyā *panar*, Kashmiri *pron* (< **paranu*). Turner’s comparison of the Romany word with OIA *pāṇḍu-* (Turner 1966: 454) is doubtful, because it implies irregular phonetic development. The regular reflex of *pāṇḍu-* in Romany would have been **panrro* (cf. *punrro* ‘leg, foot’ < OIA *piṇḍa-* ‘calf of leg’).

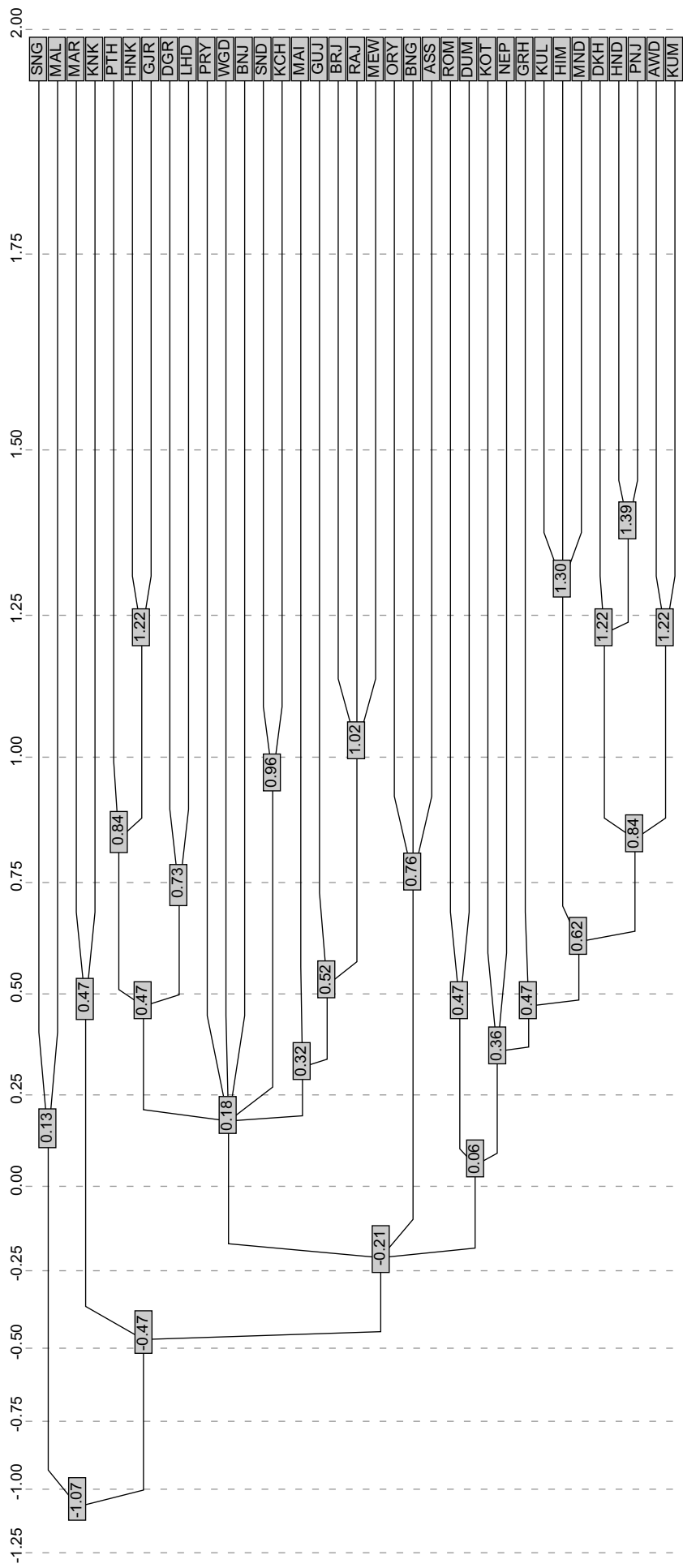
²⁷ On Burushaski loanwords in Romany see Berger 1959.

²⁸ Cf. Georgian, Laz *kviša* ‘sand’. The change *kv > k* is regular for genuine Romany words (cf. *kerel* ‘cooks’ < OIA *kvathati*). It may imply that during a certain period of time the cluster *kv* was proscribed in the language. If the borrowing of the Kartvelian word for ‘sand’ dates back to this period, the loss of *v* in the initial consonantal group is quite explicable.

Table 1. The lexicostatistical matrix of Indo-Aryan languages.

	HND	DKH	PNJ	PTH	HNK	GJR	DGR	LHD	SND	RAJ	GUJ	MAR	BNG	ASS	NEP	SNG	MAL	KOT	HIM	KUL	MND	ORY	AWD	KUM	ROM	KNK	DUM	BRJ	GRH	PRY	MAI	KCH	MEW	WGD	BNJ
HND		96	97	83	87	91	87	85	81	90	86	67	70	69	78	53	47	82	92	88	91	73	92	90	71	60	74	92	85	89	83	77	97	78	79
DKH	96		95	79	84	86	87	85	80	89	84	73	73	73	77	59	53	81	90	88	91	75	93	90	77	64	77	87	83	87	83	78	91	73	81
PNJ	97	95		89	91	93	93	90	84	92	84	68	70	72	76	57	50	80	92	85	91	72	91	89	71	67	72	88	82	93	82	81	94	75	80
PTH	83	79	89		89	93	86	82	77	80	73	69	66	63	72	50	45	75	86	79	84	69	78	75	68	58	67	77	71	77	75	75	80	64	67
HNK	87	84	91	89		95	88	87	82	81	76	67	69	67	70	54	48	77	85	81	85	68	80	78	71	59	71	78	76	83	76	77	82	67	70
GJR	91	86	93	93	95		91	83	81	84	81	69	69	67	74	52	45	79	90	84	91	70	85	82	70	64	69	81	77	86	74	75	86	73	75
DGR	87	87	93	86	88	91		87	82	84	79	67	68	67	77	55	45	79	91	86	92	69	87	80	69	66	77	81	76	83	81	79	86	70	74
LHD	85	85	90	82	87	83	87		84	80	74	63	67	67	69	50	46	76	85	79	85	67	83	78	68	62	68	77	73	80	78	86	83	68	69
SND	81	80	84	77	82	81	82	84		83	79	71	68	67	67	53	50	70	75	75	78	68	78	75	61	71	67	77	69	74	74	91	81	67	73
RAJ	90	89	92	80	81	84	84	80	83		90	75	70	70	76	57	50	78	86	86	86	76	90	84	70	72	73	91	78	81	81	86	91	75	81
GUJ	86	84	84	73	76	81	79	74	79	90		78	69	70	75	54	49	74	81	82	85	74	84	78	62	72	67	83	76	76	79	82	88	77	82
MAR	67	73	68	69	67	69	67	63	71	75	78		62	59	64	53	51	62	64	64	69	64	70	64	57	80	65	62	56	60	62	72	69	63	67
BNG	70	73	70	66	69	69	68	67	68	70	69	62		88	66	56	52	70	69	69	72	86	74	71	59	63	64	69	67	66	78	71	73	59	58
ASS	69	73	72	63	67	67	67	67	67	70	70	59	88		71	57	55	72	72	73	73	86	76	73	63	66	67	67	71	68	74	71	71	60	63
NEP	78	77	76	72	70	74	77	69	67	76	75	64	66	71		53	46	80	78	83	82	74	85	86	73	64	73	72	81	71	78	71	80	68	72
SNG	53	59	57	50	54	52	55	50	53	57	54	53	56	57	53		75	52	53	54	55	57	56	56	53	49	59	52	49	58	58	53	55	45	46
MAL	47	53	50	45	48	45	45	46	50	50	49	51	52	55	46	75		48	47	45	48	53	49	49	53	48	55	47	43	50	49	48	50	42	42
KOT	82	81	80	75	77	79	79	76	70	78	74	62	70	72	80	52	48		85	84	88	76	85	85	73	56	77	78	79	78	79	73	82	64	69
HIM	92	90	92	86	85	90	91	85	75	86	81	64	69	72	78	53	47	85		94	96	74	91	85	72	58	71	85	82	85	81	76	87	73	74
KUL	88	88	85	79	81	84	86	79	75	86	82	64	69	73	83	54	45	84	94		96	74	87	86	71	61	73	81	86	82	77	76	87	69	74
MND	91	91	91	84	85	91	92	85	78	86	85	69	72	73	82	55	48	88	96	96		75	88	86	72	65	77	84	83	86	82	77	90	73	77
ORY	72	74	70	68	67	69	68	66	67	75	73	63	87	87	73	57	53	75	73	73	74		77	74	65	63	72	72	72	70	80	72	76	68	63
AWD	92	93	91	78	80	85	87	83	78	90	84	70	74	76	85	56	49	85	91	87	88	78		95	78	67	79	88	87	83	88	79	91	76	82
KUM	90	90	89	75	78	82	80	78	75	84	78	64	71	73	86	56	49	85	85	85	86	86	75	95	77	61	78	84	90	83	86	76	91	71	75
ROM	71	77	71	68	71	70	69	68	61	70	62	57	59	63	73	53	53	73	72	71	72	65	78	77		54	82	70	75	72	68	64	74	61	60
KNK	60	64	67	60	60	66	66	62	71	70	72	82	65	66	64	51	49	57	59	61	65	63	67	61	54		60	59	53	59	59	75	66	63	66
DUM	74	77	72	67	71	69	77	68	67	75	67	65	64	67	73	59	55	77	71	73	77	72	79	78	82	62		72	72	77	76	67	81	61	67
BRJ	92	87	88	78	79	82	81	77	77	92	83	63	70	67	72	53	48	79	86	81	84	72	88	84	70	59	72		80	82	82	78	92	74	78
GRH	85	83	82	73	77	78	76	73	69	79	76	57	69	71	81	50	44	80	83	86	83	72	87	90	75	53	72	80	78	79	68	82	66	69	69
PRY	89	87	93	77	83	86	83	80	74	81	76	60	66	68	71	58	50	78	85	82	86	70	83	83	72	59	77	82	78		75	76	84	76	74
MAI	83	83	82	76	77	75	81	78	74	82	79	63	79	74	78	59	51	80	82	77	82	80	88	86	68	59	76	82	79	75		78	82	67	70
KCH	77	78	81	76	78	76	79	86	91	85	82	73	72	71	71	54	49	74	77	76	77	72	79	76	64	75	67	78	68	76	78		77	70	75
MEW	97	91	94	80	82	86	86	83	81	91	88	69	73	71	80	55	50	82	87	87	90	78	91	91	74	67	81	93	83	84	82	78		78	78
WGD	78	73	75	65	68	74	70	68	67	76	77	64	60	60	68	47	43	65	74	69	73	68	76	71	61	63	61	74	66	76	67	70	77		74
BNJ	79	81	80	67	70	75	74	69	73	82	82	67	58	63	72	46	42	69	74	74	77	63	82	75	60	66	67	78	69	74	70	75	78		74

Figure 3. The genealogical tree of Indo-Aryan languages. Figures in the nodes of the tree denote dates of separation in millennia A.D. (positive numbers) or B.C. (negative numbers), and are the results of glottochronological calculations performed in the StarLing system.



back to the 1st century A.D., and that of “Proto-Romany-Domaaki” to the close of the 5th century A.D. These figures, however, seem to be preliminary. The Swadesh lists of both Romany and Domaaki contain a number of unetymologized words. In the future, when etymologies of such words are established, the percentage of cognates may increase, and the resulting datings may appear to be somewhat younger. It is also worth noting that neither of the two above-mentioned dates should be automatically declared the date of the Gypsy exodus from India. The latter, no doubt, could date back to a later period than the linguistic split. Such a possibility is suggested by the fact that most subbranches of Continental Indo-Aryan diverged (sometimes nearly two millennia back) without mass migration of the speakers outside the subcontinent.

4. Forms of speech traditionally classified as Western Hindi do not actually form a single subgroup. Braj shows close relationship with Marwari and Mewati dialects spoken in Rajasthan, and somewhat more remote with Gujarati. Standard Hindi-Urdu and Dakhini are most closely related to Punjabi.²⁹ As for the Eastern Hindi dialects, their only representative in the database, i.e. Awadhi, is the closest relative of Kumauni.

5. Contrary to the traditional view, there is no reason to suggest a Rajasthani origin for a number of Indo-Aryan languages spoken outside Rajasthan. Thus Gojri,³⁰ classified by Grierson as a form of speech close to the Mewati dialect of Rajasthani, actually does not belong to the same group with the latter, but rather shows a close affinity to Hindko. The Banjari language,³¹ which was also usually considered as a variety of Rajasthani, actually occupies a somewhat independent position within one of the sub-branches of the Continental languages.

6. The above-stated close affinity of Kumauni to Awadhi implies that the Pahari group in the traditional sense does not exist as a genetic subdivision. The West Pahari languages except Kotgarhi (i.e., Himachali, Kului and Mandeali) do, however, form a single subgroup. The split of their ancestral language must have taken place very recently.

On the other hand, our classification does not differ from the earlier ones in those instances where the existence of subgroups is obvious or can be postulated on the basis of early linguistic evidence, as is, e.g., the case of the Eastern subgroup consisting of Oriya, Bengali and Assamese.

It should be emphasized again that the above classificatory scheme is preliminary and thus remains open to further amendments and improvements. It does not pretend to answer each and every question concerning New Indo-Aryan taxonomy. In a number of cases, it raises intriguing problems for further research. Among such problems, the rather close affinity of Maithili to Braj, Rajasthani dialects and Gujarati, as well as the somewhat isolated position of Parya and Garhwali appear to be particularly noteworthy.³² We hope that the present article will be instrumental in stimulating scholarly interest in these (and related) issues of Indo-Aryan comparative linguistics.

²⁹ The pair Hindi-Punjabi shows 97% matches, the highest percentage in the whole database.

³⁰ The language of Gujjars, a Muslim nomadic and semi-nomadic ethnic group dispersed in mountainous areas from Afghan Hindu Kush in the Northwest to the Indian state of Uttarakhand in the Southeast.

³¹ Also called Lamani and Lambadi, spoken by a semi-nomadic community of Banjaras scattered all over Central and Western India.

³² The Garhwali Swadesh list contains a significant number of unetymologized words. This fact may partly account for the relative isolation of the Garhwali language on the genealogical tree shown above. An abnormally low percentage of matches between Maldivian and Continental Indo-Aryan languages may have the same reason.

Appendix: Swadesh wordlists

The data are based on the following sources: HND, PNJ, SND, GUJ, MAR, BNG, SNG — Kogan 2005; DKH — Sibghatulla, Zamin 2000; PTH — a native speaker; HNK — Sultān Sukūn 2002; GJR — Awan 2000; DGR — Gosvāmī 2000; LHD — Saleem, Shah 2005 and Kogan 2005; RAJ — Suthar, Gahlot 1995 and Mukherji et al. 2011; ASS — Neog, Goswami 1987; NEP — Schmidt 1994; MAL — Abdulla, O’Shea 2005; KOT — Hendriksen 1976; HIM — native speakers; KUL — Mahapatra, Padmanabha, Ranganatha 1980; MND — native speakers and Mahapatra, Padmanabha, Ranganatha 1980; ORY — native speakers and Praharaj 1931–1940; AWD — Samīr 1955; KUM, GRH — native speakers and Grierson 1916; ROM — Boretzky 1994; KNK — Thali 1999–2001; DUM — Buddruss 1984 and Lorimer 1939; BRJ, MEW, WGD — Mukherji et al. 2011; PRY — Oranskiy 1977; MAI — Thakur, Jha 2012; KCH — Rohra 1965; BNJ — Ramesh 2010.

Figures in brackets after words refer to numbers of etymologies in the Indo-Aryan etymological database. They also reflect cognacy, words descending from the same OIA root having equal numbers. Negative numbers are assigned to loanwords and lacunae.

1. ALL: HND sab (1), DKH sab (1), PNJ sabh (1), PTH sāre (1), HNK sāre (1), GJR sārā (1), DGR sab (1), LHD sārā (1), SND sabhi (1), RAJ sagḷau (270), GUJ sahu (1), MAR sagḷā (270), BNG šobay (1), ASS xob (1), NEP sabai (1), SNG siyallō (270), MAL emmehā (174), KOT sarɔ (1), HIM sāre (1), KUL səbh (1), MND sabh (1), ORY sabu (1), AWD sab (1), KUM sap (1), ROM savorre (1), KNK sagḷo (270), DUM buṭā (-1), BRJ sab (1), GRH sabbi (1), PRY sare (1), MAI sab (1), KCH sab (1), MEW sabe (1), WGD sabu (1), BNJ sārī (1)

1a. ALL: SND samūro (174), SNG muḷu (381) MAL hurihā (522) AWD sagar (270)

2. ASHES: HND rākh (3), DKH rāx (3), PNJ suāh (111), PTH suhāgā (111) HNK chāi (152), GJR bhass (489), DGR kheh (440), LHD chāi (152), SND rakh (3), RAJ khe (440), GUJ rākh (3), MAR rākh (3), BNG chai (152), ASS sai (152), NEP kharāni (140), SNG aḷu (382), MAL aḷi (382) KOT kheḗ (440), HIM suāh (111), KUL ? (-1), MND swāh (111), ORY pāuñša (567), AWD rākhī (3), KUM saji (488), ROM vušar (140), KNK goboru (606), DUM čhor (140), BRJ rākh (3), GRH chāru (140), PRY čhar (140), MAI chār (140) KCH vānī (653) MEW rakhi (3), WGD pāsī (567), BNJ rāk (3)

2a. ASHES: GJR suhāgo (111), DGR bhass (489), LHD suā (111), SND chāi (152), RAJ rākh (3), KOT chār (140), MND bhās (489), ORY chāra (140), AWD chār (140), KUM chār (140)

3. BARK: HND chāl (4), DKH chilṭā (4), PNJ chill (4), PTH chilar (4), HNK chillar (4), GJR chilṛo (4), DGR chilkā (4), LHD chil (4), SND chalu (4), RAJ chilkau (4), GUJ chāl (4), MAR sāl (4), BNG chal (4), ASS bakoli (328), NEP bokro (328), SNG potta (383), MAL thoši (523), KOT chāl (4), HIM chillekar (4), KUL khol (552), MND sātū (561), ORY chāli (4), AWD chāl (4), KUM bakkhal (328), ROM skarča (-1), KNK sāl (4), DUM ? (-2), BRJ chāl (4), GRH bakkal (328), PRY pōst (-1), MAI chāl (4), KCH chall (4), MEW chāl (4), WGD sal (4), BNJ chāmbḍi (667)

3a. BARK: ASS sal (4), ORY bakkala (328), MEW bokalā (328)

4. BELLY: HND peṭ (678), DKH peṭ (678), PNJ peṭ (678), PTH ḍhiḍ (113), HNK ḍhiḍ (113), GJR ḍhiḍ (113), DGR ḍhiḍḍ (113), LHD ḍhiḍh (113), SND peṭu (678), RAJ peṭ (678), GUJ peṭ (678), MAR poṭ (-1), BNG peṭ (678), ASS pet (678), NEP peṭ (678), SNG baḍa (329), MAL ban’du (329), KOT peṭ (678), HIM peṭ (678), KUL pēṭ (678), MND pēṭ (678), ORY peṭa (678), AWD pēṭ (678),

KUM peṭ (678), ROM perr (678), KNK poṭ (-1), DUM peṭ (678), BRJ peṭ (678), GRH pyaṭ (678), PRY peṭ (678), MAI udar (-1), KCH peṭ (678), MEW peṭ (678), WGD peṭ (678), BNJ peṭ (678)

4a. BELLY: PNJ dhiḍḍ (113), GJR peṭ (678)

5. BIG: HND baṛā (5), DKH baḍā (5), PNJ vaḍḍā (5), PTH baṛā (5), HNK baḍḍā (5), GJR baṛo (5), DGR baḍḍā (5), LHD vaḍḍā (5), SND vaḍ'o (5), RAJ baḍau (5), GUJ moṭo (248), MAR moṭhā (248), BNG bəṛo (5), ASS bəṛ (5), NEP ṭhūlo (358), SNG loku (385), MAL boḍu (669), KOT bəḍḍo (5), HIM baṛā (5), KUL bəḍḍā (5), MND bəḍḍā (5), ORY baḍḍā (5), AWD baṛā (5), KUM baṛo (5), ROM baṛo (5), KNK v(h)od (5), DUM baḍā (5), BRJ bāṛo (5), GRH baṛū (5), PRY boṛo (5), MAI baṛkā (5), KCH vaḍo (5), MEW baṛo (5), WGD moṭo (248), BNJ moṭo (248)

5a. BIG: RAJ moṭau (248), NEP baṛo (5)

6. BIRD: HND ciṛiyā (6), DKH pakherū (7), PNJ pakkherū (7), PTH pakherū (7), HNK pakhrū (7), GJR pakherū (7), DGR paīchi (7), LHD pakkhī (7), SND pakhī (7), RAJ pākherū (7), GUJ pākhī (7), MAR paksī (-2), BNG pakhi (7), ASS sərai (6), NEP carā (6), SNG kurullā (-1), MAL dūni (524), KOT cəṛku (6), HIM panchī (7), KUL cīḍu (6), MND pənchi (7), ORY caḍhei (6), AWD ciraī (6), KUM callo (6), ROM čirikli (6), KNK pākko (7), DUM čai (6), BRJ paksī (-1), GRH cakhul (652), PRY kukuṛi (644), MAI ciṛai (6), KCH pakhi (7), MEW ciṛiā (6), WGD huṛo (660), BNJ gargol (668)

6a. BIRD: HND pākhī (7), DKH caṛī (6), DGR ciṛī (6), RAJ ciṛī (6), ASS pənkhī (7), KOT panchi (7), KUL pənchi (7), MND cīḍu (6), AWD panchī (7), GRH pacche (7), MAI panchī (7), MEW pakherū (7)

7. BITE: HND kāt- (8), DKH kāt- (8), PNJ kaṭṭ- (8), PTH laṛ- (153), HNK laṛ- (153), GJR khā- (490), DGR baḍḍh- (493), LHD paṭ- (501), SND caku pāi- (176), RAJ karar (249), GUJ karaḍ- (249), MAR cāv- (271), BNG kamṛa- (296), ASS kamər- (296), NEP ṭok- (114), SNG hapa- (386), MAL daiganna- (550), KOT ṭuk- (114), HIM dānd mār- (550), KUL kāt- (8), MND kāt- (8), ORY kāmur- (296), AWD ? (-1), KUM kāt- (8), ROM dandar- (550), KNK cāb- (271), DUM jan- (-3), BRJ kaṭ- (8), GRH kāt- (8), PRY kaṭ- (8), MAI bhamhor- (676), KCH cak- (176), MEW kāt- (8), WGD kaṭ- (8), BNJ cāb- (271)

7a. BITE: HNK cak mār- (176), GJR laṛ- (153), SND d'ārḥ- (177), MAR ḍāsā ghe- (177), SNG vik- (387), HIM khā- (490), MND khā- (490), PRY čak par- (176), KCH viṛ- (654), MEW cabā- (271), BNJ kāt- (8)

8. BLACK: HND kālā (9), DKH kālā (9), PNJ kālā (9), PTH kālā (9), HNK kālā (9), GJR kālo (9), DGR kālā (9), LHD kālā (9), SND kāro (9), RAJ kālāu (9), GUJ kālō (9), MAR kālā (9), BNG kalo (9), ASS kōla (9), NEP kālō (9), SNG kaḷu (9), MAL kaḷu (9), KOT kaḷo (9), HIM kālā (9), KUL kālā (9), MND kālā (9), ORY kalā (9), AWD kariā (9), KUM kālō (9), ROM kalo (9), KNK kālē (9), DUM kālā (9), BRJ kāṛo (9), GRH kālū (9), PRY kalo (9), MAI kāri (9), KCH kāro (9), MEW kālō (9), WGD kālō (9), BNJ kālō (9)

9. BLOOD: HND lahū (10), DKH lhau (10), PNJ lahū (10), PTH lahū (10), HNK rat (115), GJR rat (115), DGR lahu (10), LHD ratt (115), SND ratu (115), RAJ lohī (10), GUJ lohī (10), MAR rakt (-3), BNG rəktə (-1), ASS tez (330), NEP ragat (-1), SNG lē (10), MAL ley (10), KOT lou (10), HIM laū (10), KUL lōhu (10), MND lahū (10), ORY rakta (-1), AWD rakat (-2), KUM lwe (10), ROM rat (115), KNK ragat (-2), DUM rot (115), BRJ lohī (10), GRH loi (10), PRY law (10), MAI lohū (10), KCH ratt (115), MEW khūn (-1), WGD khōn (-1), BNJ loi (10)

9a. BLOOD: PNJ ratt (115), PTH rat (115), HNK lahū (10), DGR ratt (115), KCH lui (10)

10. BONE: HND haḍḍī (11), DKH haḍ (11), PNJ haḍḍī (11), PTH haḍī (11), HNK haḍḍī (11), GJR haḍḍī (11), DGR haḍḍī (11), LHD haḍḍī (11), SND haḍ'o (11), RAJ hāḍ (11), GUJ hāḍkū (11), MAR hāḍ (11), BNG haḍ (11), ASS haḍ (11), NEP hāḍ (11), SNG āṭaya (388), MAL kaṣi (526), KOT haḍḍ (11), HIM āḍ (11), KUL hāḍka (11), MND haḍḍī (11), ORY hāḍa (11), AWD hār (11), KUM hār (11), ROM kokalo (-2), KNK hāḍ (11), DUM hoṭ (11), BRJ haḍḍī (11), GRH hāḍ (11), PRY haḍi (11), MAI haḍḍī (11), KCH haḍo (11), MEW haḍḍī (11), WGD haḍko (11), BNJ haḍkā (11)

11. BREAST: HND chātī (12), DKH chātī (12), PNJ chātī (12), PTH chātī (12), HNK mammā (478), GJR chātī (12), DGR chātī (12), LHD bubbā (154), SND chātī (12), RAJ chātī (12), GUJ chātī (12), MAR chātī (-4), BNG buk (297), ASS buku (297), NEP chātī (12), SNG ḷaya (116), MAL buk (297), KOT cui (141), HIM chātī (12), KUL chātī (12), MND hīk (116), ORY chātī (12), AWD chātī (12), KUM chātī (12), ROM kolin (592), KNK hardhen (116), DUM titiro (-4), BRJ chātī (12), GRH ? (-1), PRY hik (116), MAI chātī (12), KCH thaṇ (655), MEW chātī (12), WGD sātī (12), BNJ chātī (12)

11a. BREAST: PNJ hikk (116), DGR hikkṛū (116), SND urahu (178), RAJ bobau (154), MAL uramathi (178), KUL hik (116), MND chātī (12), ORY buku (297)

12. BURN TR.: HND jalā- (13), DKH jāl- (13), PNJ bāl- (13), PTH sār- (117), HNK sār- (117), GJR sār- (117), DGR bāl- (13), LHD bāl- (13), SND sār- (117), RAJ phūk- (505), GUJ bāl- (13), MAR jāl- (13), BNG poṛa- (298), ASS por- (298), NEP bāl- (13), SNG dava- (217), MAL ada- (217), KOT dzaḷ- (13), HIM bāl- (13), KUL phukh- (505), MND phukh- (505), ORY poṛi- (298), AWD jarā- (13), KUM jal- (13), ROM phabar- (593), KNK jaḷayi- (13), DUM dāi- (-5), BRJ jāl- (13), GRH jagau- (653), PRY jal- (13), MAI jār- (13), KCH bār- (13), MEW jalā- (13), WGD jal- (13), BNJ bāl- (13)

12a. BURN TR.: PNJ sār- (117), bāl- (13), GJR jāl- (13), DGR sār- (117), LHD sār- (117), SND b'ār- (13), RAJ jalā- (13), BNG jala- (13), NEP ḍaṛhāu- (217), SNG pulussa- (390), KOT d5- (217), MND baḷā- (13), AWD ḍaṛhiā- (217), PRY sar бага- (117)

13. NAIL: HND nāḅun (-1), DKH nhũ (14), PNJ nahũ (14), PTH nũ (14), HNK nũh (14), GJR nahũ (14), DGR naũ (14), LHD nũ (14), SND nāhũ (14), RAJ noriyau (14), GUJ nakh (14), MAR nakh (14), BNG nōkh (14), ASS nōkh (14), NEP naṇ (14), SNG niyapotta (14), MAL niyafathi (14), KOT nōš (14), HIM nōkh (14), KUL nhōš (14), MND nēh (14), ORY nakha (14), AWD nah (14), KUM naṇ (14), ROM unjiya (-3), KNK naṇkuṭ (14), DUM naurā (-6), BRJ nōhō (14), GRH nang (14), PRY nuk (-2), MAI nah (14), KCH nō (14), MEW nũh (14), WGD nōkh (14), BNJ nak (14)

14. CLOUD: HND bādal (15), DKH abhāl (465), PNJ baddal (15), PTH baddal (15), HNK jhar (155), GJR jhar (155), DGR badal (15), LHD baddal (15), SND kakaru (179), RAJ bādḷau (15), GUJ vādḷū (15), MAR ḍhag (272), BNG barid (-2), ASS badal (15), NEP bādal (15), SNG valākula (391), MAL vilā (391), KOT badal (15), HIM baddal (15), KUL bādal (15), MND baddal (15), ORY megha (-2), AWD bādar (15), KUM bādal (15), ROM nuvero (-4), KNK moḍ (-3), DUM burinč (-7), BRJ badalore (15), GRH bādal (15), PRY badal (15), MAI bādari (15), KCH vāḍar (15), MEW bādar (15), WGD wādlo (15), BNJ abāl (465)

14a. CLOUD: PTH jhar (155)

15. COLD: HND ṭhaṇḍā (458), DKH ṭhaṇḍā (458), PNJ ṭhaṇḍā (458), PTH ṭhaṇḍā (458), HNK ṭhaṇḍā (458), GJR ṭhaṇḍo (458), DGR ṭhaṇḍā (458), LHD ṭhaḍḍhā (458), SND ṭhadho (458), RAJ

ṭhaṇḍau (458), GUJ ṭhaṇḍū (458), MAR ṭhaṇḍ (458), BNG ṭhaṇḍa (458), ASS ṭhanda (458), NEP ciso (332), SNG sisil (392), MAL fini (528), KOT šillb (218), HIM ṭhaṇḍā (458), KUL ṭhṇḍa (458), MND ṭhaṇḍa (458), ORY ṭhaṇḍā (458), AWD ṭhaṇḍh (458), KUM ṭhaṇḍo (458), ROM šilalo (218), KNK ṭhaṇḍ (458), DUM šildā (218), BRJ siāro (218), GRH ṭhaṇḍo (458), PRY yaḡ (-3), MAI ṭhār (458), KCH thadho (458), MEW siro (218), WGD ṭaḍu (458), BNJ siḷo (218)
15a COLD: KNK šeḷo (218) MEW ṭhaṇḍho (458)

16. COME: HND ā- (16), DKH ā(v)- (16), PNJ ā- (16), PTH as- (181) HNK (16), GJR ā- (16), DGR au- (16), LHD āv- (16), SND ac- (180) RAJ āv- (16), GUJ av- (16), MAR ye- (180) BNG aš- (301), ASS ah- (301), NEP āu- (16), SNG e- (180), MAL ai- (180) KOT ā- (16), HIM ā- (16), KUL āw- (16), MND āu- (16), ORY ās- (301), AWD ā- (16), KUM ū- (16), ROM av- (16), KNK yo- (180), DUM (16), BRJ ā- (16), GRH ā- (16), PRY a- (16), MAI ā- (16), KCH ac- (180), MEW (16), WGD (16), BNJ ā- (16)

16a. COME: SND āy- (181), MAR ālā- (181), BNG ail- (181), NEP āy- (181), KOT ach- (181), KUL ye- (180), ORY ail- (181), KNK āyil- (181), KCH āy- (181)

17. DIE: HND mar- (17), DKH mar- (17), PNJ mar- (17), PTH mar- (17), HNK mar- (17), GJR mar- (17), DGR mar- (17), LHD mar- (17), SND mar- (17), RAJ mar- (17), GUJ mar- (17), MAR mar- (17), BNG mār- (17), ASS mār- (17), NEP mar- (17), SNG märe- (17), MAL maru- (17), KOT mār- (17), HIM mar- (17), KUL mār- (17), MND mar- (17), ORY mar- (17), AWD mar- (17), KUM mar- (17), ROM mer- (17), KNK mar- (17), DUM mar- (17), BRJ mar- (17), GRH mar- (17), PRY mar- (17), MAI mar- (17), KCH mar- (17), MEW mar- (17), WGD mar- (17), BNJ mar- (17)

18. DOG: HND kuttā (18), DKH kuttā (18), PNJ kuttā (18), PTH kuttā (18), HNK kuttā (18), GJR kuto (18), DGR kuttā (18), LHD kuttā (18), SND kutto (18), RAJ kutau (18), GUJ kutro (18), MAR kutrā (18), BNG kukur (302), ASS kukur (302), NEP kukur (302), SNG ballā (394), MAL baḷu (394), KOT kukkər (302), HIM kuttā (18), KUL kutta (18), MND kutta (18), ORY kukkura (302), AWD kūkur (302), KUM kukur (302), ROM žukel (595), KNK kutro (18), DUM šuno (-8), BRJ kuttā (18), GRH kukur (302), PRY kuto (18), MAI kukkur (302), KCH kauto (18), MEW kutto (18), WGD kutro (18), BNJ kutrā (18)

18a. DOG: MAL kutthā (18)

19. DRINK: HND pī- (19), DKH pī(v)- (19), PNJ pī- (19), PTH pī- (19), HNK pī- (19), GJR pī- (19), DGR pī- (19), LHD pīv- (19), SND pī- (19), RAJ pī- (19), GUJ pi- (19), MAR pi- (19), BNG kha- (303), ASS pi- (19), NEP piu- (19), SNG bo- (19), MAL bo- (19), KOT pī- (19), HIM pī- (19), KUL pī- (19), MND pī- (19), ORY pī- (19), AWD pi- (19), KUM pī- (19), ROM pi- (19), KNK pī- (19), DUM pi- (19), BRJ pī- (19), GRH pya- (19), PRY pi- (19), MAI pī- (19), KCH pī- (19), MEW pī- (19), WGD pī- (19), BNJ pi- (19)

20. DRY: HND sukhā (20), DKH sukkā (20), PNJ sukkā (20), PTH sukkā (20), HNK sukkā (20), GJR suko (20), DGR sukkā (20), LHD sukkā (20), SND sukalu (20), RAJ sūkhau (20), GUJ sukū (20), MAR sukhā (20), BNG šukno (20), ASS xukan (20), NEP sukkho (20), SNG viyaḷi (456), MAL hiki (20), KOT šukkhṇ (20), HIM sukk(20), KUL ? (-2), MND sukhā (20), ORY sukhila (20), AWD sukhā(20), KUM sukko (20), ROM šuko (20), KNK sukk(h)o (20), DUM šukhā (20), BRJ sukho (20), GRH sukhyū (20), PRY suke (20), MAI sukkhā (20), KCH saukko (20), MEW sūkhā (20), WGD suko (20), BNJ suko (20)

21. EAR: HND kān (21), DKH kān (21), PNJ kann (21), PTH kan (21), HNK kaṇ (21), GJR kaṇ (21), DGR kann (21), LHD kan (21), SND kanu (21), RAJ kān (21), GUJ kān (21), MAR kān (21),

BNG kaṅ (21), ASS kaṅ (21), NEP kān (21), SNG kana (21), MAL kanfaiy (21), KOT kān (21), HIM kān (21), KUL kōn (21), MND kān (21), ORY kāṇa (21), AWD kān (21), KUM kān (21), ROM kan (21), KNK kānu (21), DUM koṅ (21), BRJ kān (21), GRH kān (21), PRY kan (21), MAI kān (21), KCH kann (21), MEW kān (21), WGD kān (21), BNJ kān (21)

22. EARTH: HND dhartī (22), DKH dharat (22), PNJ dhartī (22), PTH miṭṭī (157), HNK dhartī (22), GJR dhartī (22), DGR dhartī (22), LHD bhoē (118), SND dhartī (22), RAJ dhartī (22), GUJ dhartī (22), MAR bhuī (118), BNG maṭi (157), ASS (157), NEP dhartī (22), SNG pasa (396), MAL bin (118), KOT bhuī (118), HIM dhartī (22), KUL dhōrti (22), MND dhartī (22), ORY māṭi (157), AWD dhartī (22), KUM māṭo (157), ROM phuv (118), KNK bhūyi (118), DUM čhor (623), BRJ dhartī (22), GRH dharti (22), PRY zimin (-4), MAI māṭi (157), KCH miṭṭī (157), MEW bhūm (118), WGD dhartī (22), BNJ dharti (22)

22a. EARTH: DKH bhū (118), PNJ bhō (118), GJR miṭi (157), MAR māṭi (157), NEP māṭo (157), SNG poḷova (397), KOT maṭṭi (157), MND bhūy (118), AWD bhuī (118), KUM bhī (118), KNK māṭi (157), MEW dharatī (22)

23. EAT: HND khā- (23), DKH khā- (23), PNJ khā- (23), PTH khā- (23), HNK khā- (23), GJR khā- (23), DGR khā- (23), LHD khā- (23), SND khā- (23), RAJ khāv- (23), GUJ khā- (23), MAR khā- (23), BNG kha- (23), ASS kha- (23), NEP khā- (23), SNG ka- (23), MAL ka- (23), KOT khā- (23), HIM khā- (23), KUL khā- (23), MND khā- (23), ORY khā- (23), AWD khā- (23), KUM khā- (23), ROM xa- (23), KNK kha- (23), DUM khā- (23), BRJ khā- (23), GRH khā- (23), PRY kha- (23), MAI khā- (23), KCH khāy- (23), MEW khā- (23), WGD khāy- (23), BNJ khā- (23)

24. EGG: HND aṅḍā (24), DKH aṅḍ(r)ā (24), PNJ āṅḍā (24), PTH aṅḍā (24), HNK aṭṛā (24), GJR āṅṭro (24), DGR āṅḍā (24), LHD āṅā (24), SND āno (24), RAJ aṅḍau (24), GUJ iṅḍū (24), MAR aṅḍē (24), BNG ḍim (304), ASS ḍima (304), NEP aṅḍā (24), SNG aṅḍa (24), MAL bis (398), KOT pinne (442), HIM aṅḍā (24), KUL aṅḍa (24), MND aṅḍa (24), ORY ḍim (304), AWD aṅḍā (24), KUM aṅḍa (24), ROM anrro (24), KNK tanthe (-4), DUM tigon (-9), BRJ āṅḍā (24), GRH aṅḍru (24), PRY aṅḍō (24), MAI ḍim (304), KCH inō (24), MEW aṅḍā (24), WGD ēḍā (24), BNJ iṅḍā (24)

25. EYE: HND ākh (25), DKH ankhī (25), PNJ akkh (25), PTH akh (25), HNK akh (25), GJR akh (25), DGR akkh (25), LHD akkh (25), SND akhi (25), RAJ ākh (25), GUJ ākh (25), MAR ḍoḷā (274), BNG cokh (305), ASS soku (305), NEP ākhā (25), SNG āsa (25), MAL lō (529), KOT akkh (25), HIM ākh (25), KUL hōkhi (25), MND hākh (25), ORY ākhi (25), AWD ākhi (25), KUM ākho (25), ROM yak(h) (25), KNK doḷo (274), DUM ač (25), BRJ āk (25), GRH ākho (25), PRY ank (25), MAI ākh (25), KCH akh (25), MEW ākh (25), WGD ākh (25), BNJ āngki (25)

26. FAT N.: HND carbī (-2), DKH carbī (-1), PNJ carbī (-1), PTH thendā (474), HNK mēj (158), GJR carbī (-1), DGR mīj (158), LHD carbī (-1), SND carbī (-1), RAJ cīgtau (220), GUJ carbī (-1), MAR carbī (-5), BNG corbi (-3), ASS sārbi (-1), NEP boso (361), SNG tela (399), MAL ? (-1), KOT cikṭo (220), HIM carbī (-1), KUL ? (-3), MND miñj (158), ORY carbī (-3), AWD carbī (-3), KUM baso (361), ROM čiken (220), KNK vos (361), DUM mao (-10), BRJ carbī (-2), GRH bāvāḷi (654), PRY lant (645), MAI carbī (-2), KCH carbī (-1), MEW carbī (-2), WGD carbī (-2), BNJ maḷai (670)

26a. FAT N.: KOT mīndz (158)

27. FEATHER: HND pākh (26), DKH pākh (26), PNJ pākh (26), PTH par (-1), HNK kham (119), GJR khamm (119), DGR phāgh (26), LHD par (-2), SND khambh (119), RAJ pākh (26), GUJ pīchū (250), MAR pākh (26), BNG palok (306), ASS pakhi (26), NEP pvākh (26), SNG

pihātuva (250), MAL dūnifaiy (250), KOT pankh (26), HIM pākh (26), KUL pankh (26), MND pākh (26), ORY pālak (306), AWD paṅkh (26), KUM pankh (26), ROM por (-5), KNK pāk (26), DUM ? (-11), BRJ pākh (26), GRH ? (-2), PRY par (-5), MAI picch (250), KCH ? (-2), MEW pankhī (26), WGD pakh (26), BNJ pangkoḍā (26)

27a. FEATHER: PNJ khābh (119), GJR pakkh (26), DGR khambh (119), SND pakkh (26), GUJ pāk (26), MAR pīs (250)

28. FIRE: HND āg (27), DKH ag (27), PNJ agg (27), PTH ag (27), HNK ag (27), GJR ag (27), DGR agg (27), LHD bhāh (679), SND bāhi (679), RAJ āg (27), GUJ āg (27), MAR āg (27), BNG agun (-4), ASS zui (334), NEP āgo (27), SNG ginna (-2), MAL alifān (531), KOT āg (27), HIM āg (27), KUL āg (27), MND āg (27), ORY niā (673), AWD āgi (27), KUM āgo (27), ROM yag (27), KNK ujjo (334), DUM ak (27), BRJ āge (27), GRH āg (27), PRY ak (27), MAI āgi (27), KCH āg (-3), MEW āgi (27), WGD āgi (27), BNJ angār (671)

29. FISH: HND machlī (28), DKH macchī (28), PNJ macchī (28), PTH machlī (28), HNK macchī (28), GJR machī (28), DGR macchī (28), LHD machī (28), SND machī (28), RAJ māchlī (28), GUJ māchlī (28), MAR māsā (28), BNG mach (28), ASS mas (28), NEP māchā (28), SNG māluvā (-3), MAL mas (28), KOT macchi (28), HIM macchī (28), KUL machlī (28), MND macchī (28), ORY māccha (28), AWD macharī (28), KUM mācho (28), ROM mašo (28), KNK māsoḷi (28), DUM čumo (-12), BRJ macchī (28), GRH māchu (28), PRY mačhi (28), MAI māch (28), KCH machi (28), MEW macchī (28), WGD machlī (28), BNJ mācli (28)

30. FLY V.: HND uṛ- (29), DKH uṛ- (29), PNJ uḍḍ- (29), PTH uḍar- (29), HNK uḍ(ar)- (29), GJR uḍ- (29), DGR uḍḍ- (29), LHD uḍḍ- (29), SND uḍ'ir- (29), RAJ uḍ- (29), GUJ uḍ- (29), MAR uḍ- (29), BNG uṛe ja- (29), ASS ur- (29), NEP uḍ- (29), SNG piyāmba- (401), MAL uduh- (532), KOT ṛeu- (29), HIM uḍ- (29), KUL uḍḍ- (29), MND uḍ- (29), ORY uṛ- (29), AWD uṛ- (29), KUM uṛ- (29), ROM (h)urr- (29), KNK uḍ- (29), DUM duwal- (-13), BRJ uṛ- (29), GRH uṛ- (29), PRY uḍ- (29), MAI uṛ- (29), KCH auḍ- (29), MEW uṛ- (29), WGD uḍ- (29), BNJ waḍ- (29)

31. FOOT: HND pair (30), DKH pāō (30), PNJ pair (30), PTH pair (30), HNK pair (30), GJR pair (30), DGR pair (30), LHD pēr (30), SND per (30), RAJ pag (30), GUJ pag (30), MAR pāy (30), BNG pa (30), ASS pau (30), NEP pāu (30), SNG paya (30), MAL faithila (30), KOT pau (30), HIM pair (30), KUL per (30), MND pair (30), ORY pā (30), AWD pag (30), KUM khuṭṭo (587), ROM punrro (596), KNK pāy (30), DUM po (30), BRJ pair (30), GRH khuṭṭu (587), PRY per (30), MAI pa (30), KCH pag (30), MEW pēr (30), WGD pog (30), BNJ pak (30)

32. FULL: HND bharā (31), DKH pūrbhar (31), PNJ bhariyā (31), PTH pūrā (32), HNK pūrā (32), GJR pūro (32), DGR pūrā (32), LHD bhareā (31), SND bhariyalu (31), RAJ pūr (32), GUJ pūrū (32), MAR bharlē (31), BNG bhōra (31), ASS pur (32), NEP bharī (31), SNG piri (32), MAL furifai (32), KOT bhōra (31), HIM bhareādā (31), KUL bhōru (31), MND bharyā (31), ORY pūra (32), AWD purhar (32), KUM purro (32), ROM pherdo (31), KNK bhartī (31), DUM ? (-14), BRJ puriyo (32), GRH bharyū (31), PRY par (31), MAI bharit (31), KCH sajo (656), MEW pūro (32), WGD bhareo (31), BNJ bharo (31)

32a. FULL: HND pūrā (32), DKH pūrā (32), PNJ pūrā (32), PTH bhariyā (31), HNK bhariyā (31), GJR bhariyo (31), DGR bharā (31), LHD pūrā (32), SND pūro (32), RAJ bharpūr (31), GUJ bharpūr (31), MAR purā (32), BNG pura (32), ASS bhōra (31), KOT purō (32), HIM pūrā (32), MND pūra (32), ORY bhari (31), AWD bharā (31), KUM bhariyo (31), GRH pūru (32), KCH bhariyo (31)

33. GIVE: HND de- (33), DKH de(v)- (33), PNJ de- (33), PTH de- (33), HNK de- (33), GJR de- (33), DGR de- (33), LHD đev- (33), SND đ'i- (33), RAJ de- (33), GUJ de- (33), MAR de- (33), BNG de- (33), ASS di- (33), NEP di- (33), SNG de- (33), MAL din- (33), KOT de- (33), HIM de- (33), KUL de- (33), MND de- (33), ORY de- (33), AWD de- (33), KUM di- (33), ROM de- (33), KNK di- (33), DUM de- (33), BRJ dew- (33), GRH di- (33), PRY de- (33), MAI de- (33), KCH đe- (33), MEW de- (33), WGD de- (33), BNJ de- (33)

34. GOOD: HND acchā (34), DKH acchā (34), PNJ hachā (34), PTH acchā (34), HNK hacchā (34), GJR acho (34), DGR caṅgā (120), LHD caṅgā (120), SND cango (120), RAJ acchau (34), GUJ bhalo (184), MAR cāṅlā (120), BNG bhalb (184), ASS bhal (184), NEP ramro (362), SNG hoñda (402), MAL rangaḷu (533), KOT bhɔɔ (184), HIM acchā (34), KUL bhala (184), MND kharā (554), ORY bhala (184), AWD nīk (466), KUM niko (466), ROM lašo (34), KNK cāṅ (120), DUM šonā (624), BRJ ācho (34), GRH accho (34), PRY hačo (34), MAI nik (466), KCH cangho (120), MEW bhalo (184), WGD haue (661), BNJ āco (34)

34a. GOOD: DKH nīkā (466), PNJ caṅgā (120), PTH caṅgā (120), HNK caṅgā (120), GJR caṅgo (120), DGR kharā (494), RAJ bhalau (184), NEP bhalo (184), SNG manā (403), KUL khəra (554), MND bhalla (184), AWD acchā (34), KUM ācho (34), GRH bhalu (184)

35. GREEN: HND harā (35), DKH haryā (35), PNJ har(i)ā (35), PTH harā (35), HNK sāvā (160), GJR nīlo (253), DGR saillā (495), LHD sāvā (160), SND sāo (160), RAJ līlau (253), GUJ lilo (253), MAR hirvā (35), BNG šɔbuj (-5), ASS xeuj (-2), NEP hariyo (35), SNG palāvan (404), MAL fe-hikulaige (534), KOT hɔrɔ (35), HIM arā (35), KUL hōra (35), MND saillā (495), ORY sabujā (-4), AWD hariar (35), KUM hariyo (35), ROM zeleno (-6), KNK pācve (612), DUM jutšiqam (-15), BRJ haro (35), GRH haryū (35), PRY sōz (-6), MAI sabuj (-3), KCH sāo (160), MEW haro (35), WGD haro (35), BNJ haro (35)

35a. GREEN: GJR hariyo (35), RAJ harau (35), MND harā (35)

36. HAIR: HND bāl (36), DKH bāl (36), PNJ vāl (36), PTH bāl (36), HNK bāl (36), GJR bāl (36), DGR kes (185), LHD vāl (36), SND vār (36), RAJ bāl (36), GUJ vāl (36), MAR kēs (185), BNG cul (308), ASS suli (308), NEP bāl (36), SNG kespata (185), MAL isthaši (535), KOT bāl (36), HIM bāl (36), KUL šərāl (36), MND bālḥ (36), ORY cuḷi (308), AWD bār (36), KUM bāl (36), ROM bal (36), KNK kes (185), DUM jāt (-16), BRJ bāl (36), GRH bāl (36), PRY bal (36), MAI kes (185), KCH vār (36), MEW bāl (36), WGD wāl (36), BNJ laṭṭā (672)

36a. HAIR: DKH kes (185), SND kes (185), BNG keš (185), KUL cōḍha (308), MND kes (185)

37. HAND: HND hāth (37), DKH hāt (37), PNJ hatth (37), PTH hath (37), HNK hath (37), GJR hath (37), DGR hatth (37), LHD hatth (37), SND hathu (37), RAJ hāth (37), GUJ hāth (37), MAR hāt (37), BNG hat (37), ASS hat (37), NEP hāt (37), SNG ata (37), MAL aiy (37), KOT hatth (37), HIM āth (37), KUL hōth (37), MND hāth (37), ORY hāta (37), AWD hāth (37), KUM hāth (37), ROM vast (37), KNK hāt (37), DUM hat (37), BRJ hāt (37), GRH hāt(h) (37), PRY hat (37), MAI hāth (37), KCH hath (37), MEW hāt (37), WGD at (37), BNJ hāt (37)

38. HEAD: HND sir (38), DKH sir (38), PNJ sir (38), PTH sir (38), HNK sir (38), GJR sir (38), DGR sir (38), LHD sir (38), SND siru (38), RAJ māthau (186), GUJ māthū (186), MAR đokē (254), BNG matha (186), ASS mur (338), NEP munṭo (338), SNG hisa (38), MAL is (38), KOT mūṅḍ (338), HIM mūṅḍ (338), KUL sir (38), MND sir (38), ORY muṅḍa (338), AWD mūr (338), KUM muṅḍo (338), ROM šero (38), KNK matte (186), DUM šuṭo (38), BRJ mātho (186), GRH muṅḍ (38), PRY sar (38), MAI māth (186), KCH đogo (254), MEW māth (186), WGD monḍ (338), BNJ mātō (338)

38a. HEAD: SND matho (186), GUJ ɖokū (254), BNG šir (38), ASS xir (38), NEP sir (38), KOT šir (38), HIM sir (38), MND mūṇḍ (338), AWD kapār (581), KUM sir (38), GRH sir (38), MAI mūr (338), KCH matho (186) MEW sir (38)

39. HEAR: HND sun- (39), DKH sun- (39), PNJ suṇ- (39), PTH suṇ- (39), HNK suṇ- (39), GJR suṇ- (39), DGR sun- (39), LHD suṇ- (39), SND b'udh- (187), RAJ suṇ- (39), GUJ sām̄bhaḷ- (255), MAR aik- (275), BNG šon- (39), ASS xun- (39), NEP sun- (39), SNG asa- (39), MAL iv- (39), KOT šuṇ- (39), HIM suṇ- (39), KUL šuṇ- (39), MND suṇ- (39), ORY šuṇ- (39), AWD sun- (39), KUM suṇ- (39), ROM šun- (39), KNK aik- (275), DUM paruj- (-17), BRJ sun- (39), GRH suṇ- (39), PRY sun- (39), MAI sun- (39), KCH soṇ- (39), MEW suṇ- (39), WGD āyak- (275), BNJ sām̄aḷ- (255)

40. HEART: HND dil (-3), DKH hiyā (161), PNJ dil (-2), PTH kalejā (143), HNK dil (-2), GJR kāljo (143), DGR dil (-1), LHD hā (161), SND hō (161), RAJ hīyau (161), GUJ dil (-2), MAR hRday (-6), BNG rhitpiṇḍo (-6), ASS hiya (161), NEP muṭu (365), SNG laya (161), MAL hīy (161), KOT hio (161), HIM kāl̄jā (143), KUL dil (-4), MND dil (-1), ORY hrudaya (-5), AWD dil (-4), KUM hiyo (161), ROM ilo (161), KNK kāl̄ij (143), DUM ya (161), BRJ hiruday (-3), GRH jikuru (655), PRY hik (161), MAI hia (161), KCH hīyo (161), MEW hiye (161), WGD dil (-3), BNJ dal (-1)

40a. HEART: RAJ kāl̄jau (143), ASS koliza (143), PRY kilijo (143), MEW kāl̄ejā (143)

41. HORN: HND sīṅg (40), DKH sing (40), PNJ siṅg (40), PTH sing (40), HNK siṅg (40), GJR siṅg (40), DGR siṅg (40), LHD siṅg (40), SND siṅg (40), RAJ sīg (40), GUJ siṅgḍū (40), MAR šing (40), BNG šing (40), ASS xing (40), NEP sīṅg (40), SNG (h)anga (40), MAL daḷu (537), KOT šing (40), HIM siṅg (40), KUL sīṅgh (40), MND sīṅg (40), ORY šinga (40), AWD sīṅi (40), KUM siṅg (40), ROM šing (40), KNK šing (40), DUM šiṅ (40), BRJ sīṅg (40), GRH siṅ (40), PRY šax (-7), MAI sīg (40), KCH singh (40), MEW sīh (40), WGD hengṛo (40), BNJ singg (40)

42. I: HND maī (41), DKH maī (41), PNJ maī (41), PTH maī (41), HNK mẽ (41), GJR hũ (41), DGR aũ (41), LHD maī (41), SND āũ (41), RAJ hũ (41), GUJ hũ (41), MAR mī (41), BNG ami (309), ASS moi (41), NEP ma (41), SNG mama (41), MAL ma (41), KOT mũ (41), HIM aũ (41), KUL hāw (41), MND hāu (41), ORY mui (41), AWD mahũ (41), KUM mẽ (41), ROM me (41), KNK āv (41), DUM u (41), BRJ me (41), GRH mi (41), PRY me (41), MAI ham (309), KCH āũ (41), MEW mũ (41), WGD hu (41), BNJ ma (41)

43. KILL: HND mār- (42), DKH mār- (42), PNJ mār- (42), PTH mār- (42), HNK mār- (42), GJR mār- (42), DGR mār- (42), LHD mār- (42), SND mār- (42), RAJ mār- (42), GUJ mār- (42), MAR ṭhār mār- (42), BNG mere phel- (42), ASS mār- (42), NEP mār- (42), SNG mar- (42), MAL mara- (42), KOT mar- (42), HIM mār- (42), KUL mār- (42), MND mār- (42), ORY mār- (42), AWD mār- (42), KUM mār- (42), ROM mudar- (42), KNK mār- (42), DUM mār- (42), BRJ mār- (42), GRH mār- (42), PRY mar- (42), MAI mār- (42), KCH mār- (42), MEW mār- (42), WGD mār- (42), BNJ mār nāk- (42)

44. KNEE: HND ghuṭnā (43), DKH gurgā (122), PNJ goḍā (122), PTH goḍā (122), HNK jannū (144), GJR janũ (144), DGR goḍḍā (122), LHD goḍā (122), SND goḍ'o (122), RAJ goḍau (122), GUJ ghūṭaṇ (122), MAR guḍhā (122), BNG hāṭu (225), ASS āṭhu (225), NEP ghunḍo (43), SNG daṇa (144), MAL kakū (538), KOT dzaṇu (144), HIM jaṇṇū (144), KUL jānnhu (144), MND jānnhu (144), ORY āṭhu (225), AWD ? (-5), KUM ghūro (43), ROM čang (597), KNK dimbi

(613), DUM kuṭā (-18), BRJ ? (-4), GRH ghūḍu (43), PRY janu (144), MAI ṭhehun (636), KCH gūḍo (122), MEW ghuṭṇo (43), WGD ? (-4), BNJ goḍo (122)

44a. KNEE: HNK guḍḍar (122), DGR jāhnnū (144), MND goḍḍā (122)

45. KNOW: HND jān- (44), DKH jān- (44), PNJ jān- (44), PTH jān- (44), HNK jān- (44), GJR jān- (44), DGR jān- (44), LHD jān- (44), SND jān- (44), RAJ jān- (44), GUJ jān- (44), MAR jān- (44), BNG jan- (44), ASS zan- (44), NEP jān- (44), SNG dan- (44), MAL danna- (44), KOT dzaṅ- (44), HIM jān- (44), KUL jān- (44), MND jān- (44), ORY jān- (44), AWD jān- (44), KUM jān- (44), ROM žan- (44), KNK jān- (44), DUM ? (-19), BRJ jān- (44), GRH jaṅ- (44), PRY samj- (662), MAI jān- (44),

KCH jān- (44), MEW jān- (44), WGD jan- (44), BNJ jaṅ- (44)

46. LEAF: HND pattā (45), DKH pattā (45), PNJ pattar (45), PTH pattā (45), HNK pattar (45), GJR patar (45), DGR pattar (45), LHD pātar (45), SND panu (189), RAJ pattau (45), GUJ pāḍṛū (189), MAR pān (189), BNG pata (45), ASS pat (45), NEP pāt (45), SNG pata (45), MAL faiy (45), KOT pāč (45), HIM pattā (45), KUL pōttrə (45), MND pattar (45), ORY patra (-6), AWD pātā (45), KUM pāt (45), ROM patrin (45), KNK pān (189), DUM poṭ (-20), BRJ pattī (45), GRH pathala (45), PRY barg (-8), MAI pāt (45), KCH pann (189), MEW pattā (45), WGD patto (45), BNJ pālo (570)

46a. LEAF: RAJ pānṛau (189), GUJ pataru (45), KNK pāllo (570)

47. LIE: HND leṭ- (46), DKH paṛ- (47), PNJ leṭ- (46), PTH leṭ- (46), HNK lamā ho- (479), GJR lamū ho- (479), DGR leṭ- (46), LHD leṭ- (46), SND leṭ- (46), RAJ loṭ- (508), GUJ leṭ- (46), MAR nij- (277), BNG šue thak- (310), ASS pari thōk- (47), NEP leṭi rah- (46), SNG hansivī hiṭin- (407), MAL doguhed- (408), KOT pōṛ- (47), HIM leṭ- (46), KUL leṭ- (46), MND lambā pai- (47), ORY šo- (310), AWD paūrḥ- (582), KUM paṛ- (47), ROM pašlo- (278), KNK āḍ poḍ- (47) DUM jek- (-21), BRJ leṭ- (46), GRH paṛ- (47), PRY paṛ- (47), MAI leṭ- (46), KCH leṭ- (46), MEW leṭ- (46), WGD so- (310), BNJ āḍo paḍ- (47)

47a. LIE: HND paṛ- (47), PNJ lamme pai- (47), HNK pe- (47), GJR pe- (47), LHD pē- (47), SND pav- (47), RAJ paṛ- (47), GUJ paḍ- (47), MAR pasarlelē as- (278), NEP sut- (310), SNG digāvī hiṭin- (408), MAL oiy- (310), KOT leṭ- (46), KUL lōmma pōw- (47), MND leṭ- (46), MEW so- (310)

48. LIVER: HND kalejā (48), DKH kalejā (48), PNJ kalejā (48), PTH jigār (-2), HNK jigrā (-3), GJR kāljō (48), DGR kalejā (48), LHD kalejī (48), SND jero (190), RAJ kāljau (48), GUJ kalejā (48), MAR yakRt (-7), BNG jōkrit (-7), ASS zōkrit (-3), NEP kalejo (48), SNG akmāva (410), MAL mey (539), KOT kaldzō (-1), HIM kāljā (48), KUL kālja (48), MND kāljā (48), ORY kalijā (48), AWD karej (48), KUM phyātto (588), ROM buko (496), KNK kālij (48), DUM kaulā (625), BRJ haṅḍe (631), GRH ? (-3), PRY kiliji (48), MAI yakRt (-4), KCH jigār (-4), MEW gurdā (-3), WGD kaljā (48), BNJ koljo (48)

49. LONG: HND lambā (49), DKH lāmbā (49), PNJ lammā (49), PTH lammā (49), HNK lammā (49), GJR lamū (49), DGR lammā (49), LHD lambā (49), SND ḍigho (191), RAJ lambau (49), GUJ lāmbū (49), MAR lāb (49), BNG lōmba (49), ASS dighōl (191), NEP lamo (49), SNG dik (191), MAL digu (191), KOT lambō (49), HIM lambā (49), KUL lōmma (49), MND lambā (49), ORY lambā (49), AWD lambā (49), KUM lambo (49), ROM lungo (-7), KNK dīg (191), DUM jigā (-22), BRJ lambo (49), GRH lambu (49), PRY lamo (49), MAI lambā (49), KCH lamo (49), MEW lambo (49), WGD lambo (49), BNJ lāmbō (49)

49a. LONG: SND lambo (49), ASS lōmba (49)

50. LOUSE: HND jũ (50), DKH jũ (50), PNJ jũ (50), PTH jũã (50), HNK jũ (50), GJR jũ (50), DGR jũ (50), LHD juã (50), SND jũa (50), RAJ jũ (50), GUJ jũ (50), MAR ū (50), BNG ukuṇ (311), ASS okoṇi (311), NEP jumrā (50), SNG ukuṇā (311), MAL ukunu (311), KOT jũ (50), HIM jũ (50), KUL jũ (50), MND jũ (50), ORY ukuṇi (311), AWD juã (50), KUM jũo (50), ROM žuv (50), KNK ū (50), DUM jūya (50), BRJ jũl (50), GRH jũ (50), PRY jũ (50), MAI jũ (50), KCH jũ (50), MEW jũ (50), WGD jũ (50), BNJ jũ (50)

51. MAN: HND mard (-4), DKH purus (123), PNJ purakh (-3), PTH jaṇā (162), HNK jaṇā (162), GJR janū (162), DGR mardmāhnū (192), LHD murs (123), SND māṇuhū (192), RAJ mātī (509), GUJ bhāyḍo (256), MAR puruṣ (-8), BNG manuṣ (192), ASS munih (192), NEP logne-manche (192), SNG pirimiyā (411), MAL firihen (123), KOT mōrd (-2), HIM lok (480), KUL mōrd (-5), MND mard (-2), ORY puruṣa (-7), AWD mansedhū (192), KUM meīs (192), ROM murš (123), KNK manišu (192), DUM maniš (192), BRJ ādmī (-5), GRH ādmi (-4), PRY anmi (-9), MAI purukh (-5), KCH māṛū (192), MEW māṇas (192), WGD admī (-5), BNJ māṭi (509)

51a. MAN: BNJ māṇas (192)

52. MANY: HND bahut (51), DKH bhaut (51), PNJ bahut (51), PTH bahũ (51), HNK muc (481), GJR muc (481), DGR matā (497), LHD bahū (51), SND ghaṇā (193), RAJ ghaṇau (193), GUJ ghano (193), MAR phār (280), BNG bōhu (51), ASS bōhuto (51), NEP dherai (367), SNG bohō (51), MAL baivaru (51), KOT bōto (51), HIM bot (51), KUL bōhu (51), MND bahut (51), ORY bahuta (51), AWD bahut (51), KUM bhaut (51), ROM but (51), KNK anek (-5), DUM buṭ (51), BRJ bhaut (51), GRH bahut (51), PRY bot (51), MAI bahut (51), KCH baurā (51), MEW bhot (51), WGD ghanā (193), BNJ ghaṇo (193)

52a. MANY: HNK bahũ (51), GUJ bahu (51), MAL gina (193), MND mate (497), KCH ghaṇo (193)

53. MEAT: HND gošt (-5), DKH mās (124), PNJ mās (124), PTH mās (124), HNK mās (124), GJR mās (124), DGR mās (124), LHD gošt (-3), SND māsu (124), RAJ mās (124), GUJ mās (124), MAR mās (124), BNG mākšō (124), ASS mōngōh (124), NEP māsu (124), SNG mas (124), MAL mas (124), KOT mas (124), HIM mās (124), KUL mās (124), MND mās (124), ORY māsēsa (124), AWD māsu (124), KUM māsu (124), ROM mas (124), KNK mās (124), DUM mos (124), BRJ mās (124), GRH sikār (-5), PRY gōš (-10), MAI mās (124), KCH mas (124), MEW mās (124), WGD sīsī (662), BNJ mās (124)

54. MOON: HND cād (52), DKH cād (52), PNJ cann (52), PTH can (52), HNK can (52), GJR can (52), DGR cann (52), LHD candar (52), SND caṇḍu (52), RAJ candau (52), GUJ cādo (52), MAR cād (52), BNG cand (52), ASS zon (340), NEP jūn (340), SNG haṇḍa (52), MAL handu (52), KOT dzūṇ (340), HIM candramā (-2), KUL joth (340), MND candarmā (-3), ORY candramā (-8), AWD candā (52), KUM candā (52), ROM šonuto (340), KNK candru (52), DUM čonč (52), BRJ candā (52), GRH jūn (340), PRY čan (52), MAI candā (52), KCH candhar (52), MEW cād (52), WGD san (52), BNJ cāndā (52)

54a. MOON: KUM jūn (340)

55. MOUNTAIN: HND pahār (53), DKH p(ah)ār (53), PNJ pahār (53), PTH pahārī (53), HNK pahār (53), GJR pahār (53), DGR phār (53), LHD pahār (53), SND pahaṛu (53), RAJ pahār (53), GUJ pahāḍ (53), MAR pahāḍ (53), BNG pahaṛ (53), ASS pahar (53), NEP pahāḍ (53), SNG kanda (-4), MAL farubada (-2), KOT pār (53), HIM pār (53), KUL pahāḍ (53), MND pār (53), ORY parbata (-9), AWD pahār (53), KUM pahār (53), ROM mīnt'a (-8), KNK parvatu (-6),

DUM čišā (-23), BRJ pahār (53), GRH pār (53), PRY paṛ (53), MAI parvat (-6), KCH pahārī (53), MEW pahār (53), WGD māro (663), BNJ pād (53)

56. MOUTH: HND mūh (54), DKH mū (54), PNJ mūh (54), PTH mūh (54), HNK mūh (54), GJR mū (54), DGR mū (54), LHD mūh (54), SND mūhū (54), RAJ mūh (54), GUJ mhō (54), MAR tōḍ (281), BNG mukh (54), ASS mukh (54), NEP mukh (54), SNG muva (54), MAL anga (540), KOT mū (54), HIM mū (54), KUL mūh (54), MND mū (54), ORY muhā (54), AWD muh (54), KUM mukh (54), ROM muy (54), KNK tōḍ (281), DUM kašā (626), BRJ mūhte (54), GRH gicco (657), PRY mo (54), MAI muh (54), KCH mō (54), MEW mūh (54), WGD muḍo (54), BNJ muḍo (54)

57. NAME: HND nām (55), DKH nāō (55), PNJ nā (55), PTH nā (55), HNK nā (55), GJR nā (55), DGR nā (55), LHD nā (55), SND nālo (55), RAJ nāv (55), GUJ nām (55), MAR nāv (55), BNG nam (55), ASS nam (55), NEP nāū (55), SNG nama (55), MAL nan (55), KOT nāū (55), HIM nāū (55), KUL nā (55), MND nā (55), ORY nāma (55), AWD nāu (55), KUM nou (55), ROM nav (55), KNK nāv (55), DUM nom (55), BRJ nām (55), GRH nau (55), PRY na (55), MAI nām (55), KCH nālo (55), MEW nāv (55), WGD nām (55), BNJ nām (55)

58. NECK: HND galā (56), DKH galā (56), PNJ galā (56), PTH dhaun (125), HNK galā (56), GJR gaḷ (56), DGR dhaun (125), LHD giccī (196), SND g'icī (196), RAJ nār (514), GUJ ḍokī (258), MAR mān (229), BNG gōla (56), ASS gōl (56), NEP galām (56), SNG gela (312), MAL karu (413), KOT keṛi (228), HIM kyārī (228), KUL kyāḍi (228), MND kaiṛī (228), ORY beka (674), AWD gaṭai (197), KUM ghāṭī (312), ROM korr (599), KNK gaḷo (56), DUM koṭā (599), BRJ nāri (514), GRH dhaun (125), PRY gardan (-11), MAI dhonā (125), KCH giccī (196), MEW galo (56), WGD gābro (664), BNJ gaḷā (56)

58a. NECK: PNJ dhaun (125), HNK gāṭā (197), GJR gāṭo (197), DGR gāṭā (197), SND g'āṭo (197), RAJ dhūṇ (125), GUJ galcī (56), MAR gaḷā (56), BNG ghar (312), NEP ghāṭī (312), SNG kara (413), KOT gōḷ (56), KUL gōḷa (56), MND gōḷ (56), AWD gar (56), KUM galo (56), GRH gaḷā (56), MAI ghār (312), KCH niṛī (514), BNJ naḷḍi (514)

59. NEW: HND nayā (57), DKH navā (57), PNJ navā (57), PTH navā (57), HNK navvā (57), GJR nawū (57), DGR namā (57), LHD navā (57), SND naō (57), RAJ navau (57), GUJ navū (57), MAR navā (57), BNG nōbo (57), ASS nō (57),

NEP nayā (57), SNG nava (57), MAL au (541), KOT nōuwō (57), HIM navā (57), KUL nōwwā (57), MND navā (57), ORY nuā (57), AWD naykā (57), KUM nayyo (57), ROM nevo (57), KNK nave (57), DUM namā (57), BRJ nayo (57), GRH nayū (57), PRY navo (57), MAI nab (57), KCH nāū (57), MEW nawā (57), WGD navo (57), BNJ nawo (57)

59a. NEW: SNG alut (541)

60. NIGHT: HND rāt (58), DKH rāt (58), PNJ rāt (58), PTH rāt (58), HNK rāt (58), GJR rāt (58), DGR rāt (58), LHD rāt (58), SND rāti (58), RAJ rāt (58), GUJ rāt (58), MAR rāt (58), BNG rat (58), ASS rati (58), NEP rāt (58), SNG rāya (58), MAL reygandu (58), KOT rāč (58), HIM rattī (58), KUL rāt (58), MND rāt (58), ORY rāti (58), AWD rāti (58), KUM rāt (58), ROM rat (58), KNK rāti (58), DUM thop (-24), BRJ rāt (58), GRH rāt (58), PRY rat (58), MAI rāti (58), KCH rāt (58), MEW rāt (58), WGD rat (58), BNJ rāt (58)

61. NOSE: HND nāk (59), DKH nāk (59), PNJ nakk (59), PTH nak (59), HNK nak (59), GJR nak (59), DGR nakk (59), LHD nakk (59), SND naku (59), RAJ nāk (59), GUJ nāk (59), MAR nāk (59), BNG nak (59), ASS nak (59), NEP nāk (59), SNG nahaya (59), MAL neyfaiy (59), KOT nāk

(59), HIM nāk (59), KUL nāk (59), MND nāk (59), ORY nāka (59), AWD nāki (59), KUM nāk (59), ROM nak (59), KNK nāk (59), DUM nok (59), BRJ nāk (59), GRH nāk (59), PRY nak (59), MAI nāk (59), KCH nakk (59), MEW nāk (59), WGD nakoṛo (59), BNJ nāk (59)

62. NOT: HND na (60), DKH nakko (60), PNJ nā (60), PTH na (60), HNK nā (60), GJR na (60), DGR nā (60), LHD na (60), SND na (60), RAJ nā (60), GUJ nā (60), MAR na (60), BNG na (60), ASS no (60), NEP na (60), SNG nā (60), MAL nu (60), KOT na (60), HIM na (60), KUL nāy (60), MND na (60), ORY nā (60), AWD nāhī (60), KUM ne (60), ROM na (60), KNK nhāyī (60), DUM ni (60), BRJ nāye (60), GRH na (60), PRY na (60), MAI nahi (60), KCH na (60), MEW ni (60), WGD nā (60), BNJ ni (60)

63. ONE: HND ek (61), DKH yek (61), PNJ ik (61), PTH hik (61), HNK hik (61), GJR ek (61), DGR ikk (61), LHD ek (61), SND hiku (61), RAJ (h)ik (61), GUJ ek (61), MAR ek (61), BNG ek (61), ASS ek (61), NEP ek (61), SNG eka (61), MAL ekeh (61), KOT ēk (61), HIM ek (61), KUL yek (61), MND ek (61), ORY eka (61), AWD yak (61), KUM ek (61), ROM (y)ek (61), KNK ek (61), DUM ek (61), BRJ ek (61), GRH ek (61), PRY yek (61), MAI ek (61), KCH hikṛo (61), MEW ek (61), WGD ek (61), BNJ ek (61)

64. PERSON: HND janā (163), DKH janā (163), PNJ jaṇā (163), PTH ādmī (-3), HNK lok (461), GJR jaṇū (163), DGR māhnū (126), LHD bandā (-4), SND jaṇo (163), RAJ jaṇau (163), GUJ māṇas (126), MAR māṇūs (126), BNG manuś (126), ASS manuh (126), NEP mānis (126), 126 SNG minihā (126), MAL mēhā (126), KOT dzəṇo (163), HIM māṇū (126), KUL mhāṇu (126), MND māṇu (126), ORY maṇiša (126), AWD manaī (126), KUM jaṇ (163), ROM manuś (126), KNK maniśu (126), DUM maniś (126), BRJ mānus (126), GRH maṇāli (126), PRY jaṇo (163), MAI manukh (-7), KCH māṛū (126), MEW ādamī (-4), WGD lak (461), BNJ jaṇā (163)

64a. PERSON: DKH mānūs (126), DGR janā (163), RAJ māṇas (126), BNG lok (461), KOT maṇčh (126), MND jaṇā (163), KUM meīs (126), GRH jaṇ (163), BNJ māṇas (126)

65. RAIN: HND mēh (62), DKH barsāt (127), PNJ mīh (62), PTH baddal (127), HNK baddal (127), GJR badal (127), DGR barkhā (-2), LHD mīh (62), SND mīhu (62), RAJ meh (62), GUJ varsād (127), MAR pāūs (127), BNG badal (127), ASS bərc̣khuṇ (127), NEP jharī (369), SNG vāssa (127), MAL vārey (127), KOT pāṇī (447), HIM barkhā (-3), KUL gāś (663), MND barkhā (-4), ORY barša (-10), AWD barkhā (-4), KUM me (62), ROM beršind (127), KNK pāvsu (127), DUM orp (661), BRJ meg (62), GRH barkhā (-6), PRY mi (62), MAI barkhā (-8), KCH meh (62), MEW meh (62), WGD mehā (62), BNJ pāṇi (447)

65a. RAIN: DKH mhyūn (62), HNK mīh (62), GJR mīh (62), SND vas (127), KCH varsād (127)

66. RED: HND lāl (63), DKH lāl (63), PNJ lāl (63), PTH sūā (475), HNK rattā (128), GJR rato (128), DGR lāl (63), LHD rattā (128), SND lālu (63), RAJ rātau (128), GUJ lāl (63), MAR tābḍā (282), BNG lal (63), ASS rəṅga (128), NEP rāto (128), SNG ratu (128), MAL raiy (128), KOT ratto (128), HIM lāl (63), KUL lāl (63), MND lāl (63), ORY lāl (63), AWD lāl (63), KUM lāl (63), ROM lolo (63), KNK tābḍe (282), DUM loyā (63), BRJ lāl (63), GRH lāl (63), PRY lal (63), MAI lāl (63), KCH ratto (128), MEW lāl (63), WGD lāl (63), BNJ rātḍo (128)

66a. RED: DKH ratlā (128), PNJ rattā (128), HNK lāl (63), GJR lāl (63), LHD lāl (63), SND rato (128), RAJ lāl (63), GUJ rātū (128), ASS lal (63), KOT lāl (63), MAI rāt (128)

67. ROAD: HND saṛak (64), DKH bāt (164), PNJ saṛak (64), PTH saṛak (64), HNK širrak (64), GJR saṛak (64), DGR batt (164), LHD vāt (164), SND saṛak (64), RAJ saṛak (64), GUJ saḍak (64),

MAR sađak (64), BNG p̄oth (-8), ASS bađ (164), NEP b̄ađo (164), SNG maga (415), MAL magu (415), KOT b̄at (164), HIM b̄at (164), KUL b̄ott (164), MND peynda (448), ORY řaraka (64), AWD b̄at (164), KUM b̄ađo (164), ROM drom (-9), KNK m̄argu (-7), DUM pan̄a (-25), BRJ sarak (64), GRH b̄ađu (164), PRY ra (-12), MAI path (-9), KCH v̄at (164), MEW sarak (64), WGD w̄at (164), BNJ w̄at (164)

67a. ROAD: KOT p̄eđd (448)

68. ROOT: HND jař (65), DKH jař (65), PNJ jař (65), PTH jař (65), HNK jařh (65), GJR jař (65), DGR jařh (65), LHD jařř (65), SND jařu (65), RAJ jař (65), GUJ m̄uļ (259), MAR m̄uļ (259), BNG řik̄oř (313), ASS xipa (343), NEP jařa (65), SNG mula (259), MAL m̄o (259), KOT dz̄ořh (65), HIM jař (65), KUL j̄al̄ađa (65), MND jađ (65), ORY m̄uļa (259), AWD jari (65), KUM jař (65), ROM koreno (-10), KNK m̄uļ (259), DUM moli (-26), BRJ jař (65), GRH jař (65), PRY rešo (-13), MAI jaři (65), KCH ? (-5), MEW jař (65), WGD m̄ur (259), BNJ jađ (65)

68a. ROOT: GUJ jađ (65), MEW m̄uļ (259)

69. ROUND: HND gol (66), DKH gol (66), PNJ gol (66), PTH gol (66), HNK gol (66), GJR gol (66), DGR gol (66), LHD gol (66), SND golu (66), RAJ goļ (66), GUJ gol (66), MAR gol (66), BNG gol (66), ASS gol (66), NEP b̄ađuło (283), SNG vađa (283), MAL vah (283), KOT ḡoļ (66), HIM gol (66), KUL ḡoļ (66), MND ḡoļ (66), ORY gola (66), AWD gol (66), KUM golo (66), ROM rotato (-11), KNK ur̄uđ (-8), DUM ? (-27), BRJ gol (66), GRH gulgaņđo (66), PRY ? (-14), MAI gol (66), KCH gird (-6), MEW ? (-5), WGD ? (-6), BNJ gol (66)

69a. ROUND: MAR v̄ađuļa (283)

70. SAND: HND ret (67), DKH b̄al̄u (68), PNJ ret (67), PTH ret (67), HNK ret (67), GJR ret (67), DGR ret̄a (67), LHD ret (67), SND ret̄i (67), RAJ b̄al̄u (68), GUJ ret̄i (67), MAR ret̄i (67), BNG bali (68), ASS bali (68), NEP ret (67), SNG v̄ali (67), MAL veli (68), KOT bađu (68), HIM ball̄u (68), KUL r̄et (67), MND b̄allu (68), ORY b̄ali (68), AWD b̄ar̄u (68), KUM balw̄a (68), ROM kiřay (-12), KNK rev (67), DUM b̄ali (68), BRJ b̄al̄u (68), GRH b̄alo (68), PRY ? (-15), MAI b̄al̄u (68), KCH ? (-7), MEW b̄al̄u (68), WGD ret (67), BNJ retu (67)

70a. SAND: HND b̄al̄u (68), PNJ b̄al̄u (68), SND v̄ar̄i (68), RAJ ret (67), GUJ v̄alu (68), MAR v̄al̄u (68), NEP b̄aluv̄a (68), HIM ret (67), MND ret (67), AWD ret (67), MEW ret (67)

71. SAY: HND kah- (69), DKH ka- (69), PNJ kah- (69), PTH ākh- (468), HNK āχ- (468), GJR keh- (69), DGR ākh- (468), LHD ākh- (468), SND al- (166), RAJ kah- (69), GUJ kah- (69), MAR mhaņ- (231), BNG k̄oh- (69), ASS k̄o- (69), NEP bhan- (231), SNG kiya- (69), MAL buna- (231), KOT kiu- (69), HIM bol- (70), KUL b̄oļ-(70), MND b̄oļ- (70), ORY kah- (69), AWD kah- (69), KUM k̄u- (69), ROM phen- (231), KNK mh̄aņ- (231), DUM mun- (231), BRJ keh- (69), GRH bval- (70), PRY ke- (69), MAI kah- (69), KCH co- (657), MEW kah- (69), WGD kah- (69), BNJ ke- (69)

71a. SAY: HND bol- (70), DKH bol- (70), PNJ ākh- (468), PTH bol- (70), HNK bol- (70), GJR bol- (70), DGR bol- (70), LHD bol- (70), SND b'ol- (70), RAJ bol- (70), GUJ bol- (70), MAR bol- (70), BNG b̄oļ- (70), ASS bol- (70), NEP bol- (70), MAL kiya- (69), KOT bol- (70), ORY bol- (70), AWD bol- (70), KUM bol- (70), MAI bol- (70), KCH bol- (70)

72. SEE: HND dekh- (71), DKH dekh- (71), PNJ dekh- (71), PTH tak- (476), HNK deχ- (71), GJR dekh- (71), DGR dikkh- (71), LHD đekh- (71), SND đ'is- (129), RAJ dekh- (71), GUJ dekh- (71), MAR p̄ah- (198), BNG dekh- (71), ASS dekh- (71), NEP dekh- (71), SNG daki- (71), MAL duř- (129), KOT dekh- (71), HIM dekh- (71), KUL d̄ekh- (71), MND dekh- (71), ORY dekh- (71),

AWD dekh- (71), KUM dekh- (71), ROM dikh- (71), KNK paļe- (617), DUM dekh- (71), BRJ dek- (71), GRH dekh- (71), PRY dekh- (71), MAI dekh- (71), KCH đis- (129), MEW dekh- (71), WGD dekh- (71), BNJ dek- (71)

72a. SEE: PNJ diṭh- (129), HNK diṭṭh- (129), GJR diṭṭh- (129), DGR diṭṭh- (129), LHD diṭṭh- (129), SND pas- (198), RAJ diṭh- (129), GUJ diṭh- (129), MAL deke- (71), KNK đis- (129), BNJ diṭ- (129)

73. SEED: HND bij (72), DKH biṅ (72), PNJ bi (72), PTH bi (72), HNK bi (72), GJR bij (72), DGR bi (72), LHD bij (72), SND b'iju (72), RAJ bij (72), GUJ bi (72), MAR bi (72), BNG bij (72), ASS biz (72), NEP biū (72), SNG bijaya (-5), MAL oh (543), KOT bīdz (72), HIM bij (72), KUL bejja (72), MND biu (72), ORY bija (72), AWD biyā (72), KUM byū (72), ROM semınca (-13), KNK bi (72), DUM yono (-28), BRJ bij (72), GRH bij (72), PRY ? (-16), MAI bij (72), KCH bijj (72), MEW bihan (72), WGD ? (-7), BNJ bijā (72)

74. SIT: HND baiṭh- (73), DKH baiṭ(h)- (73), PNJ baiṭh- (73), PTH beh- (73), HNK baiṭh- (73), GJR bais- (73), DGR beh- (73), LHD beh- (73), SND vih- (73), RAJ baiṭh- (73), GUJ bes- (73), MAR bas- (73), BNG bōš- (73), ASS bōh- (73), NEP basi rah- (73), SNG iṅdagan hiṭi- (418), MAL inn- (418), KOT bēš- (73), HIM baiṭh- (73), KUL bešš- (73), MND beyṭh- (73), ORY baš- (73), AWD baiṭh- (73), KUM beiṭh- (73), ROM beš- (73), KNK bais- (73), DUM beš- (73), BRJ beṭh- (73), GRH baiṭh- (73), PRY beṭ- (73), MAI baiṭh- (73), KCH ve- (73), MEW bēṭh- (73), WGD way- (73), BNJ bes- (73)

75. SKIN: HND camṛī (74), DKH camṛī (74), PNJ camṛī (74), PTH camṛī (74), HNK camṛā (74), GJR camṛī (74), DGR camm (74), LHD camṛī (74), SND camṛī (74), RAJ khāl (75), GUJ camḍī (74), MAR camḍī (74), BNG camṛa (74), ASS sal (348), NEP chālā (348), SNG hama (74), MAL han (74), KOT camṛī (74), HIM camṛī (74), KUL camḍī (74), MND khāl (75), ORY camaṛā (74), AWD khāl (75), KUM khāl (75), ROM morči (-14), KNK cām (74), DUM čom (74), BRJ cam (74), GRH camṛī (74), PRY pōst (-17), MAI camṛā (74), KCH camṛī (74), MEW khāl (75), WGD samḍī (74), BNJ khālḍi (75)

75a. SKIN: HND khāl (75), DKH khararī (75), PNJ khall (75), DGR khall (75), LHD khal (75), SND khal (75), RAJ cāmṛī (74), GUJ khāl (75), MAR kātḍī (284), KUL khōlḍi (75), MND cāmṛī (74), AWD camṛā (74), KUM cām (74), KNK sālī (348), MAI khāl (75)

76. SLEEP: HND so- (76), DKH so- (76), PNJ so- (76), PTH seh- (76), HNK sē- (76), GJR so- (76), DGR sau- (76), LHD summ- (76), SND sumh- (76), RAJ sūv- (76), GUJ su- (76), MAR nij- (285), BNG ghuma- (314), ASS xo- (76), NEP sut- (76), SNG nid- (199), MAL nid- (199), KOT sut- (76), HIM sut- (76), KUL so- (76), MND sau- (76), ORY šo- (76), AWD so- (76), KUM nīn kar- (199), ROM so- (76), KNK nīd- (199), DUM so- (76), BRJ sow- (76), GRH sī- (76), PRY so- (76), MAI sut- (76), KCH sum- (76), MEW so- (76), WGD so- (76), BNJ so- (76)

76a. SLEEP: SND ninḍ' kar- (199), NEP nidau- (199), KUM sit- (76)

77. SMALL: HND choṭā (77), DKH lhorā (470), PNJ choṭā (77), PTH nikkā (130), HNK nikkā (130), GJR nikko (130), DGR lauḥkā (470), LHD nikkā (130), SND naṅḍho (200), RAJ nain (261), GUJ nānū (261), MAR lahān (261), BNG choṭo (77), ASS xoru (349), NEP sāno (261), SNG kuḍā (420), MAL kuḍa (420), KOT choṭṭo (77), HIM nikkājā (130), KUL choṭṭa (77), MND halkā (565), ORY choṭa (77), AWD choṭ (77), KUM nāno (261), ROM cigno (-15), KNK sān (261), DUM čunā (-29), BRJ choṭṭo (77), GRH chvaṭṭu (77), PRY nuko (130), MAI choṭ (77), KCH niṅḍho (200), MEW choṭo (77), WGD nēḍho (200), BNJ nānkyā (261)

77a. SMALL: DKH choṭā (77), PNJ nikkā (130), HNK choṭā (77), GJR choṭo (77), DGR nikkā (130), LHD choṭā (77), SND choṭo (77), RAJ choṭau (77), HIM choṭā (77), AWD nanhakā (261), KUM choṭo (77), MAI nānh (261)

78. SMOKE: HND dhuā (78), DKH dhuvā (78), PNJ dhūā (78), PTH dhūā (78), HNK dhūvā (78), GJR dhūā (78), DGR dhū (78), LHD dhū (78), SND dūhō (78), RAJ dhūārau (78), GUJ dhumāḍo (78), MAR dhūr (78), BNG dhōva (78), ASS dhōva (78), NEP dhuvā (78), SNG duma (78), MAL dun (78), KOT dhū (78), HIM dhūā (78), KUL dhūwā (78), MND dhū (78), ORY dhūā (78), AWD dhuwā (78), KUM dhuo (78), ROM thuv (78), KNK dhuv(o)ru (78), DUM dum (78), BRJ dhūo (78), GRH dhuā (78), PRY tōva (78), MAI dhuā (78), KCH dhūo (78), MEW dhūā (78), WGD dūā (78), BNJ ? (-2)

79. STAND: HND kharā ho- (79), DKH khar- (79), PNJ khalo- (79), PTH khal- (79), HNK khal- (79), GJR khal- (79), DGR kharo- (79), LHD khar- (79), SND bih- (201), RAJ ūbh- (262), GUJ ubhū rah- (262), MAR khaḍā as- (79), BNG dāra- (315), ASS thiyoi thok- (350), NEP ubhi- (262), SNG nāgiṭa hiṭi- (350), MAL theduve hur- (350), KOT khər- (79), HIM khar- (79), KUL khəḍ- (79), MND khar- (79), ORY thiā he- (350), AWD ṭhārḥ ho- (516), KUM ṭhar- (516), ROM terd- (516), KNK ubbī- (262), DUM cak hut- (-30), BRJ ṭhāre ho- (514), GRH ṭhāḍu ho- (514), PRY kalo ho- (79), MAI ṭhārḥ bhe- (514), KCH aubhū- (262), MEW kharo ho- (79), WGD ub- (262), BNJ hubar- (262)

79a. STAND: HNK uṭh- (350), RAJ ṭhāḍau hov- (516), GUJ khaḍū rah- (79), MAR ubhā rāh- (262)

80. STAR: HND tāra (80), DKH tāra (80), PNJ tāra (80), PTH tāra (80), HNK tāra (80), GJR tāro (80), DGR tāra (80), LHD tāra (80), SND tāro (80), RAJ tārau (80), GUJ tāro (80), MAR tāra (80), BNG tara (80), ASS tōra (80), NEP tāra (80), SNG taruva (80), MAL thari (80), KOT taro (80), HIM tāra (80), KUL tāra (80), MND tāra (80), ORY tāra (80), AWD taraī (80), KUM tāro (80), ROM čehran (-16), KNK nakṣatr (-9), DUM tāro (80), BRJ tāre (80), GRH gainu (658), PRY taro (80), MAI tāra (80), KCH tāro (80), MEW tārai (80), WGD tāro (80), BNJ tāra (80)

81. STONE: HND patthar (81), DKH pāthar (81), PNJ patthar (81), PTH baṭṭā (167), HNK baṭā (167), GJR baṭo (167), DGR baṭṭā (167), LHD patthar (81), SND pattharu (81), RAJ bhāṭau (167), GUJ patthar (81), MAR dagaḍ (286), BNG pathər (81), ASS xil (351), NEP ḍhungo (371), SNG gala (-6), MAL hila (351), KOT ḍhol (451), HIM pāthar (81), KUL pāthər (81), MND pāthar (81), ORY pathara (81), AWD pathrā (81), KUM ḍhuṇo (371), ROM barr (167), KNK patthoru (81), DUM boṭ (167), BRJ patthar (81), GRH ḍhūgu (371), PRY patar (81), MAI pāhan (677), KCH pathrī (81), MEW pāthar (81), WGD patthar (81), BNJ bhāṭā (167)

81a. STONE: HNK gaṭṭā (483), GJR pathar (81), DGR patthar (81), RAJ pātharau (81), ASS pathər (81), KUM pāthar (81), MEW bhaṭṭo (167), BNJ pattar (81)

82. SUN: HND sūraj (82), DKH sūrj (82), PNJ sūraj (82), PTH dīhū (168), HNK dī (168), GJR dī (168), DGR sūraj (82), LHD sijh (82), SND siju (82), RAJ sūraj (82), GUJ sūraj (82), MAR sūrya (-9), BNG šurjə (-9), ASS beli (352), NEP ghām (236), SNG ira (82), MAL iru (82), KOT dheṛo (168), HIM sūraj (82), KUL surjə (82), MND sūraj (82), ORY sūrya (-11), AWD suruj (82), KUM sūraj (82), ROM kham (236), KNK sūryu (-10), DUM to (628), BRJ sūraj (82), GRH ghām (236), PRY tup (646), MAI surūj (82), KCH sūraj (82), MEW sūraj (82), WGD dāṛobāwsi (168), BNJ dāḍo (168)

82a. SUN: ASS xuruz (82), MND dhairā (168), GRH sūraj (82), BNJ surjyā (82)

83. SWIM: HND tair- (83), DKH tir- (83), PNJ tar- (83), PTH tar- (83), HNK tar- (83), GJR tar- (83), DGR tar- (83), LHD tar- (83), SND tar- (83), RAJ tir- (83), GUJ tar- (83), MAR poh- (287), BNG šātōr de- (83), ASS xātor- (83), NEP tar- (83), SNG pīna- (421), MAL fath- (545), KOT tæer- (83), HIM tar- (83), KUL tōr- (83), MND tar- (83), ORY pahār- (83), AWD pāvar- (83), KUM teir- (83), ROM pliv- (-17), KNK paū- (287), DUM ? (-31), BRJ ter- (83), GRH tar- (83), PRY ? (-18), MAI tair- (83), KCH tar- (83), MEW tar- (83), WGD pōhow- (287), BNJ tar- (83)

83a. SWIM: MAR tar- (83)

84. TAIL: HND pūch (84), DKH dum (-2), PNJ pucch (84), PTH pucchaṛ (84), HNK pūchaṛ (84), GJR pūchaṛ (84), DGR pucch (84), LHD pucchaṛ (84), SND puch (84), RAJ pūchrau (84), GUJ puchdū (84), MAR šēpūt (288), BNG langul (317), ASS negur (317), NEP pucchar (84), SNG naguṭa (317), MAL nagō (317), KOT pundzhəṛ (84), HIM pūnch (84), KUL phunjit (84), MND pūch (84), ORY langula (317), AWD pūchi (84), KUM punch (84), ROM pori (601), KNK bāl (-11), DUM čipoy (288), BRJ pūc (84), GRH puchaṛu (84), PRY dum (-19), MAI pūch (84), KCH pucch (84), MEW pōch (84), WGD pochṛī (84), BNJ puncḍi (84)

84a. TAIL: MAI lāgari (317)

85. THAT: HND vah (85), DKH (v)o (85), PNJ o (85), PTH oh (85), HNK o (85), GJR vo (85), DGR oh (85), LHD o (85), SND hū (85), RAJ (v)o (85), GUJ te (203), MAR to (203), BNG o (85), ASS xi (353), NEP u (85), SNG ō(ka) (85), MAL e (289), KOT so (353), HIM se (353), KUL sō (353), MND sē (353), ORY se (353), AWD u (85), KUM u (85), ROM (k)odo (85), KNK then (203), DUM hei (289), BRJ bū (85), GRH vu (85), PRY u (85), MAI ū (85), KCH hū (85), MEW wo (85), WGD o (85), BNJ u (85)

85a. THAT: DKH ti- (203), SND ta (203), RAJ tikau (203), MAR jo (289), BNG ta- (203), ASS teō (203), NEP tyo (203), SNG ē(ka) (289), HIM vo (85), KUL te- (203), MND te- (203), ORY tāhā (203), MAI soi (353), KCH ta (203)

86. THIS: HND yah (86), DKH (y)e (86), PNJ e (86), PTH eh (86), HNK e (86), GJR yo (86), DGR eh (86), LHD e (86), SND hī (86), RAJ yau (86), GUJ ā (265), MAR hā (265), BNG e (86), ASS i (86), NEP yo (86), SNG mē(ka) (86), MAL mi (86), KOT jō (86), HIM yeh (86), KUL ye (86), MND e (86), ORY ehā (86), AWD yai (86), KUM yo (86), ROM kado (602), KNK he(ṅ) (86), DUM tahei (86), BRJ ī (86), GRH yū (86), PRY ya (86), MAI ī (86), KCH hī (86), MEW i (86), WGD to (665), BNJ i (86)

86a. THIS: ROM -le- (86)

87. THOU: HND tū (87), DKH tū (87), PNJ tū (87), PTH tū (87), HNK tū (87), GJR tū (87), DGR tū (87), LHD tū (87), SND tū (87), RAJ tū (87), GUJ tū (87), MAR tū (87), BNG tumi (87), ASS tōi (87), 87 NEP tā (87), SNG tō (87), MAL thiya (87), KOT tū (87), HIM tū (87), KUL tū (87), MND tū (87), ORY tume (87), AWD tū (87), KUM tē (87), ROM tu (87), KNK tu (87), DUM tu (87), BRJ tu (87), GRH tū (87), PRY tu (87), MAI tū (87), KCH tū (87), MEW tu (87), WGD tu (87), BNJ tū (87)

88. TONGUE: HND jībh (88), DKH jīb(h) (88), PNJ jībh (88), PTH jīv (88), HNK jībh (88), GJR jībh (88), DGR jibbh (88), LHD jibbh (88), SND j'ibh (88), RAJ jībh (88), GUJ jībh (88), MAR jībh (88), BNG jib (88), ASS zibha (88), NEP jibro (88), SNG diva (88), MAL dō (88), KOT dzibbh (88), HIM jīb (88), KUL jībh (88), MND jībh (88), ORY jibha (88), AWD jībhi (88), KUM jiboṛo (88), ROM čhib (88), KNK jīb(h) (88), DUM jibā (88), BRJ jībh (88), GRH jīb (88), PRY jip (88), MAI jībh (88), KCH jibbh (88), MEW jībh (88), WGD jab (88), BNJ jībh (88)

89. TOOTH: HND dāt (89), DKH dāt (89), PNJ dand (89), PTH dand (89), HNK dand (89), GJR dand (89), DGR dand (89), LHD dand (89), SND d'andu (89), RAJ dāt (89), GUJ dāt (89), MAR dāt (89), BNG dāt (89), ASS dāt (89), NEP dāt (89), SNG data (89), MAL daiy (89), KOT dānd (89), HIM dānd (89), KUL dānd (89), MND dānd (89), ORY dānta (89), AWD dāt (89), KUM dāt (89), ROM dand (89), KNK dāt(u) (89), DUM don (89), BRJ dāto (89), GRH dāt (89), PRY dan (89), MAI dāt (89), KCH dandh (89), MEW dāt (89), WGD dāt (89), BNJ dānt (89)

90.TREE: HND peṛ (90), DKH rūk (131), PNJ rukkh (131), PTH būṭā (373), HNK būṭā (373), GJR būṭo (373), DGR būṭā (373), LHD vaṇ (169), SND vaṇu (169), RAJ rūkh (131), GUJ jhāḍ (266), MAR jhāḍ (266), BNG gach (318), ASS gōs (318), NEP rukh (131), SNG gaha (318), MAL gahā (318), KOT pēṛ (90), HIM ḍāl (373), KUL buṭṭa (373), MND ḍāl (373), ORY gaccha (318), AWD rūkh (131), KUM rukh (131), ROM kaṣṭ (603), KNK rūku (131), DUM tom (-32), BRJ peṛ (90), GRH ḍāl (373), PRY darayṭ (-20), MAI gāch (318), KCH vaṇ (169), MEW rukh (131), WGD peḍ (90), BNJ jhāḍ (266)

90a. TREE: DGR rukkh (131), NEP boṭ (373), SNG ruka (131), KOT buṭṭi (373), KUM boṭ (373), KNK jhāḍ (266), KCH jhār (266)

91. TWO: HND do (91), DKH do (91), PNJ do (91), PTH do (91), HNK do (91), GJR do (91), DGR do (91), LHD ḍū (91), SND b'a (91), RAJ be (91), GUJ be (91), MAR don (91), BNG dui (91), ASS dui (91), NEP duī (91), SNG deka (91), MAL dey (91), KOT dui (91), HIM do (91), KUL duy (91), MND do (91), ORY dui (91), AWD dui (91), KUM dwi (91), ROM duy (91), KNK doni (91), DUM dui (91), BRJ dui (91), GRH dvī (91), PRY do (91), MAI dui (91), KCH b_a (91), MEW do (91), WGD be (91), BNJ dī (91)

92. WALK (GO): HND jā- (92), DKH jā(v)- (92), PNJ jā- (92), PTH gas- (93), HNK jul- (484), GJR ga- (93), DGR jā- (92), LHD vanj- (204), SND vañ- (204), RAJ jā- (92), GUJ jā- (92), MAR jā- (92), BNG ja- (92), ASS za- (92), NEP jā- (92), SNG ya- (92), MAL dā- (92), KOT ṛeu- (459), HIM jā- (92), KUL jā- (92), MND jā- (92), ORY jī- (92), AWD jā- (92), KUM jā- (92), ROM ža- (92), KNK vas- (204), DUM jā- (92), BRJ jā- (92), GRH jā- (92), PRY ja- (92), MAI jā- (92), KCH vin- (204), MEW jā- (92), WGD jā- (92), BNJ jā- (92)

92a. WALK (GO): HND ga- (93), DKH ga- (93), PNJ ge- (93), HNK ga- (93), GJR jā- (92), DGR ge- (93), RAJ ga- (93), MAR gel- (93), BNG gel- (93), ASS gal- (93), NEP ga- (93), SNG gi- (93), MAL g- (93), HIM ga- (93), KUL gə- (93), MND ge- (93), ORY gal- (93), KUM ga- (93), ROM gel- (93), DUM gi- (93), BRJ ga- (93), GRH ga- (93), PRY gi- (93), MAI ge- (93), MEW ga- (93), WGD ga- (93), BNJ g- (93)

93. WARM: HND garm (-6), DKH tattā (170), PNJ kosā (133), PTH tattā (170), HNK tattā (170), GJR tato (170), DGR tattā (170), LHD tattā (170), SND koso (133), RAJ ūnau (133), GUJ ūnū (133), MAR ūn (133), BNG gəɾɔm (-10), ASS umi (133), NEP tāto (170), SNG uṇu (133), MAL hōnu (133), KOT nettō (170), HIM tattā (170), KUL tōtta (170), MND tātā (170), ORY ušuma (-12), AWD garam (-7), KUM tāto (170), ROM tato (170), KNK hūn (133), DUM tatā (170), BRJ chīṭṭaiyo (632), GRH tātu (170), PRY garm (-21), MAI tāt (170), KCH koso (133), MEW tātā (170), WGD ōno (133), BNJ garam (-3)

93a. WARM: HNK kosā (133), GJR koso (133), SND tato (170), RAJ tātau (170)

94. WATER: HND pānī (94), DKH pānī (94), PNJ pānī (94), PTH pānī (94), HNK pānī (94), GJR pānī (94), DGR pānī (94), LHD pānī (94), SND pānī (94), RAJ pānī (94), GUJ pānī (94), MAR pānī (94), BNG jol (319), ASS pani (94), NEP pānī (94), SNG pān (94), MAL fen (94), KOT pānī

(94), HIM pāṇī (94), KUL pāṇi (94), MND pāṇī (94), ORY pāṇi (94), AWD pāṇī (94), KUM pāṇī (94), ROM pay (94), KNK udak (-12), DUM pāni (94), BRJ pāṇī (94), GRH pāṇi (94), PRY paṇi (94), MAI pāni (94), KCH pāṇī (94), MEW pāṇī (94), WGD pāṇī (94), BNJ pāṇi (94)

94a. WATER: ASS zəl (319), KCH jar (319)

95. WE: HND ham (95), DKH hame (95), PNJ asī (95), PTH assā (95), HNK asī (95), GJR ham (95), DGR as (95), LHD assā (95), SND asī (95), RAJ mhe (95), GUJ ame (95), MAR āmhī (95), BNG amra (95), ASS ami (95), NEP hāmī (95), SNG api (95), MAL aharumen (95), KOT hamme (95), HIM ase (95), KUL asse (95), MND asē (95), ORY āme (95), AWD ham (95), KUM ham (95), ROM ame(n) (95), KNK āmmī (95), DUM ame (95), BRJ hum (95), GRH ham (95), PRY ham (95), MAI ham sab (95), KCH asī (95), MEW ham (95), WGD hamu (95), BNJ ham (95)

96. WHAT: HND kyā (96), DKH kyā (96), PNJ kī (96), PTH keh (96), HNK ke (96), GJR ke (96), DGR keh (96), LHD kyā (96), SND kahiṛo (96), RAJ kāi (96), GUJ šū (96), MAR kāy (96), BNG ki (96), ASS kih (96), NEP ke (96), SNG mokada (96), MAL kēkey (96), KOT kε (96), HIM kyā (96), KUL kī (96), MND kyā (96), ORY kana (96), AWD kā (96), KUM kī (96), ROM so (96), KNK kasane (96), DUM kisek (96), BRJ kae (96), GRH kyā (96), PRY ka (96), MAI kī (96), KCH kuro (96), MEW kā (96), WGD kae (96), BNJ kāi (96)

97. WHITE: HND ciṭṭā (97), DKH ujlā (98), PNJ ciṭṭā (97), PTH ciṭṭā (97), HNK ciṭṭā (97), GJR ciṭo (97), DGR ciṭṭā (97), LHD ciṭṭā (97), SND accho (205), RAJ dhoḷau (267), GUJ dholū (267), MAR pāḍhrā (290), BNG dhobol (267), ASS boga (134), NEP seto (375), SNG sudu (428), MAL hudu (428), KOT šuklo (452), HIM ciṭṭā (97), KUL šetta (375), MND ciṭṭā (97), ORY dhalā (267), AWD ujar (98), KUM seto (375), ROM parno (-18), KNK dhave (267), DUM šukulā (452), BRJ dhulo (267), GRH saphed (-7), PRY bago (134), MAI ujar (98), KCH acho (205), MEW dhawaro (267), WGD doḷo (267), BNJ dhoḷo (267)

97a. WHITE: HND ujlā (98), DKH pāḍū (290), PNJ baggā (134), HNK gorā (485), SND bago (134), ASS dhowol (267), SNG āli (375), KUM sukilo (452), MAI dhaval (267), KCH dhauro (267)

98. WHO: HND kaun (99), DKH kon (99), PNJ kauṇ (99), PTH kuṇ (99), HNK koṇ (99), GJR koṇ (99), DGR kun (99), LHD koṇ (99), SND keru (99), RAJ kuṇ (99), GUJ koṇ (99), MAR koṇ (99), BNG ke (99), ASS kon (99), NEP ko (99), SNG kavuda (99), MAL kāku (99), KOT kuṇ (99), HIM kuṇ (99), KUL kūṇ (99), MND kūṇ (99), ORY kie (99), AWD ko (99), KUM ko (99), ROM kon (99), KNK koṇ (99), DUM koṇo (99), BRJ kōn (99), GRH ko (99), PRY koṇ (99), MAI kon (99), KCH ker (99), MEW koṇ (99), WGD kun (99), BNJ kuṇ (99)

99. WOMAN: HND aurat (-7), DKH tiyā (135), PNJ tivī (135), PTH jaṇāṇī (-4), HNK trīmat (135), GJR janī (491), DGR janāṇī (-3), LHD tremit (135), SND nāri (206), RAJ lugāi (472), GUJ baiḍī (268), MAR bāyko (268), BNG meye (321), ASS tirota (135), NEP svāsṇimanche (244), SNG gāṇī (429), MAL anhenmēhā (244), KOT chēuṛi (453), HIM tīmī (135), KUL beṭṭaḍi (557), MND janāṇ (-5), ORY bāilī (268), AWD joy (471), KUM seṇmeis (244), ROM žuvli (471), KNK bāil- (268), DUM joi (471), BRJ lugāi (472), GRH janāṇī (-8), PRY ran (136), MAI tia (135), KCH bāyṛī (-8), MEW berbānī (268), WGD buirī (268), BNJ bāi (268)

99a. WOMAN: DKH joī (471), PNJ rann (136), ORY māi (321), AWD tiy (135), ROM manušni (244), KNK manišī (244)

100. YELLOW: HND pīlā (100), DKH pīlā (100), PNJ pīlā (100), PTH pīlā (100), HNK pīlā (100), GJR pīlo (100), DGR pīlā (100), LHD pīlā (100), SND pīlo (100), RAJ pīlau (100), GUJ pīlū (100),

MAR pīvḷā (100), BNG hōlōde (322), ASS halōdhiya (322), NEP pahēlo (100), SNG kaha (431), MAL rēndōkula (546), KOT piūḷo (100), HIM pīlā (100), KUL piwḷa (100), MND piulā (100), ORY haḷadiā (322), AWD piyar (100), KUM pīlo (100), ROM galbeno (–19), KNK haḷduve (322), DUM haližžā (–33), BRJ pīlo (100), GRH pingḷu (659), PRY zard (–22), MAI pīyar (100), KCH pīro (100), MEW piyaro (100), WGD halojjo (322), BNJ piḷo (100)

Abbreviations for languages and dialects

ASS – Assamese	HNK – Hindko	NEP – Nepali
AWD – Awadhi	KCH – Kutchi	NIA – New Indo-Aryan
BNG – Bengali	KNK – Konkani	OIA – Old Indo-Aryan
BNJ – Banjari	KOT – Kotgarhi	ORY – Oriya
BRJ – Braj	KUL – Kului	PNJ – Punjabi
DGR – Dogri	KUM – Kumauni	PTH – Pothohari
DKH – Dakhini	LHD – Lahnda (Multani)	PRY – Parya
DUM – Dumaki (Domaaki)	MAI – Maithili	RAJ – Rajasthani (Marwari)
GJR – Gojri	MAL – Maldivian (Dhivehi)	ROM – Romany
GRH – Garhwali	MAR – Marathi	SND – Sindhi
GUJ – Gujarati	MEW – Mewati	SNG – Sin(g)halese
HIM – Himachali	MIA – Middle Indo-Aryan	WGD – Wagdi
HND – Hindi-Urdu	MND – Mandeali	

References

- Abdulla, F., M. O’Shea. 2005. *English Dhivehi, Dhivehi-English dictionary*. Version 1.0. Belconnen.
- Awan, Anjum Rafiq 2000. *Concise Gojri-English dictionary*. Rajouri: Anjuman Taraqi Gojri Adab.
- Berger, H. 1959. Die Burušaski-Lehnwörter in der Zigeunersprache. *Indo-Iranian Journal* 3: 17–43.
- Bloch, J. 1920. *La formation de la langue marathe*. Paris: É. Champion.
- Bloch, J. 1965. *Indo-Aryan: from the Vedas to modern times*. Paris: Libr. d’Amérique et d’Orient Adrien-Maisonneuve.
- Boretzky, Norbert. 1994. *Romani. Grammatik des Kalderaš-Dialekts mit Texten und Glossar*. Wiesbaden: Harrassowitz Verlag.
- Buddruss, G. 1984. Domaaki-Nachträge zum Atlas der Dardsprachen. *Münchener Studien zur Sprachwissenschaft* 43: 9–24.
- Chatterji, S. K. 1926. *The origin and development of the Bengali language*. Part 1: *Introduction. Phonology*. Calcutta: Calcutta University Press.
- Burrow T., M. B. Emeneau. 1961. *A Dravidian etymological dictionary*. Oxford: Clarendon Press.
- Gosvāmī, Om. 2000. *Ḍogri-hindī śabdkoś*. Jammū: J&K Academy of Art, Culture and Languages.
- Grierson, G. A. 1916. *Linguistic survey of India*. Vol. 9. Part 4: *Specimens of the Pahari languages and Gujuri*. Calcutta: Superintendent Government Printing.
- Grierson, G. A. 1927. *Linguistic survey of India*. Vol. 1. Part 1. *Introductory*. Calcutta: Government of India, Central Publication Branch.
- Hendriksen, H. 1976. *Himachali studies*. Vol 1. *Vocabulary*. København.
- Hoernle, A. F. Rudolf. 1880. *A comparative grammar of the Gaudian languages, with special reference to the Eastern Hindi, accompanied by a language-map and a table of alphabets*. London: Trübner & Co.
- Kogan, A. I. 2005. *Dardskie yazyki. Geneticheskaya kharakteristika*. Moskva: Vostochnaya literatura.
- Lorimer, D. L. R. 1939. *The Dumaki language: outlines of the speech of the Doma, or Bericho, of Hunza*. Nijmegen: Dekker & Van De Vegt.
- Mahapatra, B.P., B. Padmanabha, M.R. Ranganatha. 1980. *Census of India 1971. Ser. 1. Language monograph. Monogr. No 7. Survey of Mandeali and Kului in Himachal Pradesh*. New Delhi.

- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge University Press.
- Mukherji, Kakali et al. 2011. *Linguistic survey of India. Rajasthan*. Part 1. Language Division, Office of the Registrar General & Census Commissioner, India. s.l.
- Neog, Maheswar, Upedranath Goswami. 1987. *Chandrakanta abhidhan: a comprehensive dictionary of the Assamese language with etymology and illustrations of words with their meanings in Assamese and English*. Guwahati: Gauhati University.
- Oranskiy, I. M. 1977. *Fol'klor i yazyk gissarskikh par'ya (Srednyaya Aziya). Vvedenie, teksty, slovar'*. Moskva: Nauka, Glavnaya redakciya vostochnoy literatury.
- Praharaj, G. C. 1931–1940. *Purnnacandra odia bhashakosha*. Cuttack: Utkal Sahitya Press.
- Ramesh, G. V. 2010. *Banjara multilingual dictionary with semantic category section and indexes for Telugu, English and Hindi*. Hyderabad: Rajiv Vidya Mission (Education Department of Andhra Pradesh State Government).
- Rohra, S. K. 1965. *Descriptive analysis of Kachi (Kacchi)*. Thesis submitted to the University of Poona for the degree of Doctor of Philosophy in Linguistics. Poona: Deccan College, Post-graduate and Research Institute.
- Saleem, Teyyeb, Wazir Shah. 2005. *Conversational Saraiki, Urdu, English*. Karachi: Institute of Applied Linguistics.
- Samīr, Rāmājñā Dvivedī. 1955. *Avadhī koś*. Ilāhābād: Hindustānī Ekeḍemī.
- Schmidt, Ruth Laila. 1994. *A practical dictionary of modern Nepali*. Ratna Sagar.
- Sibghatulla, Mohd., Syed Abu Turaab Khataai Zamin. 2000. *Dakhni lughaat (Classical Urdu dictionary). With Dakhni idioms and phrases*. Bangalore: Malik Publications.
- Sukūn, Sultān. 2002. *Hindko Urdū luyat*. Pišāvar: Gandhārā Hindko Borḍ Pākistān.
- Suthar, B. L., S. S. Gahlot. 1995. *Rajasthani-Hindi-English dictionary*. Jodhpur: Rajasthani Sahitya Sansthan.
- Thakur, Gajendra, Nagendra Kumar Jha, Panjīkar Vidyanand Jha. 2012. *Videha English-Maithili dictionary*. Delhi: Shruti Publication.
- Thali, Mukesa. 1999–2001. *Rajhauns Konkani-English illustrated dictionary*. Panaji: Rajahasa.
- Turner, R. L. 1921a. Gujarati phonology. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 3: 329–365.
- Turner, R. L. 1921b. Gujarati phonology. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 4: 505–544.
- Turner, R. L. 1924. The Sindhi recursives or voiced stops preceded by glottal closure. *Bulletin of the School of Oriental Studies, University of London* 3(2): 301–315.
- Turner, R. L. 1926. The position of Romany in Indo-Aryan. *Journal of the Gypsy Lore Society*, 3rd Ser. 5(4): 145–188.
- Turner, R. L. 1931. *A comparative and etymological dictionary of the Nepali language*. London: K. Paul, Trench, Trübner.
- Turner, R. L. 1966. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

А. И. Коган. Генеалогическая классификация индо-арийских языков и лексикостатистика.

Генетические отношения индоарийских языков до сих пор остаются неясными. Существующие классификации зачастую носят интуитивный характер и не основываются на надежных сравнительно-исторических критериях. В статье делается попытка классификации новоиндийских языков по обновленным лексикостатистическим данным. Сопоставление полученного в итоге родословного древа с традиционными классификационными схемами позволило автору сделать ряд выводов относительно наиболее вероятной таксономии языков индоарийской группы.

Ключевые слова: индоарийские языки, генеалогическая классификация языков, лексикостатистика, глоттохронология.

К вопросу о точности глоттохронологии: датирование процесса лексических замен по данным романских языков

Статья представляет собой первую часть исследования, посвященного проблеме достоверности лингвистических датировок, получаемых с помощью метода глоттохронологии. В предлагаемой работе рассматривается процесс лексических замен, происходящих в базисной лексике одного языка с течением времени. В качестве исходных данных нами использовались 110-словные списки, собранные на материале 54 современных и нескольких исторических романских идиомов. При этом для измерения скорости замен списки современных языков сравнивались со списками классической и архаический латыни, а также старофранцузского и староитальянского. Временная дистанция между сопоставляемыми идиомами определялась с помощью трех различных глоттохронологических методов: классического уравнения М. Сводеша, модифицированной формулы С. А. Старостина, а также недавно предложенной потоковой модели. Сравнение полученных результатов позволило сделать ряд важных выводов о характере лексических изменений, адекватности существующих глоттохронологических моделей, а также численно оценить точность и надежность глоттохронологических расчетов при датировании общего процесса замен. Вторую часть исследования планируется посвятить проблеме датирования относительной дивергенции двух родственных языков.

Ключевые слова: глоттохронология, лексикостатистика, список Сводеша, романские языки.

Одной из основных аксиом лексикостатистики является равномерность процесса замен базисной лексики, описанная М. В. Араповым и М. М. Херц следующим образом:

Доля p слов из O [сновного] S [писка], которые сохранятся (не будут заменены другими словами) на протяжении интервала времени Δt (равного, например, году, столетию или тысячелетию) постоянна (т. е. зависит только от величины выбранного промежутка, но не от того, как он выбран, или слова какого языка рассматриваются). (Арапов, Херц 1974: 22)

Математическим соответствием этой аксиомы является коэффициент сохраняемости, используемый в глоттохронологических формулах вычисления времени. Впервые коэффициент сохраняемости был вычислен М. Сводешем путем сравнения двухсотсловных списков ряда языков с долгой письменной историей в ранней форме их существования и на более позднем этапе развития: табл. 1.

Как можно заметить, процент сохранившейся лексики при данной методике подсчета колеблется в интервале 76—85 % слов за тысячелетие (Сводеш 1960а: 47).

В дальнейшем, используя модифицированный список базисной лексики, состоящий из 100 слов, Сводеш пересчитывает на его основании коэффициент сохраняемости, предварительно исключив из рассмотрения коптский (так как он может быть непрямой

Таблица 1. Значения коэффициента сохраняемости, рассчитанные Сводешем на основании двухсотсловных списков (Сводеш 1960а: 34).

Ранняя форма	Поздняя форма	Интервал времени	Процент сохранившихся слов на 1000 лет
Среднеегипетский	Коптский	2300 лет	76
Классическая латынь	Румынский	2000 лет	77
Древневерхненемецкий	Немецкий	1100 лет	78
Классический китайский	Северокитайский	1000 лет	79
Латынь Плавта	Французский Мольера	1850 лет	79
Доминика кариб 1650 н.э.	Современная форма	300 лет	80
Классическая латынь	Португальский	2000 лет	82
Койне	Кипрский диалект	[без даты]	83
Койне	Афинский диалект	[без даты]	84
Классическая латынь	Итальянский	2000 лет	85
Древнеанглийский	Английский	1000 лет	85
Латынь Плавта	Ранний новоиспанский 1600 г.	1800 лет	85

Таблица 2. Значения коэффициента сохраняемости, рассчитанные Сводешем на основании стословных списков (Сводеш 1960б: 72).

Язык	Интервал времени	Процент сохранившихся слов на 1000 лет
Шведский	1020	94,3
Немецкий	1100	89,0
Английский	1000	76,6
Румынский	2150	76,4
Французский	1850	77,6
Афинский	2070	83,6
Китайский	1000	79,6

потомком среднеегипетского), испанский, итальянский, португальский, каталанский (поскольку уже есть подсчет по родственному ему французскому), кипрский (поскольку есть подсчет по афинскому): табл. 2.

Однако уже в 1958 г. появилась статья Дж. Ри, в которой сравниваются стословные списки восьми романских языков и при использовании коэффициента сохраняемости $r = 0,85$ (максимальный среди вычисленных Сводешем) были получены нелепые даты расхождения, например, испанского и португальского — 370 лет назад, румынского и итальянского — 826 лет назад. Поскольку история романских языков довольно хорошо известна, очевидно, что эти результаты не отвечают действительности. Исходя из этого, Дж. Ри делает вывод о некорректности всей методики глоттохронологии. В ответной статье А. Крубер возражает, что в данном случае нужно не сходу отказываться от глоттохронологии вообще, а попытаться ее доработать. В частности, для романских языков

Таблица 3. Расчеты К. Бергсланда и Х. Фогта (Bergsland, Vogt 1962: 117–125).

Язык-1	Язык-2	Временное расстояние (100-словник)	Временное расстояние (200-словник)	Временное расстояние (215-словник)	Фактическая временная дистанция
Древне-скандинавский	Исландский	63	130	194	1000
Древне-скандинавский	диалект Гьестал	345	799	901	1000
Древне-скандинавский	диалект Санднес	412	861	964	1000
Древне-скандинавский	Риксмол	637	930	1000	1000
Древне-грузинский	Грузинский	338	750	861	1500
Грузинский	Мегрельский	1316	1004	1033	1800–1900
Древне-армянский	Армянский	211	437	437	1500

Таблица 4. Значения коэффициента сохраняемости, полученные Старостиным на материале стословных списков (Starostin 2000: 230).

Язык	Интервал времени	λ_1 (с учетом заимствований)	λ_2 (без учета заимствований)
Японский	1200 лет	0,11	0,06
Китайский	2600 лет	0,1	0,1
Английский	1300 лет	0,14	0,1
Немецкий	1200 лет	0,08	0,05
Французский	1500 лет	0,09	0,07
Испанский	1500 лет	0,07	0,06
Румынский	1500 лет	0,09	0,06

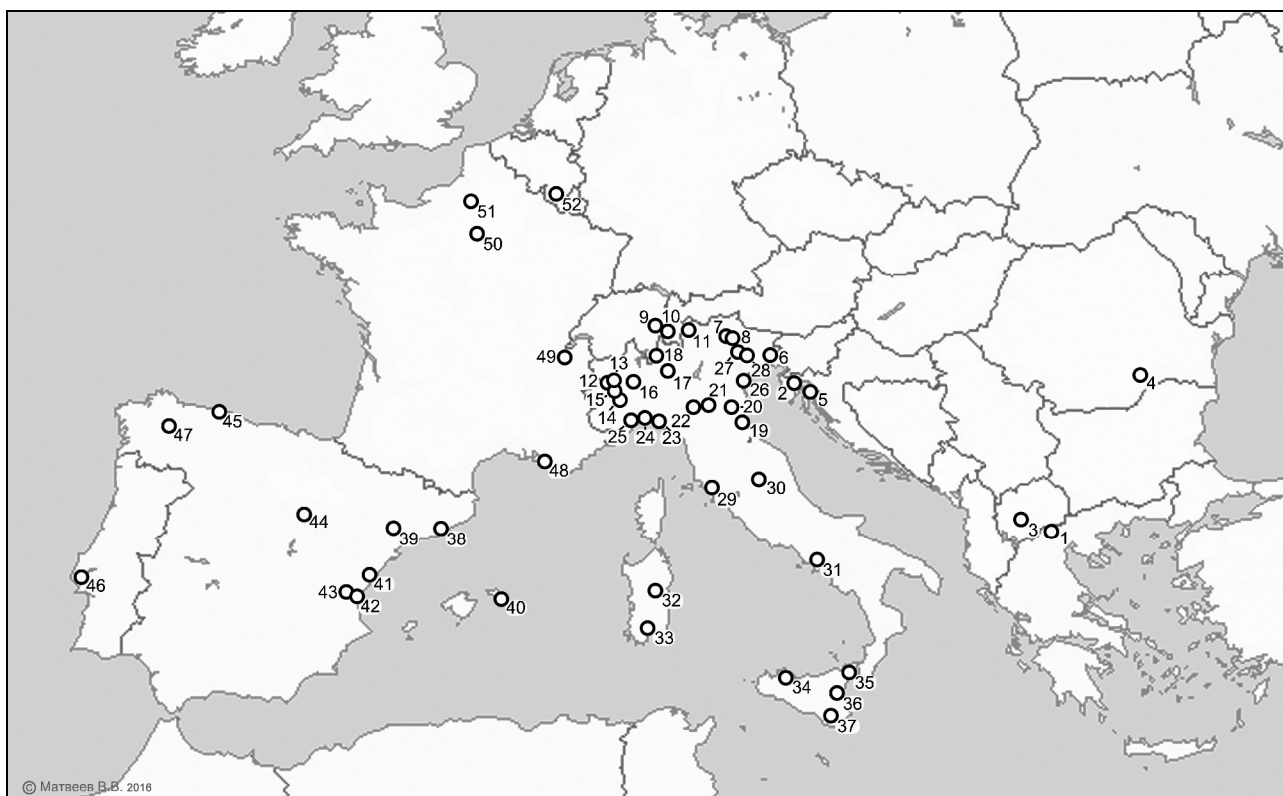
при введении поправочного коэффициента 2,2 Крубнером были получены вполне приемлемые даты (Kroeber 1958).

Тем не менее, последующие подсчеты, произведенные по методике Сводеша К. Бергсландом и Х. Фогтом (Bergsland, Vogt 1962) на материале риксмолы, исландского, армянского и грузинского языков, снова дали результат, существенно расходящийся с известными значениями: табл. 3.

Как видно из табл. 3, кроме единственного исключения, расчетные значения временной дистанции ни в одном случае не совпадают с фактическими. Особенно разительным оказалось расхождение при сравнении исландского и древнескандинавского языков.

В середине 80-х годов прошлого века усовершенствованием метода глоттохронологии занялся С. А. Старостин, который улучшил формулу дивергенции языков и выдвинул требование исключать заимствования из списков сопоставляемой лексики, поскольку заимствования являются следствием внешних контактов, а не внутреннего изменения языка: табл. 4.

Рисунок 1. Географическое распространение идиомов, используемых в исследовании. Номерами на карте обозначены: 1. мегленорумынский; 2. истрорумынский; 3. арумынский; 4. румынский (литературный); 5. далматинский; 6. фриульский (центральный); 7. ладинский (гарденский); 8. ладинский (фассанский); 9. руманшский (сурсельский); 10. руманшский (сурмиранский); 11. руманшский (нижнеэнгадинский); 12. пьемонтский (Ланцо-Торинезе); 13. пьемонтский (Барбания); 14. пьемонтский (Карманьола); 15. пьемонтский (Турин); 16. пьемонтский (верчельский); 17. ломбардский (Бергамо); 18. ломбардский (Плезियो); 19. эмилиано-романьольский (Равенна); 20. эмилиано-романьольский (Феррера); 21. эмилиано-романьольский (Карпи); 22. эмилиано-романьольский (Реджо); 23. лигурийский (Рапалло); 24. лигурийский (Генуя); 25. лигурийский (Стелла); 26. венецкий (Венеция); 27. венецкий (Примьеро); 28. венецкий (Беллуно); 29. тосканский (Гроссето); 30. умбрийский (Фолиньо); 31. неаполитанский; 32. логудорский; 33. кампиданский; 34. сицилийский (Палермо); 35. сицилийский (Мессина); 36. сицилийский (Катания); 37. сицилийский (юго-восточный; объединены данные, полученные от информантов из Рагузы и Агридженто); 38. каталанский (центральный); 39. каталанский (северо-западный); 40. каталанский (Менорка); 41. каталанский (Кастельон-де-ла-Плана); 42. каталанский (Валенсия); 43. каталанский (Манисес); 44. кастильский (Сория); 45. астурийский (центральный); 46. португальский (литературный); 47. галисийский (центральный); 48. окситанский (провансальский); 49. франко-провансальский (савойский); 50. французский (литературный); 51. пикардский (южный); 52. валлонский (южный).



За последние годы в рамках проекта «Глобальная лексикостатистическая база данных» было накоплено большое количество списков базисной лексики, качество которых стоит на более высоком уровне, чем у материала, доступного Сводешу и Старостину.

В частности, в 2015—2016 гг. одним из авторов данной статьи (М. Н. Саенко) были собраны аннотированные 110-словные списки базисной лексики для 54¹ романских идиомов (а также 4 списка для староитальянского Данте, старофранцузского Кретьена де Труа, ла-

¹ В дальнейшем будут использоваться 52 идиома, так как списки для руманч грижун и итальянского литературного были исключены в силу искусственного характера первого и чрезвычайной архаичности второго. Полные списки с источниками, таблицами транслитерации и описанием доступны на сайте «Глобальная лексикостатистическая база данных» <http://starling.rinet.ru/cgi-bin/main.cgi?root=new100>.

тыни Плавта и Апулея). При этом предпочтение отдавалось диалектам и «малым» языкам без строгой литературной нормы, из 54 идиомов лишь 5 являются строго нормированными (румынский, руманч грижун, итальянский, португальский, французский). Географически исследование было ограничено только Старым Светом (см. рис. 1). Данные были получены как из диалектных словарей романских языков, так и от информантов. В связи с необходимостью за короткий срок охватить большое количество материала, работа с носителями велась через Интернет (рассылался опросник, а после его заполнения информантам задавались уточняющие вопросы). Всего было опрошено 76 информантов.

Используя материал романских языков в качестве тестового, мы последовательно применяем к нему два наиболее известных глоттохронологических метода: классическую глоттохронологию М. Сводеша, усовершенствованную методику С. А. Старостина, а также недавно предложенную потоковую модель. При этом определим основные цели и задачи нашего исследования следующим образом:

- 1) Проверить применимость существующих глоттохронологических моделей (М. Сводеша, С. А. Старостина, потоковой модели) для датирования изменений в базисной лексике романских языков;
- 2) Определить оптимальные параметры моделей, обеспечивающие наилучшее соответствие расчетных значений и исходных лексикостатистических данных. При необходимости произвести калибровку моделей с учетом новых параметров и сравнить результаты, полученные с использованием калиброванных и некалиброванных моделей;
- 3) Установить объективные и теоретические пределы точности при вычислении лингвистических датировок с применением рассматриваемых моделей.

В первой части данной статьи мы рассмотрим процесс замен в лексике одного языка по мере его развития и подробно остановимся на методике вычисления временной дистанции между языком-предком и языком-потомком. Вторая часть работы будет посвящена относительной дивергенции двух родственных языков и проблемам ее датирования.

1. Глоттохронологические модели общего процесса лексических замен

Исходные данные для определения скорости изменений в базисной лексике романских языков были получены путем сравнения современных идиомов с ближайшими родственниками их непосредственных предков: архаической латынью Плавта, поздней классической латынью Апулея, староитальянским Данте и старофранцузским Кретьена де Труа, датировки которых можно установить по историческим источникам. По результатам сравнения составлена табл. 5, где для каждой пары (или нескольких пар)² языков указан процент совпадений³ между соответствующими основными списками, а также временной интервал между датами их фиксации.

² При сопоставлении нескольких пар языков (строки 3, 4, 10, 12), в соответствующих столбцах приводится минимальное, максимальное и среднее арифметическое значение. Последнее рассчитывается как сумма всех известных процентов совпадений, деленная на количество слагаемых, что в общем случае не совпадает со средним значением между максимальным и минимальным процентом совпадений. Например, в строке 4 для трех пар языков с долями совпадений 86%, 83% и 87% *среднее* между минимальным и максимальным значением составит $(83+87)/2=85$, а *среднее арифметическое* — $(83+86+87)/3=85,3$.

³ Значения процентов совпадений приводятся по данным табл. 8, полученной путем попарного сравнения всех собранных списков в программе Starling (см. Дополнительные материалы).

Таблица 5. Лексикостатистические данные о скорости замен в базисной лексике романских языков.

№	Сравниваемые языки	Мин. % совп.	Средн. % совп.	Макс. % совп.	Интервал времени, лет
1	Исходное значение (для любого идиома)	—	100	—	0
2	Архаическая латынь (Плавт, 250 г. до н.э.) — поздняя классическая латынь (Апулей, 150 г. н.э.)	—	98	—	400
3	Староитальянский (Данте, 1270 г.) — современные итальянские (тосканский, умбрийский)	85	86,5	88	730
4	Старофранцузский (Кретьен де Труа, 1140 г.) — современные французские (литературный французский, пикардский, валлонский)	83	85,3	87	860
5	Поздняя классическая латынь (Апулей, 150 г.) — старофранцузский (1140 г.)	—	84	—	990
6	Поздняя классическая латынь (150 г.) — староитальянский (1270 г.)	—	85	—	1120
7	Архаическая латынь (250 г. до н.э.) — старофранцузский (1140 г.)	—	78	—	1390
8	Архаическая латынь (250 г. до н.э.) — староитальянский (1270 г.)	—	78	—	1520
9	Поздняя классическая латынь (150 г.) — далматинский (1900 г.)	—	81	—	1750
10	Поздняя классическая латынь (150 г.) — современные романские (52 идиома, 2000 г.)	69	74,6	83	1850
11	Архаическая латынь (250 г. до н.э.) — далматинский (1900 г.)	—	76	—	2150
12	Архаическая латынь (250 г. до н.э.) — современные романские (52 идиома, 2000 г.)	64	69,5	77	2250

Рисунок 2. Изменение доли сохранившейся лексики романских языков в зависимости от времени. Нумерация рядом с точками указывает на соответствующие строки табл. 5. Для точек 3, 4, 10 и 12 показан диапазон разброса долей совпадений и среднеарифметическое значение.



Для наглядности полученные данные можно представить в виде диаграммы, которая отражает уменьшение доли сохранившейся лексики $N(t)$ с течением времени (t): рис. 2.

Приведенная диаграмма свидетельствует о ярко выраженном статистическом характере исходных данных, который проявляется в значительном разбросе процентов совпадений, полученных для точек с одинаковыми или хронологически близкими датировками. В то же время, очевидно, что все (даже наиболее выделяющиеся) значения хорошо сгруппированы вокруг некоторой средней величины на всем рассматриваемом интервале времени, что позволяет говорить о существовании зависимости случайного процесса лексических замен от времени. Для определения характера этой зависимости и ее параметров перейдем к рассмотрению конкретных глоттохронологических моделей: классическому уравнению М. Сводеша, усовершенствованной формуле С. А. Старостина и потоковой модели.

1.1. Глоттохронологическая модель М. Сводеша.

Классический метод глоттохронологии, предложенный М. Сводешем в середине XX в., построен по аналогии с методом радиоуглеродного датирования и базируется на четырех основных допущениях (постулатах⁴):

- а) наличие в словаре каждого языка некоторого устойчивого подмножества слов — базисной лексики, из которой можно выделить универсальный список значений, обладающий повышенной стабильностью в любом языке⁵;
- б) постоянная скорость лексических изменений в основном списке, не зависящая от выбранного языка и временного периода;
- в) одинаковая стабильность всех элементов основного списка;
- г) независимость замен в списках языков-потомков после их разделения.

В качестве математического аппарата, отражающего содержание постулатов глоттохронологии, была использована формула радиоактивного распада, описывающая процесс замен в базисной лексике в виде экспоненциальной зависимости с коэффициентом сохраняемости λ , определяющим темп замен:

$$N(t) = e^{-\lambda \cdot t}.$$

Как уже говорилось, согласно подсчетам Сводеша, которые проводились на разнообразном материале (в том числе романском), за 1000 лет различные языки в среднем сохраняют около 85% основного списка (Swadesh 1952: 456—460), что соответствует коэффициенту $\lambda=0,16$ ($e^{-0,16 \cdot 1}=0,852$). Подставив данное значение λ в исходное выражение, получаем формулу для датирования процесса замен в лексике одного языка:

$$N_{sw}(t) = e^{-0,16t}.$$

Для калибровки формулы Сводеша по новым данным, полученным на основе романских языков, воспользуемся методом наименьших квадратов. Смысл данного метода сводится к поиску таких параметров модели (в данном случае — коэффициента λ), при

⁴ Эти и другие положения глоттохронологии Сводеша более подробно излагаются в работе Арапов, Херц 1974: 21—25.

⁵ Наибольшее распространение получил 100-словный список, зачастую называемый также «списком Сводеша».

Рисунок 3. Соответствие исходной и калиброванной модели М. Сводеша исходным данным: $N_{Sw}(t) = e^{-0,16 \cdot t}$; $N_{SwC}(t) = e^{-0,16 \cdot t}$.

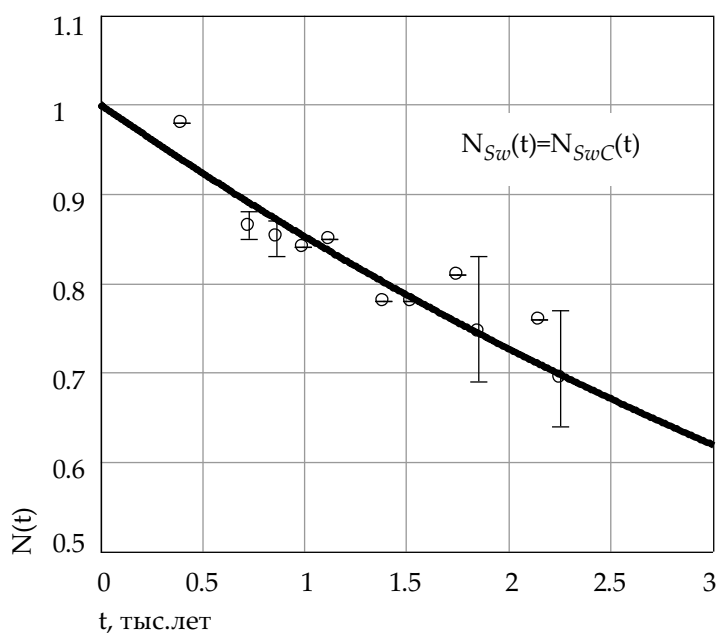
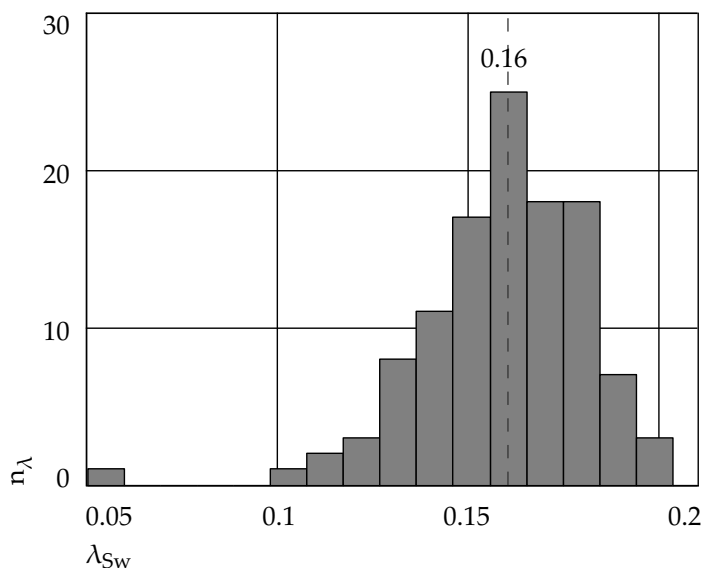


Рисунок 4. Количество пар языков (n_λ) с коэффициентом λ , полученным для калиброванной модели Сводеша по исходным данным (табл. 5).



которых суммарная разница (или отклонение) между фактическими и расчетными долями совпадений, вычисленными для каждой из опорных точек, является минимальной.

Например, в соответствии с табл. 5 (строка 2), процент совпадений между списками арахаической латыни Плавта и классической латыни Апулея составляет 98 ($N_\phi=0,98$) при разделяющем их временном интервале 400 лет ($t=0,4$). Подставляя это значение времени в формулу Сводеша, получим выражение: $N_p = e^{-\lambda \cdot 0,4}$. Теперь, для того чтобы обеспечить наилучшее совпадение фактического (N_ϕ) и расчетного (N_p) значений, необходимо найти такую величину λ , при которой квадрат разности между фактическим и расчет-

ным значениями $(N_\phi - N_p)^2$ окажется наименьшим. Таким образом, задача поиска минимального отклонения ε для данной опорной точки будет иметь вид:

$$\varepsilon = (e^{-\lambda \cdot 0,4} - 0,98)^2 \rightarrow \min.$$

Суммируя отклонения, полученные аналогичным способом для каждой опорной точки, получим общую формулу для расчета оптимального коэффициента λ :

$$\varepsilon = \sum_i (N_p - N_\phi)^2 \rightarrow \min,$$

где N_p — расчетное значение доли общей лексики, вычисленное по формуле $N_{sw}(t) = e^{-\lambda t}$, i — номер опорной точки, а N_ϕ и t — фактические значения доли совпадений и времени, представленные в табл. 5.

Вычисления, проведенные с помощью пакета Mathcad, показали, что минимальное отклонение между расчетными и фактическими значениями достигается при коэффициенте сохраняемости равном 0,16 (при этом величина суммарного отклонения составляет $\varepsilon=0,094$).

Таким образом, калиброванное значение λ совпало с константой Сводеша, а полученная модель оказалась идентична исходной (см. рис. 3):

$$N_{swC}(t) = N_{sw}(t) = e^{-0,16t}.$$

Как видно на приведенном графике (рис. 3), расчетные значения модели хорошо соответствуют опорным точкам на всем временном интервале (до 2,5 тыс. лет). При этом распределение значений λ , полученное для всех опорных точек по формуле Сводеша, оказалось близким к нормальному распределению⁶ с математическим ожиданием $\bar{\lambda}_{sw} = 0,16$ и средним квадратическим отклонением $\sigma_\lambda \approx 0,02$ (рис. 4).

Воспользуемся теперь другой глоттохронологической моделью, которая была предложена С. А. Старостиным.

1.2. Усовершенствованная глоттохронология С. А. Старостина.

Анализируя критику методики Сводеша, С. А. Старостин приходит к выводу о несостоятельности 2-го и 3-го постулатов глоттохронологии и указывает на необходимость их пересмотра. В частности, он приводит следующие аргументы (Starostin 2000: 229—230, 236—237):

- а) слова в базисной лексике языка со временем устаревают, и чем больше рассматриваемый промежуток времени, тем больше вероятность замены слова в основном списке, а следовательно — тем выше скорость распада.
- б) слова в основном списке неоднородны и обладают разной стабильностью, поэтому с течением времени общая скорость распада снижается из-за повторных замен наименее устойчивых значений и увеличивающейся доли более устойчивых.

Чтобы учесть эти особенности в математическом аппарате глоттохронологии, С. А. Старостин предлагает ввести в формулу Сводеша две поправки, одна из которых отражает замедление процесса замен, связанное с проявлением в списке наиболее стабильной

⁶ После исключения из рассмотрения выделяющегося значения $\lambda_1=0,051$, (строка 2 табл. 5), найденный при калибровке коэффициент $\lambda=0,16$ не изменился.

Рисунок 5. Сравнение исходной $N_{St}(t)$ и калиброванной $N_{StC}(t)$ моделей Старостина: $N_{St}(t) = e^{-0,05 \cdot N_{St} \cdot t^2}$; $N_{StC}(t) = e^{-0,11 \cdot N_{StC} \cdot t^2}$.

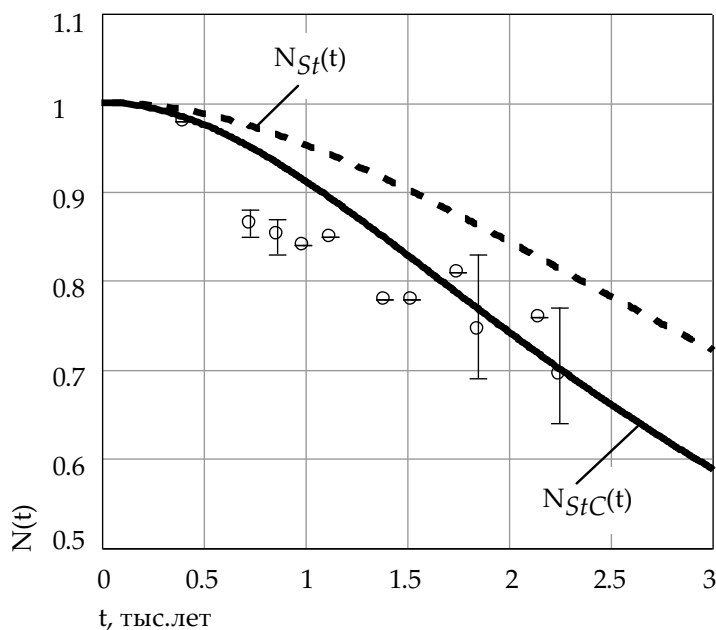
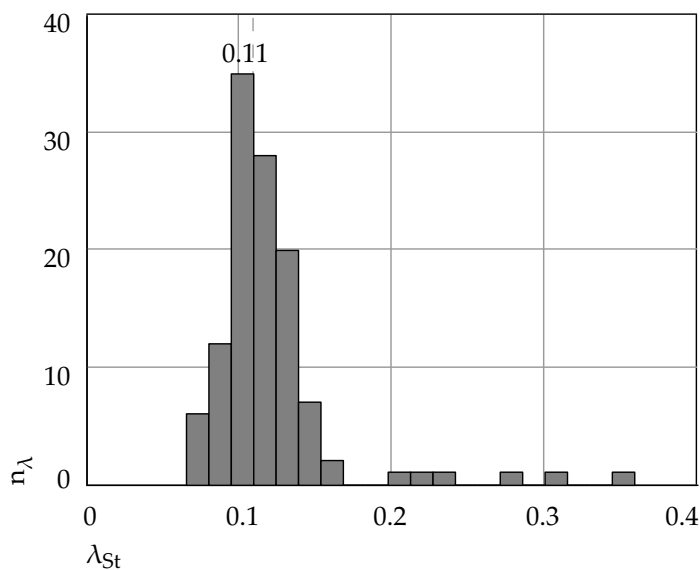


Рисунок 6. Количество пар языков (n_λ) с коэффициентом λ , полученным для калиброванной модели Старостина по исходным данным (табл. 5).



части лексики ($\lambda = \lambda \cdot N(t)$), а вторая — наоборот, его ускорение, обусловленное устареванием сохранившихся значений ($\lambda = \lambda \cdot t$). В результате новая глоттохронологическая модель принимает следующий вид:

$$N_{St}(t) = e^{-\lambda \cdot N_{St} \cdot t^2}.$$

При этом величина константы λ , согласно С. А. Старостину, должна составлять около 0,05 для большинства языков:

$$N_{St}(t) = e^{-0,05 \cdot N_{St} \cdot t^2}.$$

Для калибровки параметров рассмотренной модели по данным романских языков, как и в предыдущем случае, воспользуемся методом наименьших средних квадратов и получим следующие значения коэффициента λ и минимального суммарного отклонения ε :

$$\begin{aligned}\lambda &= 0,108; \\ \varepsilon &= 0,214.\end{aligned}$$

Очевидно, что найденный коэффициент $\lambda \approx 0,11$ более чем в два раза отличается от исходного (0,05), что заставляет нас перейти к уточненной модели вида:

$$N_{Sc}(t) = e^{-0,11 \cdot N_{Sc} \cdot t^2}.$$

Как показывает рассмотрение графиков (рис. 5), переход к калиброванной модели позволяет добиться гораздо лучшего совпадения расчетных значений с опорными точками в диапазоне от 1,5 до 2,5 тыс. лет (в отличие от исходной формулы, дающей заметно «заглубленные» датировки практически на любом временном интервале).

Распределение значений λ , полученное по исходным данным для модели Старостина выглядит менее равномерным, чем в случае с формулой Сводеша, однако также достаточно близко к нормальному: математическое ожидание составляет $\bar{\lambda}_{St} = 0,11$, среднее квадратическое отклонение $\sigma_\lambda \approx 0,02$ (рис. 6). При этом наличие шести сильно выделяющихся значений, соответствующих $\lambda > 0,2$ (точки 3, 4 и 5 на рис. 2), не повлияло на результат калибровки и конечный вид модели⁷.

Наряду с описанными выше традиционными глоттохронологическими моделями М. Сводеша и С. А. Старостина рассмотрим также статистический метод, основанный на потоковой интерпретации процесса лексических замен.

1.3. Потоковая глоттохронологическая модель.

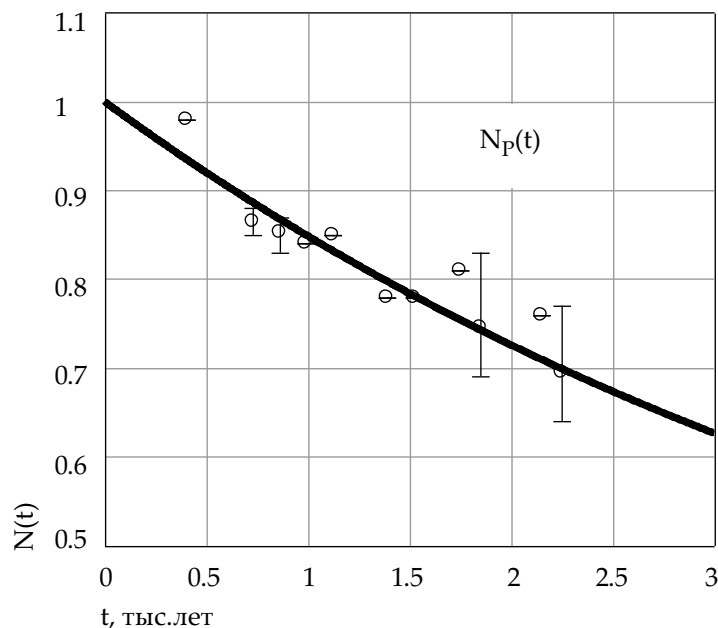
В основе потоковой модели лежит представление о том, что процесс замен каждого из значений основного списка является потоком случайных событий, которые происходят с малой интенсивностью и не влияют одно на другое. Таким образом, развитие всего основного списка можно описать как сумму нескольких независимых экспоненциальных потоков с различным, но постоянным темпом распада⁸. Каждая из таких составляющих соответствует группам значений или отдельным значениям в составе основного списка, которые обладают одинаковой или близкой устойчивостью. При этом число групп и их коэффициенты стабильности подбираются в зависимости от данных, используемых для калибровки модели. Например, принимая исходное количество составляющих модели равное трем, получаем выражение следующего вида:

$$N_p(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t},$$

где c_1 , c_2 и c_3 — доля слов в каждой из составляющих основного списка, а λ_1 , λ_2 и λ_3 — соответствующие коэффициенты стабильности каждой из них. При этом сумма всех най-

⁷ После исключения всех выделяющихся значений из расчетов величина λ_{St} по-прежнему составила 0,11.

⁸ Подробное теоретическое обоснование предлагаемой модели приводится в статье Васильев, Милитарев 2008: 518–523, а практические примеры ее применения в работах Васильев, Старостин 2013, Васильев, Коган 2014. Можно заметить, что данный подход, как и методика С. А. Старостина, подразумевает отказ от 3-го и 4-го постулатов Сводеша, однако использует при этом совершенно другие исходные посыпки.

Рисунок 7. Соответствие между полученной потоковой моделью и исходными данными: $N_p(t) = 0,2 + 0,8e^{-0,21t}$.

денных компонент $c_1+c_2+c_3$ должна равняться единице (т.е. образовывать полный основной список), а все коэффициенты — иметь неотрицательные значения. Так, например, классическую формулу Сводеша можно представить как частный случай потоковой модели, которая содержит всего одну значимую составляющую $c_1=1$ с коэффициентом $\lambda_1=0,16$.

Определим оптимальные параметры потоковой модели по исходным данным (табл. 5 и рис. 2) с помощью метода наименьших средних квадратов. Полученные значения приводятся ниже:

$$\begin{aligned} c_1 &= 0,200; c_2 = 0,238; c_3 = 0,562; \\ \lambda_1 &= 0,000; \lambda_2 = 0,210; \lambda_3 = 0,210; \\ \varepsilon &= 0,094. \end{aligned}$$

Нулевой коэффициент первой составляющей ($\lambda_1=0$), указывает на высокую устойчивость входящих в нее значений (около 20% основного списка), которые сохраняются в языке с течением времени. В то же время равенство коэффициентов $\lambda_2=\lambda_3=0,210$ свидетельствует о том, что заданное число слагаемых было избыточным, и модель без потери точности может быть сведена к двум содержательным компонентам: 20% ($c_1=0,2$) — сверхстабильная часть списка, 80% ($c_2+c_3=0,8$) — экспоненциально убывающая часть значений с коэффициентом $\lambda_{2,3}=0,21$:

$$N_p(t) = 0,2 + 0,8e^{-0,21t}.$$

Представленный график (рис. 7) позволяет убедиться в хорошем соответствии потоковой модели фактическим данным о процессе замен в романских языках, что численно подтверждается незначительной суммарной ошибкой ($\varepsilon=0,094$), полученной в ходе калибровки.

Завершив рассмотрение основных глоттохронологических методов и моделей, перейдем к сравнению и анализу полученных результатов.

Рисунок 8. Сравнение исходных и калиброванных моделей глоттохронологии, полученных по данным романских языков:

- $N_{Sw}(t) = e^{-0,16t}$ — исходная/калиброванная модель Сводеша,
- $N_{St}(t) = e^{-0,05 \cdot N_{St} \cdot t^2}$ — исходная модель Старостина,
- $N_{StC}(t) = e^{-0,11 \cdot N_{StC} \cdot t^2}$ — калиброванная модель Старостина,
- $N_p(t) = 0,2 + 0,8e^{-0,21t}$ — калиброванная потоковая модель.

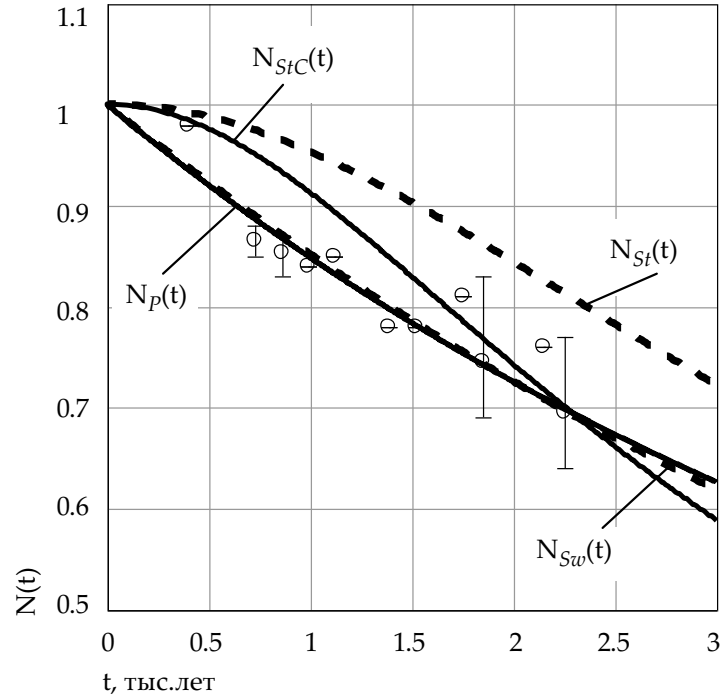


Таблица 6. Сравнение исходных и калиброванных моделей.

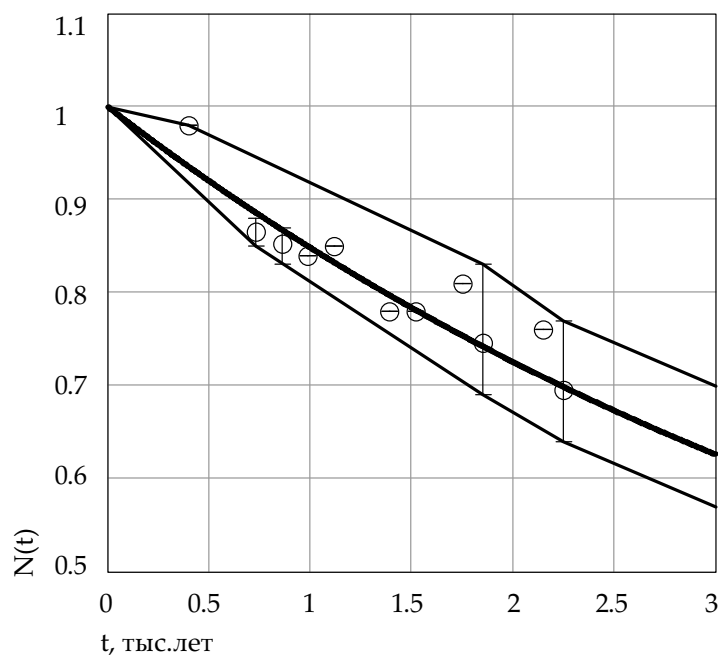
Общий вид и название модели	Вид исходной модели	Вид калиброванной модели
$N(t) = e^{-\lambda \cdot t}$ модель Сводеша	$N_{Sw}(t) = e^{-0,16t}$ $\lambda_{Sw} = 0,16$	$N_{SwC}(t) = e^{-0,16t}$ $\lambda_{SwC} = 0,16$ ($\epsilon_{Sw} = 0,094$)
$N_{St}(t) = e^{-\lambda \cdot N_{St} \cdot t^2}$ модель Старостина	$N_{St}(t) = e^{-0,05 \cdot N_{St} \cdot t^2}$ $\lambda_{St} = 0,05$	$N_{StC}(t) = e^{-0,11 \cdot N_{StC} \cdot t^2}$ $\lambda_{StC} = 0,11$ ($\epsilon_{St} = 0,214$)
$N_p(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t}$ потоковая модель	$N_p(t) = 0,2 + 0,8e^{-0,21t}$ $c_1 = 0,20; c_2 = 0,80;$ $\lambda_1 = 0,00; \lambda_2 = 0,21;$ ($\epsilon_p = 0,094$)	

1.4. Сравнение полученных моделей и их оценка.

Обратимся к сравнительной таблице (табл. 6), которая содержит как исходные, так и калиброванные модели, а также значения соответствующих параметров.

Сопоставление столбцов таблицы показывает, что существенные изменения в ходе калибровки моделей по опорным точкам произошли только в формуле Старостина: полученный коэффициент λ_{St} составил 0,11 при исходном значении 0,05. При этом калиброванная модель Сводеша оказалась неизменной с константой $\lambda_{Sw}=0,16$.

Рисунок 9. Фактический разброс значений $N(t)$, используемых в качестве исходных данных для калибровки моделей.



Если мы сравним графики представленных моделей (рис. 8), то обнаружим, что потоковая модель и модель Сводеша (несмотря на принципиальное отличие используемых подходов), дают практически идентичные датировки на всем временном диапазоне с одинаковой суммарной погрешностью $\varepsilon_{sw}=\varepsilon_p=0,094$.

При этом, как следует из рисунка, обе эти модели обеспечивают наилучшее совпадение с опорными точками, в то время как применение модели Старостина приводит к существенным неточностям как с исходным коэффициентом ($\lambda_{st}=0,05$), так и после его корректировки в соответствии с исходными данными ($\lambda_{stc}=0,11$). И в том, и в другом случае расчетные датировки в интервале 0,5...1,5 тыс. лет оказываются завышенными по отношению к фактическим. Главной причиной этих расхождений является неравномерный «замедленно-ускоренный» характер распада, вызванный введением ускоряющей и замедляющей поправок.

Таким образом, можно сделать вывод, что процесс изменений в базисной лексике одного языка наиболее корректно описывается с помощью экспоненциальной зависимости с постоянным (но не обязательно одинаковым!) темпом замен отдельных значений или частей списка.

Убедившись в принципиальной адекватности математического аппарата глоттохронологии для описания процесса словарных замен, мы можем перейти к численной оценке точности и надежности различных моделей при датировании лексической дивергенции.

1.5. Оценка точности глоттохронологических моделей.

Представляется очевидным, что точность расчетных датировок в первую очередь обусловлена точностью исходных данных, которые используются для идентификации и калибровки параметров моделей. Следовательно, измерение точности следует начинать с оценки фактического разброса значений, объективно присутствующего в исходных данных в силу их статистического характера. Графически данный разброс можно представить в виде кривых, соединяющих крайние значения опорных точек на всем интервале времени: рис. 9.

Рисунок 10 а, б. Иллюстрация разброса фактических долей совпадений и соответствующих датировок по отношению к расчетным значениям $N(t)$ и времени t , полученным по калиброванной модели $N_{sw}(t) = e^{-0,16t}$.

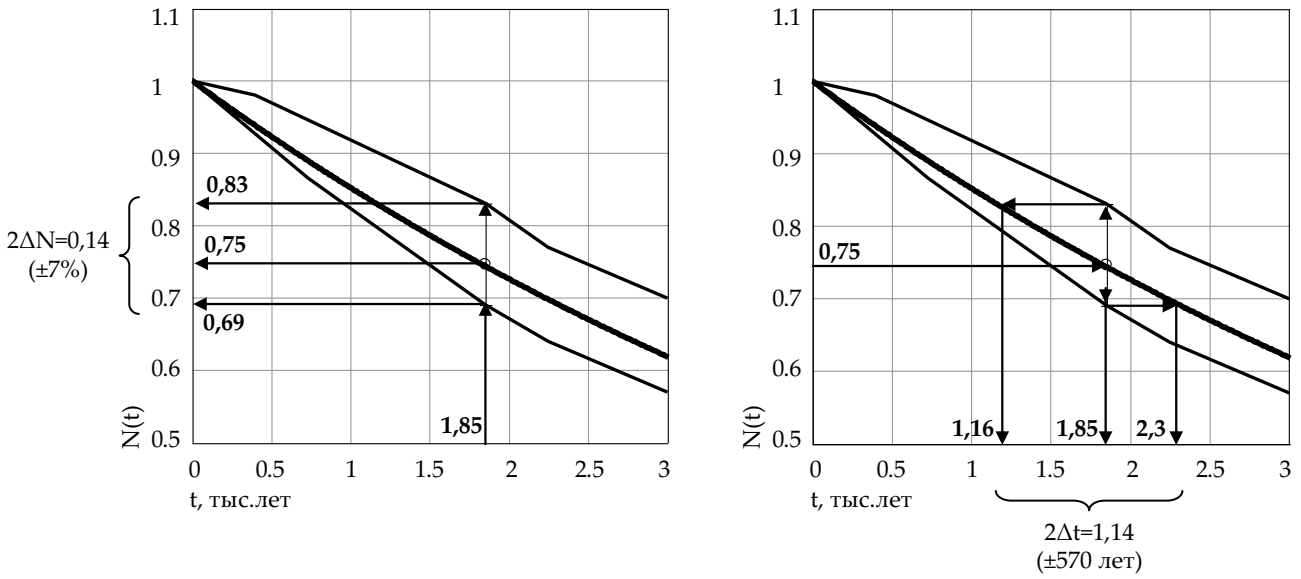
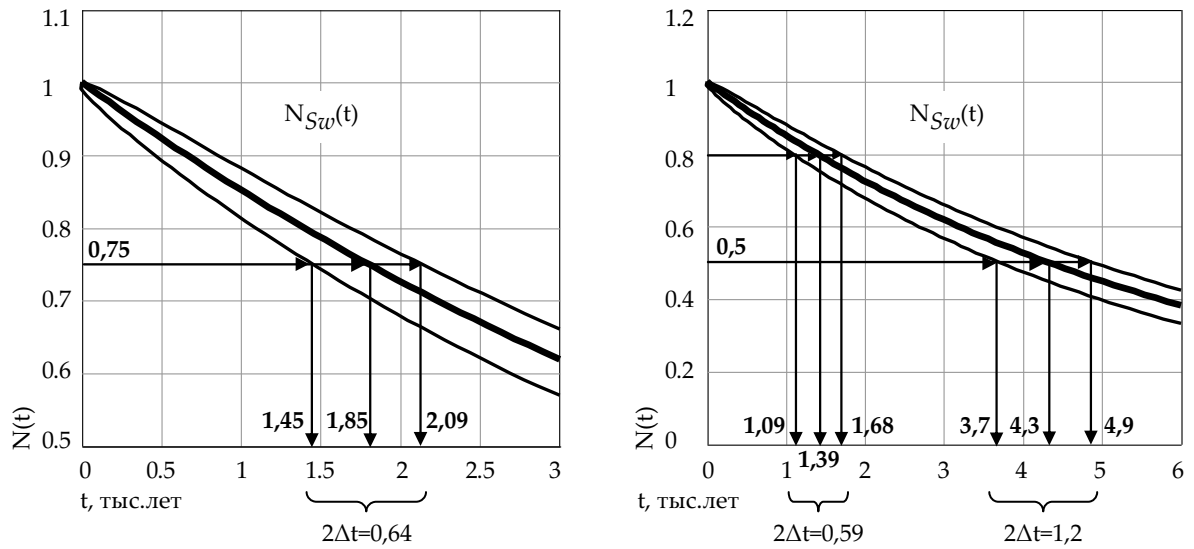


Рисунок 11 а, б. Доверительный интервал, рассчитанный для усредненной модели $N_{sw}(t) = e^{-0,16t}$ с коэффициентом $\lambda=0,16$ и заданной вероятностью 0,7.



Используя диаграммы на рис. 10 (а и б), можно численно оценить разброс процентов совпадений для выбранного значения времени или наоборот — интервал неопределенности датировки для известного процента совпадений между сравниваемыми языками. Например, для временного отрезка 1850 лет разброс доли совпадающих значений составит около 14% (т.е. в среднем ± 7 слов при использовании 100 словных списков). Аналогичным образом для доли совпадений $N(t)=0,75$ диапазон временной неопределенности составит 1140 лет (± 570 лет).

Изложенный способ оценки точности глоттохронологических моделей подразумевает наличие достаточно большого объема фактических данных о разбросе процентов

Рисунок 12. Доверительный интервал, рассчитанный для усредненной модели $N_{Sw}(t) = e^{-0,16t}$ с коэффициентом $\lambda=0,16$ и заданной вероятностью 0,95.

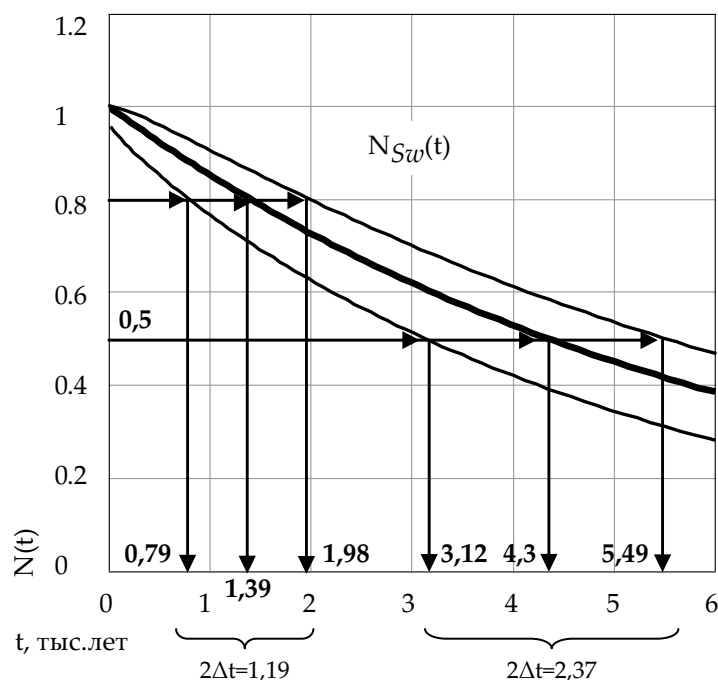


Таблица 7. Зависимость величины доверительного интервала от выбранного времени и вероятности. Приведены усредненные значения. В скобках указан процент величины интервала от значения расчетной датировки.

Расчетная датировка, лет	1000	2000	3000	4000	5000	6000
Величина доверительного интервала для вероятности $p = 0,7$	± 250 (25%)	± 360 (18%)	± 470 (16%)	± 560 (14%)	± 650 (13%)	± 750 (13%)
Величина доверительного интервала для вероятности $p = 0,95$	± 500 (50%)	± 720 (36%)	± 940 (32%)	± 1120 (28%)	± 1300 (26%)	± 1500 (25%)

совпадений сравниваемых списков и соответствующих достоверных датировок — т.е. является эмпирическим.

Вместе с тем, на основе описанного выше потокового подхода оценка точности моделей может быть произведена теоретическими методами — благодаря известным статистическим свойствам этого процесса. А именно, если представить общий поток замен в одном списке как сумму потоков замен его значений (каждый из которых является экспоненциальным), то для суммарного потока первых замен по известным формулам можно рассчитать *доверительный интервал* (Вентцель, Овчаров 1969: 235) значений, полученных с использованием модели. Зная величину доверительного интервала, вычисленную для некоторого процента совпадений, мы можем указать диапазон времени, в который с заданной вероятностью⁹ укладывается искомая датировка (рис. 11).

Например, при известной доле совпадений $N(t)=0,75$, пользуясь моделью $N(t) = e^{-0,16t}$, получаем расчетную датировку 1850 лет и доверительный интервал 640 лет, из чего сле-

⁹ Например, вероятность $p=0,7$ указывает на то, что в 70 случаях из 100 фактическая датировка будет находиться в пределах нижней и верхней границы рассчитанного доверительного интервала.

дует, что фактическое значение с вероятностью 0,7 может варьироваться в диапазоне от 1450 лет до 2090 лет (рис. 11а). С увеличением временной дистанции ширина доверительного интервала также будет расти. Так, для значения $N(t)=0,5$ и той же вероятности ($p=0,7$) его величина достигает уже 1200 лет (рис. 11б). Однако следует отметить, что в процентном отношении величина доверительного интервала постепенно снижается по мере удреждения датировок: например, от 25% (для $t=1000$ лет) до 13% (для $t > 4$ тыс. лет), (см. табл. 7, $p=0,7$). Это обстоятельство хорошо объясняет трудности глоттохронологического сопоставления списков языков с малыми глубинами расхождений.

Численно оценить величину доверительного интервала для разных значений времени и выбранной вероятности можно с помощью табл. 7 и рис.12.

Сопоставив рис. 11 с рис. 12 и значениями из табл. 7, можно убедиться, что фактический разброс значений хорошо совпадает с теоретическим доверительным интервалом, установленным для вероятности $p=0,95$, что позволяет сделать вывод о статистической адекватности используемых моделей, а также их практической пригодности для получения лингвистических датировок. При этом, однако, следует помнить, что все результаты вычислений будут иметь вероятностный характер, т.е. чем выше желаемая надежность датировки, тем больше величина ее неопределенности. На практике это означает, что при датировании лексических процессов с использованием глоттохронологии следует говорить не о конкретной дате, а о диапазоне дат с известной вероятностью. Например: «нами получена датировка 2000 ± 360 лет с вероятностью 70%» или « 2000 ± 720 лет с вероятностью 95%».

Несмотря на то, что исследование процесса лексических замен, происходящих в списке одного языка с течением времени, имеет большое теоретическое значение и составляет основу любых лексикостатистических методов, его применение на практике весьма ограничено. Действительно, случаи, когда необходимо определить временную дистанцию между языком-предком и его потомком, встречаются в компаративистике довольно редко. Гораздо более распространенной задачей является датирование разделения двух современных языков, обладающих предположительным или установленным генетическим родством. Таким образом, с практической точки зрения было бы интересно провести анализ глоттохронологических моделей, описывающих относительную языковую дивергенцию. Этому анализу, как уже говорилось выше, будет посвящена вторая часть настоящей статьи.

Тем не менее, полученные результаты позволяют сформулировать некоторые важные выводы уже по итогам первой части проведенного исследования:

1. Процесс лексических изменений, наблюдаемый в базисной лексике романских языков, наиболее корректно описывается экспоненциальной зависимостью с постоянной скоростью замен отдельных значений или частей списка. Данная зависимость реализована, в частности, в классической модели М. Сводеша, а также, в общем случае, — в потоковой модели.
2. Калибровка параметров глоттохронологических моделей позволяет добиться хорошего соответствия получаемых датировок исходным данным. При этом отдельные фактические значения могут существенно отличаться от расчетных, что свидетельствует, с одной стороны, о ярко выраженной статистической природе и неравномерности лексического процесса, а с другой — о недостаточном количестве опорных точек, используемых при калибровке.
3. Лингвистические датировки, полученные с использованием глоттохронологических методов, имеют *вероятностный характер* и представляют собой не точную величину, а диапазон (доверительный интервал) значений, к которому с известной вероятностью будет принадлежать искомая дата.

4. Величина доверительного интервала может быть определена по фактическим исходным данным, а также рассчитана теоретически — на основе установленных статистических свойств процесса лексических замен. При этом она зависит от измеряемого временного отрезка и желаемой надежности расчетных значений. По мере их увеличения доверительный интервал растет, а точность датировки, соответственно, снижается.
5. Повышение надежности лингвистических датировок, получаемых с использованием глоттохронологии, возможно за счет привлечения дополнительных исходных данных из различных языковых групп и семей, что является важной задачей для будущих лексикостатистических исследований.

Дополнительные материалы доступны на:

- <http://jolr.ru/>

Файл MS Excel содержит:

- Таблица 8. Проценты совпадений между 110-словными списками романских идиомов.

Литература

- Арапов, М. В., М. М. Херц. 1974. *Математические методы в исторической лингвистике*. Москва: Наука.
- Васильев, М. Е., А. Ю. Милитарев. 2008. Глоттохронология в сравнительно-историческом языкознании. Модели дивергенции языков. *Orientalia et Classica: Труды Института восточных культур и античности* 19: 509—536.
- Васильев, М. Е., А. И. Коган. 2013. К вопросу о восточнодардской языковой общности. *Journal of Language Relationship* 10: 149—177.
- Васильев, М. Е., Г. С. Старостин. 2014. Лексикостатистическая классификация нубийских языков: к вопросу о нильско-нубийской языковой общности. *Journal of Language Relationship* 12: 51—72.
- Вентцель, Е. С., Л. А. Овчаров. 1969. *Теория вероятностей*. Москва: Наука.
- Сводеш, М. 1960а. Лексикостатистическое датирование доисторических этнических контактов. *Новое в лингвистике* 1: 23—52.
- Сводеш М. 1960б. К вопросу о повышении точности в лексикостатистическом датировании. *Новое в лингвистике* 1: 53—87.

References

- Arapov, M. V., M. M. Herz. 1974. *Matematicheskiye metody v istoricheskoy lingvistike*. Moskva: Nauka.
- Bergsland, K., H. Vogt. 1962. On the validity of glottochronology. *Current anthropology* 3(2): 115—153.
- Kroeber, A. L. 1958. Romance history and glottochronology. *Language* 34(4): 454—457.
- Rea, J. A. 1958. Concerning the Validity of Lexicostatistics. *International Journal of American Linguistics* 24(2): 145—150.
- Starostin, S. 2000. Comparative-historical linguistics and lexicostatistics. In Colin Renfrew et al. (eds.). *Time Depth in Historical Linguistics*. Cambridge: McDonald Institute for Archaeological Research. Vol. 1: 233—259.
- Swadesh, M. 1952. Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American philosophical society* 96(4): 452—463.
- Swadesh, M. 1960а. Leksikostatisticheskoye datirovaniye doistoricheskikh etnicheskikh kontaktov. *Novoye v lingvistike* 1: 23—52.
- Swadesh, M. 1960b. K voprosu o povyshenii tochnosti v leksikostatisticheskom datirovanii. *Novoye v lingvistike* 1: 53—87.
- Vasilyev, M. E., A. I. Kogan. 2013. K voprosu o vostochnodardskoy yazykovoy obshchnosti. *Journal of Language Relationship* 10: 149—177.

- Vasilyev, M. E., A. Yu. Militaryov. 2008. Glottokhronologiya v sravnitel'no-istoricheskom yazykoznanii. Modeli divergentsii yazykov. *Orientalia et Classica: Trudy Instituta vostochnykh kultur i antichnosti* 19: 509–536.
- Vasilyev, M. E., G. S. Starostin. 2014. Leksikostatisticheskaya klassifikatsiya nubiyskikh yazykov: k voprosu o nil'sko-nubiyskoy yazykovoy obshchnosti. *Journal of Language Relationship* 12: 51–72.
- Venzel, E. S., L. A. Ovcharov. 1969. *Teoriya veroyatnostey*. Moskva: Nauka.

Mikhail E. Vasilyev, Mikhail N. Saenko. How accurate glottochronology can be? Dating the lexical replacement process in the Romance languages.

In this paper we discuss the accuracy of glottochronology, a lexicostatistical method used in the dating of linguistic divergence. Our study provides a detailed analysis of the process of lexical replacement in the basic lexicon of one language over the course of time. To measure replacement rates and determine other statistic features of lexical change we use 110-item wordlists, compiled over the past two years for 54 modern and several historically attested Romance languages. Pairwise comparison of modern wordlists with those of Archaic Latin, Late Classical Latin, Old French, and Old Italian allows to obtain several control points suitable for calibration of glottochronological equations. To estimate the time distance between the compared idioms, three different methods have been applied: the classic formula of M. Swadesh, the modified glottochronology of S. Starostin and a recently proposed approach based on simulation of lexical changes of every meaning on the Swadesh list as stationary Poisson processes. Further analysis resulted in several important conclusions concerning the following questions: (a) what are the main characteristics of lexical divergence in one language; (b) which of the existing models maps these characteristics more efficiently; (c) how precise and reliable glottochronological dating can be in general. We plan to follow this research by another study in which the process of relative divergence between two or more languages with the same ancestor will be considered.

Keywords: Romance languages, lexicostatistics, glottochronology, Swadesh wordlist.

Славянский чертежник: этимология слав. **čьrtь* ‘черт’

В настоящей работе обсуждается славянское наименование черта, которое реконструируется как **čьrtь* на общеславянском уровне, но не имеет очевидных индоевропейских когнатов. Типологические параллели из других индоевропейских мифологических традиций, в первую очередь армянские и анатолийские данные, приводят автора к заключению, что праславянское **čьrtь* может быть интерпретировано как обозначение неземного чертежника, определяющего человеческие судьбы. Постепенно этот бог судьбы стал восприниматься как ангел смерти, что и привело к его синкретизму с дьяволом — воплощением зла в христианской традиции.

Ключевые слова: этимология, черт, славянский, армянский, анатолийский.

1. Введение.

Убедительная этимология лексемы с нетривиальным семантическим развитием не должна сводиться к внешним лингвистическим соответствиям¹. Следует также представить сценарий того, как могли развиваться новые значения, в идеале сопроводив их типологическими параллелями из других языков. Корректное лингвистическое сравнение не обязательно влечет за собой правильное понимание эволюции значения. Для иллюстрации этой идеи я хотел бы обратиться к случаю, когда когнаты были подобраны корректно, но этимология опирается на неубедительную семантическую реконструкцию, и предложить лучшую альтернативу. Хочется надеяться, что данный пример окажется небезынтересным также и для индоевропейской сравнительной мифологии.

2. Славянский ‘черт’.

Рефлексами праслав. **čьrtь* являются польское *czart* ‘черт, дьявол, злой дух’, чешское и словацкое *čert* ‘черт, злой дух; беда’, верхнелужицкое *čert* ‘черт, демон’, нижнелужицкое *cart* ‘черт, дьявол’, русское *чёрт*, украинское и белорусское *чорт* ‘черт, дьявол’ (Трубачев 1977: 164). Наличие потомков этой лексемы в южнославянских языках является предметом дискуссии. Проще всего считать словенское *črt* ‘черт’ ученым заимствованием из других славянских языков (Bezljaj 1977: 89), которое могло распространиться посредством искусственно

¹ Первая версия настоящей статьи была представлена на XIV Конгрессе Indogermanische Gesellschaft в Копенгагене (2012 г.). Я благодарен участникам конгресса за комментарии и предложения, послужившие улучшению настоящей статьи. Особую благодарность я выражаю Вацлаву Блажеку (Брно), Александру Лубоцкому (Лейден), Крэгу Мелчерту (Лос Анджелес) и, особенно, Грачу Мартиросяну (Лейден). Алексей Касьян (Москва), Армен Петросян (Ереван) и Крешимир Вукович (Оксфорд) помогли мне с рядом библиографических ссылок. Ответственность за высказанные в данной работе суждения лежит, разумеется, на ее авторе. Работа над финальной версией статьи была поддержана стипендией имени Гумбольдта во время моего пребывания в Марбургском университете. Задержка с публикацией материалов конгресса, а также особый интерес, который данная работа может представлять для русскоязычного читателя, побуждают меня представить ее русскую доработанную версию для публикации в «Вопросах языкового родства».

созданного личного имени *Črtomir*². Заимствованное происхождение труднее продемонстрировать для лексемы *črt* ‘ненависть, вражда’, зафиксированной в словенском языке XIX века (см. Трубачев 1977: 164 и ссылки). Хорватское *črt* ‘черт’ также встречается в источниках XIX века, и его объяснение через раннее заимствование из словенского не является убедительным ввиду наличия этого корня в местных топонимах (Katičić 2003—2004: 260—263). Таким образом, нельзя исключать гипотезу о том, что рефлекс праславянского **čvrtь* встречался и на Балканах до того, как был там замещен различными эвфемизмами, такими, как *vrag*.

Был ли исконный **čvrtь* утерян в общеюжнославянском или в отдельных южнославянских языках, сравнение между восточно- и западнославянскими языками оправдывает его реконструкцию для праславянского³. Однако в самом праславянском языке он, по-видимому, являлся инновацией по сравнению с более архаичным обозначением злого сверхъестественного существа, сохранившемся в балтийских языках. Литовское *vėlnias* и латышское *velns* ‘черт’ связаны на синхронном уровне со словами, обозначающими духи мертвых: лит. *vėlės*, лат. *velis*. Эти лексемы относятся к тому же индоевропейскому корню, что и древнеисландское *valr*, древнесаксонское и древневерхненемецкое *wal* ‘воин, убитый на поле брани’, откуда, несомненно, происходит и название валькирий, досл. «выбирающих мертвых» (Fraenkel 1962—65: 1218—19; Smoczyński 2007: 731, 732). Несмотря на невозможность провести формальную реконструкцию общего теонима из балтийских и германских когнатов со значением бога — покровителя мертвецов, концепт такого божества вполне вероятно является по происхождению северноиндоевропейским⁴. В то же время, праслав. **čvrtь* представляется совершенно изолированным в рамках индоевропейского религиозного лексикона⁵.

3. Формально неприемлемые этимологии

Если **čvrtь* представляет собой исключительную славянскую инновацию в области индоевропейской демонологии, весьма вероятно, что его деривационная основа также сохраняется в славянских языках. Ранние гипотезы относительно славянской деривации

² Период формирования словенского литературного языка совпал с периодом творчества поэта-романтика Франца Прешерна (1800—1849), чье влияние на последующую литературу в Словении трудно переоценить. Его эпическая поэма *Krst pri Savici* («Крещение при Савице») изображает протагониста с именем *Črtomir*, формально являющимся дериватом от *črt*. Языческий воин Чртомир в конце концов принимает христианство под влиянием своей возлюбленной Богомилы. Я глубоко признателен Гашперу Бегушцу (Любляна), который подтвердил, что имя Чртомир, повидимому, не использовалось в Словении до Прешерна, несмотря на его популярность в наши дни.

³ Несмотря на то, что рус. *чёрт*, мн. *черти* засвидетельствован лишь с XVI или даже начала XVII веков, гипотеза о том, что это существительное было заимствовано из польского *czart* (Раденкович 2014) представляется неприемлемой по формальным соображениям. Чередование *e / ě* в корне и архаическое множественное число на *-и* указывают на то, что данная лексема уже существовала в общевосточнославянском.

⁴ Несколько иной подход к этимологии балтийского слова для ‘черта’ был предложен Ивановым и Топоровым (1974: 31—74). Российские ученые сравнивают лит. *vėlnias* со славянским теонимом Велес / Волос и ведическим демоном Вала, предполагая, что они представляют собой различные трансформации индоевропейского хтонического божества, главного оппонента бога грозы, а также властелина (или похитителя) скота. Доказательство этой гипотезы поднимает большой пласт вопросов, касающихся реконструкции праиндоевропейского пантеона, рассмотрение которых находится за рамками настоящей статьи. Хотелось бы, однако, подчеркнуть, что гипотеза Иванова и Топорова 1977 года только подкрепляет архаический характер лит. *vėlnias* и лат. *velns* по сравнению с праслав. **čvrtь* ‘черт’.

⁵ См. также попытку Дуковой (1984) найти мифологические существа, которые обозначались бы когнатом лексемы **čvrtь* вне славянских языков. Дукова предлагает древнеисл. *skrati* ‘тролль’ и древневерхненем. *scraz* ‘фавн’ как возможные соответствия. Это сравнение невозможно, поскольку германские формы требуют реконструкции **scrod°* (vel sim.).

*čьrtь собраны, например, в словарях Vasmer 1953—58, III: 328 и Bezlaj 1977: 89, но ни одна из них не убедила авторов соответствующих словарей и не стала более вероятной в свете новейших открытий⁶. В самом деле, две из них имеют очевидные формальные недостатки. Гипотеза о родстве с унаследованным праслав. *čьrнь 'черный' (~ вед. *kr̥ṣṇá-* 'id.') ostавляет без объяснения суффикс *-t-*, а связь с русским *терять* невозможна не только ввиду необъяснимой палатализации *k > č*, но и поскольку примеры данного глагола ограничены восточнославянскими языками. Какие бы семантические соображения ни приводились бы в пользу *черта* как «черного» или «потерянного», соответствующие этимологии сталкиваются с непреодолимыми формальными препятствиями.

То же самое можно сказать и об этимологии *čьrtь, которой отдали предпочтения Vasmer и Bezlaj (loc. cit.), исходя из допущения, что словообразовательная база славянского существительного наличествует в балтийском. Цитирую Фасмера: «Ursl. *čьrtь wird als *-to- Part. 'der Verwünschte' und für verwandt angesehen mit lit. *kyrėti* 'böse werden', *i-kỳrti* 'sich ekeln', *ap-kyrėti* 'überdrüssig werden', *i-kyrùs* 'lästig, aufdringlich', *kerėti* 'mit bösem Blick bezaubern'». Отсутствие славянских когнатов само по себе говорит против данной корневой этимологии. Что, однако, более существенно, ни одна из балтийских форм не является семантически совместимой с Фасмеровской реконструкцией *черта* как 'проклятого'. К тому же, все процитированные формы, за исключением *i-kỳrti* 'заикаться', представляют собой стивные образования на *-ē-, не образующие *-to- причастий в праславянском. Таким образом, можно лишь согласиться с оценкой Трубочева (1974—, IV: 165): «Каким образом при этом могло получиться причастие на *-to-* со значением 'проклятый' ... остается для нас загадкой, авторы, кажется, не очень считаются с семантикой и морфологией привлекаемых для сравнения основ»⁷.

Единственная этимология *čьrtь, упомянутая Фасмером и Безлаем, которую можно считать формально приемлемой, связывает эту лексему с лат. *curtus* 'короткий, обрубленный', вероятно, произведенным от и.-е. *(s)ker 'обрезать, обрубить' (Rix 2001: 556—57). С формальной точки зрения, можно предполагать, что мы имеем дело с лексическими когнатами и прямыми рефлексамы этимологического причастия *kr-to-. Эпитет *черта* «куцый», разделяемый рядом славянских традиций, мог бы представлять собой семантическую мотивацию для обсуждаемой этимологии (Черных 1993, II: 384). Против данного решения, однако, говорит тот факт, что единственным славянским глагольным рефлексом и.-е. *(s)ker является диалектное укр. *čru (Rix 2001: 557), а сама лексема *čьrtь не встречается в функции прилагательного. Заметим, что праслав. *kortькь 'короткий' произведено от и.-е. *(s)kert 'резать', вероятно, вариант и.-е. *(s)ker с расширителем, выделившийся в отдельный корень уже на праиндоевропейском уровне (Rix 2001: 559—60). Если бы *čьrtь 'куцый' > 'черт' было внутриславянским развитием, следовало бы ожидать больше когнатов данного корня в славянских языках.

4. Семантически проблематичные этимологии.

В последние десятилетия в славистических кругах завоевала популярность другая этимология *čьrtь. С формальной точки зрения, славянская лексема может представлять

⁶ Ср. также современный обзор этимологий *черта* в статье Березович, Виноградов 2012.

⁷ Следует отметить, что лит. *kerėti* 'ворожить' этимологически не связано с другими литовскими лексемами, приводимыми Фасмером. Это прозрачный когнат лит. *kerai* 'колдовство' и менее очевидный когнат лит. *kùrti* 'строить, сооружать' (Smoczyński 2007: 277). Относительно семантического перехода 'строить' > 'колдовать', см. ниже. Что касается лит. *i-kyrùs* 'назойливый, навязчивый' и родственных лексем, их индоевропейская этимология не установлена, однако они не могут являться когнатами *kùrti* 'строить, сооружать' ввиду различий в значении и в корневом вокализме.

собой отглагольное имя, производное от праслав. *čersti / čbrtq ‘чертить, бороздить’, родственное лит. kiřsti / kertù ‘рубить’. Обсуждаемый балтославянский корень является еще одним рефлексом и.-е. *(s)kert ‘резать, рубить’, упомянутого в предыдущем разделе. Подобное существительное едва ли могло быть унаследовано из праиндоевропейского, где простые тематические именные производные отглагольных корней чаще всего характеризовались *o*-вокализмом (ср. тип τόμος / τομός в древнегреческом). В истории славянских языков, однако, получили развитие вторичные тематические производные с нулевой ступенью корня. Они, как правило, соответствовали нулевой ступени корневого вокализма производящего глагола. Например, от праслав. *smьrděti / smьrdjō ‘смердеть, вонять’ было образовано первичное имя действия *smordь ‘смерд, вонь’ и вторичное имя деятеля др.-рус. smьrdь ‘смерд’, досл. «вонючка» (Vasmer 1953—58, II: 671, 676). Другое инновативное образование с нулевой ступенью корня отражено в ст.-слав. tlьkvь ‘толмач’, образованном от и.-е. корня *tlok^w, сохраненного в лат. loquī ‘говорить’ (Vasmer 1953—58, III: 115) или от праслав. *telkti / tlьkq ‘толочь’ (Vaillant 1974: 249)⁸. Еще чаще инновативные производные с нулевой ступенью корня используются в значении имен действия или объекта, ср., напр., др.-рус. tlьkvь ‘толк’ (Vasmer 1953—58, III: 115). Другим дериватом подобного же рода является сербскохорватское и словенское *o*-črt ‘контур’, относящееся к тому же корню, что и вышеупомянутое *čersti / čbrtq ‘чертить, бороздить’ (Vaillant 1974: 239). Наконец, славянские производные *a*-основы с нулевой ступенью корня, подобные *čbrta ‘черта’ (Vasmer 1953—58, III: 328) можно рассматривать как дополнительный вариант той же словообразовательной модели. Очевидно, что если от глагольной основы можно было произвести вторичное *nomen rei actae* с нулевой ступенью корня, от нее также можно было произвести и *nomen agentis* с нулевым корневым вокализмом.

Таким образом, сопоставление между славянским обозначением злого духа и производящего глагола со значением ‘чертить, бороздить’ является формально безупречным. В самом деле, праслав. *čersti / čbrtq является единственной распространенной глагольной основой, от которой может быть образовано производное существительное *čbrtь. Однако, семантика предложенной деривации ни в коей мере не является очевидной и в прошлом служила основанием для различных интерпретаций.

Первая такая попытка была предпринята в работе Brückner 1918: 174 и разработана в статье Jacobson 1959: 276 (без ссылки на Брюкнера). Согласно Брюкнеру, *čbrtь первоначально обозначало ‘колдун’. И Брюкнер, и Якобсон связывают данное существительное с праслав. *čbrta ‘черта’, а также *čara ‘черта’, засвидетельствованном в западославянских. Якобсон определяет чеш. čára как «a borderline up to which something is permitted or magically prohibited, e.g. the line which marks the so-called magic circle (where the evil demons retain or lose their power)». Мы увидим, что волшебство, осуществляемое за защитной чертой, релевантно и для моих собственных построений. Но этимологическая связь между *čara и *čbrta представляется проблематичной, поскольку *čara непосредственно связана со ст.-слав. čarь ‘колдовство, чары’, лит. kerėti ‘ворожить’ и другими архаичными производными от и.-е. *k^wer ‘строить, сооружать’, не сохранившегося или плохо сохранившегося в виде финитного глагола в праславянском. В качестве типологической параллели к подобному семантическому переходу можно привести др.-исл. gørningar ‘дела, деяния, магия, колдовство’, производное от gørva ‘делать’ или скр. kr̥tya- ‘дело, деяние, магия, колдовство’ от kar ‘делать’, в конечном итоге восходящего к тому же и.-е. *k^wer ‘строить, сооружать’. Таким образом, связь между чеш. čára и черчением, вероятно, вто-

⁸ Семантический переход ‘толочь’ > ‘переводить’ обсуждается в работе Vaillant 1974 (loc. cit). В качестве дополнительной типологической параллели ср. рус. *молоть чепуху*.

рична, как и связь между *čьrta ‘черта’ и колдовством, хотя, как мы увидим, вторая ассоциация и восходит к праславянской эпохе. Поэтому в той мере, в какой сценарий Брюкнера и Якобсона поддается реконструкции, он представляется основанным на народной этимологии⁹.

Другая реконструкция того, каким образом *čьrtь может быть связан с *čьrsti / čьrtq ‘чертить, бороздить’, была детально разработана Трубачевым (Трубачев 1974—, IV: 166). Отправной точкой данного сценария является слов. čft ‘раскорчеванное место, межа’. Трубачев полагает, что данная лексема совместима со значением производящего глагола ‘копать’, каковое значение сосуществовало со значением ‘чертить, бороздить’ или даже ему предшествовало. Если это так, то *čьrtь мог исходно являться именем деятеля со значением ‘землекоп’, уместным для подземного духа. Особым указанием на склонность черта к земляным работам является распространенный топоним *čьрто-ryjь, досл. ‘прорытый чертом’, употребляемый по отношению к различным оврагам и промоинам (Трубачев 1974—, IV: 163—4). Наконец, в той же работе приводятся другие предполагаемые производные от обсуждаемого корня, а именно, ц.-слав. krьtь ‘крот’ и лит. kertùkas ‘землеройка’, которые также могли исходно значить ‘землекоп’ (Трубачев 1974—, XIII: 58—9).

Оценка данной этимологии напрямую зависит от убедительности реконструкции значения ‘копать’ для *čьrsti / čьrtq в праславянском. Насколько я могу судить, в анализе Трубачева она не подкрепляется вескими аргументами. Прямое сравнением между праслав. *čьrtь ‘черт’ and *krьtь ‘крот’ невозможно по формальным причинам, как и признает и сам Трубачев, поскольку первая лексема отражает палатализацию *k > č перед рефлексом слогового *r̥, тогда как вторую можно возводить только к *krutu/o¹⁰. Спекуляции относительно «единства корня kūr и ker», подкрепляемые ссылками столетней давности, могут лишь скомпрометировать современное лингвистическое исследование. Словенский ландшафтный термин čft, несомненно, этимологически связан с праслав. *čьrsti / čьrtq, однако, не требует изоцирковой семантической реконструкции, поскольку его значение тривиально выводится из производящего глагола ‘рубить (дерева)’ или ‘бороздить (межу)’. С другой стороны, хотя демоны и ассоциируются с подземным миром в славянской и во многих других традициях, отсюда отнюдь не следует, что их основной функцией считались земляные работы. Для древних славян было естественным приписывать овраги и промоины (*čьрто-ryjь) дьяволу, а не Богу по причине их уродливого вида и потенциального вреда для сельского хозяйства. Однако славяне также использовали рефлекс *čьрто-polxь досл. ‘чертов ужас’ для различных неприятных растений и ядовитых грибов (Трубачев 1974—, IV: 163). Суммируя вышеизложенное, гипотеза Трубачева является формально ущербной и семантически избыточной.

⁹ С чисто формальной точки зрения, *čьrtь можно реконструировать как страдательное причастие *k^wr̥-to- от того же корня, что и ст.-слав. čarь ‘колдовство, чары’. Недостатком данной этимологии является, с одной стороны, отсутствие финитных рефлексов и.-е. *k^wer в праславянском, а с другой, сложности с возведением обозначения злого духа к страдательному причастию ‘заколдованный’ (скорее, ожидалось бы ‘колдующий’!). Заметим, что *čьrtь не мог быть произведен от дескриптора ‘находящийся за чертой’, поскольку здесь ожидалось бы префиксальное образование, напр., **za-čьrtь или **otь-čьrtь.

¹⁰ Напротив, лит. kertùkas ‘землеройка’, действительно, может являться производным от kīrsti / kertū ‘рубить’ (Fraenkel 1962—65: 223 sub kařstas). Этого, однако, недостаточно для реконструкции исходного значения ‘копать’ для более ранней стадии семантического развития kīrsti / kertū, поскольку обозначение землеройки вполне могло быть основано на метафоре разрубания земли. Сходным образом, англ. cutter ‘катер’ не предполагает, что исходным значением производящего глагола to cut является ‘плыть’.

5. Предпосылки для семантического анализа.

Общей особенностью этимологий, обсуждавшихся в предыдущем разделе, было то, что все они исходят из специфических особенностей сверхъестественного существа, известного как черт. Такой подход, разумеется, вполне оправдан, если иметь дело с хорошо сохранившимися религиозными традициями. Его достоинство менее очевидно в применении к славянской религии, которая перестала существовать как таковая с приходом христианства и не оставила за собой первичных письменных источников. Хотя некоторые славянские демоны и сохранили за собой специфические функции после христианизации Восточной Европы, это в меньшей мере относится **čbrtbʹu*, который подвергся синкретизации с падшими ангелами христианства и их предводителем Дьяволом / Сатаной. В частности, ни защитная черта, упоминаемая Якобсоном, ни хтонические ассоциации черта, используемые Трубачевым в своей работе, не являются особенностями славянского демона, а, скорее, отражают общие черты средневековой христианской демонологии. Поэтому ни одна из этих особенностей, вероятно, не была связана с **čbrtbʹom* в дохристианской религии славян.

По существу, для реконструкции исходной функции **čbrtbʹa* существенны лишь два соображения. С одной стороны, поскольку **čbrtb* не является калькой какого-либо известного эпитета Сатаны, данный термин, вероятно, уже употреблялся славянами-язычниками для обозначения сверхъестественного существа. С другой стороны, это имя или титул не имеет очевидных индоевропейских когнатов, относящихся к той же сфере, а его внутренняя форма предполагает реконструкцию производного имени деятеля от производящей основы **čersti / čbrtq* ‘чертить, бороздить’. Поэтому **čbrtb* может быть реконструирован как «чертежник», исходя из чисто структурных соображений. Для того же, чтобы оценить вероятность данной гипотезы, следует опираться на типологические параллели. Если подобных «чертежников» можно обнаружить в других религиозных традициях с коннотациями, способствующими их демонизации, это может как подкрепить обсуждаемую этимологию черта, так и помочь реконструировать функции **čbrtbʹa* в дохристианском славянском пантеоне.

6. Иранские параллели?

Одна из потенциальных параллелей засвидетельствована в иранских языках. Иранский глагольный корень **kart* ‘резать’, рефлекс того же самого и.-е. **(s)kert*, который дал и праслав. **čersti / čbrtq*, приобрел вторичное терминологическое значение ‘стряпать (творить) злые существа’ в зороастрийской традиции. Обычным творцом в таких случаях является Ахра Манью / Ахриман — воплощение и создатель зла. В авестийском языке сотворение злых существ обычно выражается глаголом *kart* с префиксом *fra-* или его производными (добрые акты творения передаются глаголом *θβarš* с теми же приставками). Например, оппонент героя Фраетаона описывается как «весьма могучий демон-друдж, которого состряпал (*fracā kərəntat*) Ахра Манью против материального мира, ради сокрушения творений Правды» (Ясна 9.8, ср. Pirart 2004: 66—67). Родственный среднеперсидский глагол *kirrēn-* ‘стряпать (творить) злые существа’ используется без префикса (Расторгуева и Эдельман 2000 —, IV: 312)¹¹.

Существуют обильные свидетельства доисторических контактов между иранцами и славянами, отразившиеся в заимствовании ряда религиозных терминов, таких как **bogʹь*

¹¹ Можно осторожно предположить, что иранский корень **kart* скрывается в первой части среднеперсидского композита *kerdagār* (пехлеви ⟨krtkʹl⟩, манихейское ⟨kyrdgʹr⟩) ‘творец’ (ср. MacKenzie 1971: 49, где данное слово переводится как ‘могучий’).

‘бог’ and **rajь* ‘рай’¹². Тем не менее, гипотеза о заимствовании славянами зороастрийского концепта злого творца сталкивается с непреодолимыми трудностями. С одной стороны, нет никаких оснований полагать, что те иранцы, от скифов до аланов, которые находились в контакте со славянами, следовали учению Заратуштры. Вместе с тем, маловероятно, что терминологическое использование иранского корня *kart* для сотворения злых существ, или даже религиозная доктрина о такого рода деятельности, существовали до развития зороастрийского дуализма. С другой стороны, семантика праслав. **čersti / čьrtь* не содержит и тени указания на акт творения и не обладает пейоративными коннотациями, тогда как авестийский язык не сохранил производного имени деятеля для злого демиурга. Данная комбинация исторических и лингвистических нестыковок заставляет меня пойти по другому пути в поисках славянского чертежника.

7. Анатолийские чертежницы.

Сравнительная информация иного рода доступна из анатолийских текстов бронзового века. В анатолийских языках встречаются генетически родственные теонимы, такие как хеттское ^d*Gulses* (мн.ч.), лувийское ^d*GUL-za-* = /*kwanza-*/ (ед.ч), и палайской ^d*Gulzannikes* (мн.ч.), связанные с лувийским глагольным корнем, засвидетельствованным в хеттском *guls-* и лувийском *GUL-za(i)-* ‘царапать, писать’, а также в лувийском *GUL-zattar* ‘чертеж, черновик’ (ср. Kloekhorst 2008: 492—93)¹³. Многочисленные контексты, в которых хетт. ^d*Gulses* встречается в клинописных текстах из Хаттусы, позволяют неплохо реконструировать их функции (ср. Naas 1994 и, в особенности, Otten 1971). Данная группа богинь происходит из подземного мира, привлекается к заботе о новорожденных и часто ассоциируется с месопотамской богиней-матерью DINGIR.MAḪ. Их основной задачей было, однако, записывать (*guls-*) человеческие судьбы. Вероятно, по этой причине их иногда называли «злыми» (KUB 58.108 i 5), тогда как к неудачнику можно было обратиться с утешением ‘тебя обидели Гулсы’ (KUB 23.85 rev. 6). Кроме того, они определяли время смерти каждого человека. Аккадско-хеттская билингва содержит аккадское выражение *a[-na] u₄-mi [ši-ma-]ti-ka* ‘в день твоей судьбы’, чему в хеттском тексте соответствует *GIM-an=ma=(t)ta* ^d*Gulsas* UD.KAM-*us tianzi* ‘когда для тебя наступят дни Гулс’ (KBo 12.70 obv.² 14 ff., ср. Taracha 2000: 186, fn. 89)¹⁴.

¹² Ряд ученых утверждал, что праслав. **bogь* ‘бог’ не заимствован из праир. **baga-* ‘бог’, а является его когнатом (ср. Трубачев 1974—, II: 161—63). Решающий аргумент против исходного генетического родства между этими двумя терминами оказался доступным с открытием закона Винтера. В качестве *tertium comparationis* в данном случае выступает ведийский теоним Бхага, подтверждающий непридыхательный характер индоиранского *-g-. Поэтому славянский когнат данной лексемы звучал бы как ***bagь* с рефлексом компенсаторного удлинения.

¹³ В соответствии с гипотезой, развитой в работе Waal 2014, теоним ^d*Gulses* следует читать как ^d*GUL-ses*, где *GUL* — шумерограмма с базовым значением ‘ударять’. Ваал также предлагает сопоставить лувийский глагол, известный как *GUL-za(i)-* в клинописной передаче, с глаголом *REL-za-* (= *kwal-i-za-*) иероглифических надписей и приписать получившемуся глаголу терминологическое значение ‘писать иероглифами’. Моя реакция на данную гипотезу, которую я принимаю для лувийского, но не для хеттского языка, представлена в работе Yakubovich 2013. Для целей настоящей работы достаточно отметить, что принятие или отклонение гипотезы Ваал не отменяет этимологической связи между анатолийскими именами божеств судьбы и глаголом ‘писать’. Между тем, для аргументации настоящего раздела существенна лишь данная этимология.

¹⁴ Перевод Тарахи отличается в случае хеттского предложения: ‘когда Гулсы установят твои дни’. Перевод, приведенный выше, любезно предложенный мне К. Мелчертом, предполагает, что сказуемым в данном предложении является *tiya-mi* ‘ступать’, а не *dai / tiya-hi* ‘ставить’. Данная гипотеза является предпочтительной, поскольку окончания именительного падежа множественного числа на *-as* в хеттском являются редкими и нерегулярными, тогда как окончания на *-us* широко распространены и даже обобщены в позд-

Синхронная связь между именами анатолийских богов судьбы и анатолийским глаголом для письма требует отдельной дискуссии. Хеттская цивилизация в позднем бронзовом веке была знакома с грамотой и даже использовала две различные системы письма: месопотамскую клинопись и туземные анатолийские иероглифы. Тем не менее, не-тривиальные фонетические различия между хетт. ^d*Gulses*, лув. ^d*GUL-za-* = /kwanza-/, и пал. ^d*Gulzannikes* предполагают, что если эти теонимы действительно связаны друг с другом, речь не идет о недавних заимствованиях, а скорее о рефлексах общеанатолийского прототипа или результате древней диффузии. Каким же образом могли пониматься имена божеств судьбы в конце III — начале II тыс. до н.э.? Можно предположить, что праанатол. **guls-* (vel sim.) обладал значением ‘чертить, рисовать пиктограммы’, тогда как производный теоним имел значение ‘чертежник’ или ‘чертежница’, представляя тем самым типологическую параллель к праслав. **čbrtvь*.

Разумеется, невозможно строго доказать, что анатолийским мойрам приписывалась способность влиять на человеческую судьбу посредством черчения или манипуляции пиктограммами. Однако, известно, что сами хетты пользовались подобными средствами, чтобы получить представление о будущем. Популярным методом гадания среди хеттов была клеромантия, известная также как KIN-оракулы в хеттологической литературе. Сущностью данной практики, проинтерпретированной в работе Archi 1974, была манипуляция случайными символами при помощи случайного «медиума», приводящего их к случайной цели. Символы, «медиумы» и цели могли быть как положительными, так и отрицательными. К последней группе принадлежали, например, такие понятия, как Враг, Зло или Великий грех (термин для обозначения смерти царя или царицы). Archi (ibid: 130—131) осторожно предполагает, что процедура гадания заключалась в наблюдении над животными. Определенное животное могло попасть на площадку для гадания через те или иные ворота, тем самым взяв на себя функции «медиума», дотронуться до определенных символов и метафорически привести их к определенной цели, то есть к выходу с площадки для гадания. Можно предположить, что набор пиктограмм использовался для обозначения различных входов и выходов, а также символов на площадке для гадания. Насколько можно судить, подобный тип гадания был крайне редким на Ближнем Востоке за пределами Анатолии. Все вышеизложенное согласуется с гипотезой о том, что идея «предначертания судьбы» предшествовала грамотности на территории Малой Азии.

8. Черт и гадание у славян.

Теперь уместно обратиться к предсказанию будущего у славян в дохристианскую эпоху. Ключевая информация по данному вопросу содержится в полемическом трактате *О письменах*, предположительно написанном в 893 году монахом Храбром, учеником св. Мефодия, но дошедшем до нас в более поздних списках. Он открывается словами: «Прежде славяне не имели книг, а читали и гадали чертами и резами, будучи язычниками»¹⁵. Можно оставить в стороне спекуляции о дохристианской грамоте среди славян, поскольку навыки чтения будущего не требуют грамотности в обыденном смысле, как это уже было отмечено в связи с ситуацией в Анатолии. У нас нет точной информации

нехеттском для определенных морфологических классов, включая сюда и консонантные основы, к которым принадлежит UD.KAM-us = *siwattus*.

¹⁵ Ср. старославянский текст: *Прѣжде оубо словѣне не имѣахъ книгъ, нѣ чрѣтами и рѣзами чѣтѣахъ и гатаахъ погани сѣще*. Относительно возможных этимологических связей ст.-слав. *rězy*, см. Топоров 1974: 12—19.

о том, как черты и «резы» использовались при гадании. Однако нет оснований подвергать сомнению свидетельство Храбра, который жил на рубеже христианизации славян и поэтому должен был обладать определенными познаниями в их дохристианских традициях.

Ц.-слав. *čьrta* 'черта', упоминаемая Храбром как инструмент гадания, исторически связана с праслав. *čьrtь, если его первоначальным значением было «чертежник». Соотношение между этими двумя существительными функционально напоминает соотношение лув. *GUL-zattar* 'чертеж, черновик' и *GUL-za* 'божество судьбы'. Развивая данную параллель, можно предположить, что праслав. *čьrtь также считался божеством судьбы. Подобно тому, как в праанатолийском глагол **guls-* использовался для предначертания судьбы, праслав. глагол *čьrtati мог использоваться в том же значении (ср. рус. *предначертать*). Начертание пиктограмм, возможно, ассоциировавшееся с процессом предначертания судьбы у древних анатолийцев и древних славян, могло также использоваться гадателями, желавшими проникнуть в замыслы предначертателей судьбы. В империи Хаттусы подобная гадательная практика приняла облик оракулов KIN, тогда как славянский источник сохранил лишь смутное упоминание о чертах и «резах».

С другой стороны, связь между гаданием, чертом и защитной чертой хорошо отражена в русском фольклоре. Несмотря на неодобрение православной церкви, в сельской местности была широко распространена практика гадания из-за защитной черты, в особенности среди девушек, желавших увидеть своего суженого или узнать его имя. Андрей Трофимов (Москва / Марбург) любезно поделился типичным заклинанием, сопровождающим данный тип гадания, которое он записал в Каргопольском районе Архангельской области: *Чёрти, по черте. По закругу три чёрта, в кругу нет никого*¹⁶. Предполагалось, что чертей можно принудить выдать заветную информацию, если ритуал правильно совершался на сакральном пространстве за защитной чертой, однако при нарушении правильной процедуры ритуала черти могли навредить девушкам¹⁷. Из-за большого промежутка времени к любым параллелям между современными ритуалами и гаданиями древних славян следует относиться с осторожностью. Существенно, однако, что русский фольклор наделяет черта знанием будущего, что как будто бы не вытекает из других его известных свойств. Вполне вероятно, что данная особенность могла быть унаследована от его языческого предшественника.

Для понимания того, как концепт предначертания судьбы мог независимо распространиться в Анатолии бронзового века и в юго-восточной Европе в эпоху раннего средневековья, следует напомнить, что оба региона развивались в тени письменных цивилизаций, соответственно, Месопотамии и Византийской империи. Языковые сообщества, не обладающие навыками письма, а лишь смутными представлениями о данной практике, легко могут ассоциировать его с волшебством. Классическим мотивом в поддержку данного наблюдения является история Беллерофонта, отправленного в Ликию с запечатанной табличкой, содержащей инструкцию по его собственному убийству. Данный эпизод является единственным упоминанием письма в Илиаде (II. VI: 166—170), причем содержимое таблички определяется там как «злосоветные знаки» (σῆματα λυγρὰ). Было бы неудивительно, если бы все письменные тексты воспринимались сходным образом в

¹⁶ Данное заклинание было записано в этнолингвистической экспедиции РГГУ в 1995 году. Информантом была М. В. Крапивина (1918 г.р.)

¹⁷ Это не единственный пример применения защитной черты в восточнославянской магии. Как любезно напомнил мне М. Живлов, русская идиома *очертя голову* восходит к жесту древних воинов, проводивших условную защитную черту головы перед тем, как ринуться в бой.

гомеровской Греции, где микенская слоговая письменность уже вышла из употребления, алфавит финикийского происхождения еще не приобрел популярности, но тем не менее имелись смутные представления об идее письма¹⁸. Сходным образом, поверхностное знакомство с идеей передачи письменных сообщений могло побудить как анатолийцев, так и славян приписать им магические свойства, а покровителями данного типа магии стали, соответственно, ^d*Gulsēs* и *čbrtv¹⁹.

9. Армянский чертежник.

Обсуждаемый сценарий вызывает естественный вопрос о путях демонизации *čbrtv'a. Можно предположить, что функция ангела смерти, естественная для божества судьбы, могла играть ключевую роль в данном процессе. Подобная функция уже упоминалась в контексте хеттских ^d*Gulsēs*, а одна из греческих мойр, Атропос, специализировалась на перерезании нити жизни. Здесь, однако, уместно остановиться еще на одной параллели, которая иллюстрирует пейорацию божества судьбы в контексте христианской культуры.

Согласно Расселу (Russell 1987: 305), армянская мифологическая фигура *Groł*, досл. «писатель» (ср. арм. *grel* 'писать'), представляет собой адаптацию западноиранского бога Тира, покровителя писцов и помощника при толковании сновидений. Вероятно также, что он воплотил черты урартского бога ^d*Huṭuini*, если последнего действительно можно отождествить с богом судьбы (ср. Petrosyan 2007: 181)²⁰. В средневековом армянском фольклоре *Groł* обладал четкими функциями: он записывал человеческие дела, забирал души тех, для кого настал час смерти, и передавал их на небо для суда²¹. Это была почтенная роль, которая привела к его отождествлению с архангелом Гавриилом в армянской придворной поэзии (Russell 1987: 302—303). Однако в современном армянском фольклоре *Groł* обладает совершенно иными коннотациями. Его боятся и ненавидят как ангела смерти, что и приводит к его частому появлению в проклятиях (ср. Russell 1987: 310). Два современных восточноармянских мягких ругательства, упоминающих *Groł* > *Groy* находят прямые параллели в русском языке: арм. *groyə tani* 'черт возьми' и арм. *groyə k^hez het* 'черт с тобой'. Данное сходство, возможно, вызвано контактным схождением

¹⁸ Возможным косвенным свидетельством негативных коннотаций индоевропейского корня **peik-* 'рисовать, писать' возможно, являются начальные строчки гимна Сапфо, посвященного Афродите, в котором содержится призыв к «бессмертной Афродите, обладающей расписным треном» (Ποικίλοθρον', ἀθάνατ' Ἀφροδίτα) не подавлять дух поэтессы. Можно лишь задаваться вопросом, обладало ли греческое прилагательное ποικίλος 'расписной' путающими ассоциациями в лесбийском диалекте древнегреческого. Я благодарен Н. Эттингеру (Эрланген) за данную параллель, предложенную с должной осторожностью.

¹⁹ Независимым доводом в пользу пассивного знакомства древних славян с письменностью до Св. Кирилла и Мефодия является германское происхождение общеславянского **buky* 'буква (ед.); книга, документ (мн.)'. Германские формы, обычно приводимые в этой связи, включают готское *boka* 'буква (ед.); книга, документ (мн.)' и древневерхненемецкое *buoh* 'буква (ед.); книга (мн.)'. Данные германские существительные являются производными от названия бука, что обычно объясняется в контексте использования буковой коры для древнегерманских рунических и алфавитных надписей (Pronk-Tiethoff 2013: 80—82). Как отметил один из рецензентов, сами германские племена могли наделять руны магическими чертами, ср. др.-исл. *rún* 'буква (ед.); магия (мн.)'.

²⁰ Функции ^d*Huṭuini* не могут быть восстановлены по контекстам, хотя этимологически можно сравнить урартское абстрактное существительное *huṭutuhi* с позитивным значением ('счастье, удача'(?)). Основным аргументом в пользу интерпретации *Huṭuini* как бога судьбы является сходство его имени с хурритскими теонимами ^d*Hutena* и ^d*Hutellura*, которые, в свою очередь, принадлежат к функциональным эквивалентам хетт. ^d*Gulsēs*, согласно текстам из Хаттусы (Otten 1971).

²¹ Армен Петросян (личн. сообщ.) напоминает мне о внутренней форме древнеармянского *čakata-gir* 'судьба', досл. 'то, что написано на лбу'.

между армянской и русской эксплетивной идиоматикой, но такой процесс едва ли был бы возможен без предварительного отождествления *Groγ*'а и черта.

Тот же маршрут семантической пейорации можно предположить и для славянского Чертежника. Из всех событий человеческой жизни, предопределяемых судьбой, славяне, вероятно, ассоциировали с чертом самое последнее. Не исключено, что это было результатом синкретизма между чертом и иным божеством, относящимся к миру мертвых, возможным когнатом лит. *vėlnias*²². Тем не менее, роль мифологического ангела смерти едва ли более способствует улучшению репутации, чем роль палача в этом подлунном мире. От нее был один лишь шаг до отождествления с воплощением зла, что и произошло с приходом христианства. Единственное, что осталось от старой функции черта — это, возможно, его связь с гаданием.

10. Заключение.

Сравнительная мифология является областью, где строгие доказательства едва ли возможны. Хотелось бы, однако, надеяться, что предлагаемая этимология праслав. *čьrtь обладает двумя ключевыми преимуществами перед альтернативами, обсуждаемыми в начале статьи. Во-первых, она постулирует производящую основу *čьrsti / čьrtь 'чертить, бороздить', которая хорошо засвидетельствована в славянских языках и прозрачный дериват со значением 'чертежник', который хорошо согласуется с правилами славянской морфологии. Во-вторых, все постулируемые гипотезы относительно религии славян находят типологические параллели в других индоевропейских традициях. Обозначение бога судьбы как 'чертежника' обнаруживает полную параллель в анатолийских языках и приблизительную параллель в армянском, тогда как предполагаемое развитие бога судьбы в демоническую фигуру прямо повторяет эволюцию армянского *Groł*. Некоторые альтернативные этимологии для праслав. *čьrtь также можно считать формально безупречными, другие являются семантически удовлетворительными, но ни одна из них не обладает обоими качествами. Поэтому предлагаемая реконструкция славянского чертежника является наилучшим доступным решением.

Литература

- Березович, Е. Л., Л. Н. Виноградова. 2012. Черт. *Славянские древности*, под ред. Н. И. Толстого. Том 5: 519—527. М.: Международные отношения.
- Дукова, Уте. 1984. Праслав. *čьrtь 'черт, злой дух' / герм. *skrat*- 'лесной дух, черт. *Этимология* 1982: 61—3. М.: Наука.
- Черных, П. Я. 1994. *Историко-этимологический словарь современного русского языка*. 2-е изд. М.: Русский язык.
- Иванов, В. В., В. Н. Топоров. 1974. *Исследования в области славянских древностей: лексические и фразеологические вопросы реконструкции текстов*. М.: Наука.
- Раденкович, Л. 2014. Черт и/или межевой. Наименование беса «черт» у славян. *Wiener Slawistisches Jahrbuch* (Neue Folge) 2: 152—162.
- Расторгуева, В. С., Дж. И. Эдельман. 2000—. *Этимологический словарь иранских языков*. Т. 1—. М.: Наука.
- Топоров, В. Н. 1974. Славянские комментарии к нескольким латинским архаизмам. *Этимология* 1972: 3—19. М.: Наука.
- Трубачев, О. Н. (ред.). 1974—. *Этимологический словарь славянских языков*. Т. 1—. М.: Наука.

²² В этой связи уместно упомянуть демонизацию славянского бога Велеса / Волоса в ряде традиций, в особенности в древнечешской (Иванов, Топоров 1974: 66). Можно предположить, что частичный синкретизм черта и Велеса имел место уже в дохристианский период, вне зависимости от того, принимать ли этимологическую связь между праслав. *welsь и лит. *vėlnias*.

References

- Archi, Alfonso. 1974. Il sistema KIN della divinazione hittita. *Oriens Antiquus* 13: 113–144.
- Berezovich, E. L., L. N. Vinogradova. 2012. Chert. *Slavyanskije drevnosti*, pod red. N. I. Tolstogo. Tom 5: 519–527. Mocsow: Mezhdunarodnye otnosheniya.
- Bezljaj, France. 1977. *Etimološki slovar slovenskega jezika*. Vol. 1: A–J. Ljubljana: Slovenska akademija znanosti in umetnosti.
- Brückner, Aleksander. 1918. Über Etymologien und Etymologisieren. *Zeitschrift für vergleichende Sprachforschung* 48 (3/4): 161–229.
- Chernykh, Pavel J. 1994. *Istoriko-etimologičeskij slovar sovremennogo russkogo jazyka*. 2nd ed. Moscow: Russkij jazyk.
- Dukova, Ute. 1984. Praslav. *čbrtv ‘chert, zloj dukh’/germ. *skrat-* ‘lesnoj dukh, chert’. *Etimologija* 1982: 61–3. Moscow: Nauka.
- Fraenkel, Ernst. 1962–65. *Litauisches etymologisches Wörterbuch*. 2 vols. Heidelberg: Winter.
- Haas, Volkert. 1994. *Geschichte der hethitischen Religion*. Leiden: Brill.
- Ivanov, Vyacheslav V., Vladimir N. Toporov. 1974. *Issledovanija v oblasti slavianskikh drevnostej: leksičeskije i frazeologičeskije voprosy rekonstrukcii tekstov* Moscow: Nauka.
- Jacobson, Roman. 1959. Marginalia to Vasmer’s Russian etymological dictionary (p-я). *International Journal of Slavic Linguistics and Poetics* 1959 (1/2): 266–277.
- Katičić, Radoslav. 2003–2004. Uz rub rječnika hrvatskoga kajkavskoga kniževnog jezika. *Folia Onomastica Croatica* 12–13: 259–274.
- Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- MacKenzie, David N. 1971. *A Concise Pahlavi dictionary*. Oxford: Oxford University Press.
- Otten, Heinrich. 1971. Gulš-Gottheiten. *Reallexikon der Assyriologie* 3: 598.
- Petrosyan, Armen. 2007. State pantheon of Greater Armenia: earliest sources. *Aramazd (Armenian Journal of Near Eastern Studies)* 2: 174–201
- Pirart, Éric. 2004. *L’éloge mazdéen d’ivoresse : édition, traduction et commentaire du Hōm Stōd*. Paris: L’Harmattan.
- Pronk-Tiethoff, Saskia. 2013. *The Germanic Loanwords in Proto-Slavic*. Amsterdam: Rodopi.
- Radenković, Ljubinko. 2014. Chert i/ili mezhevoj. Naimenovanje besa «chert» u slavjan. *Wiener Slawistisches Jahrbuch (Neue Folge)* 2: 152–162.
- Rastorgujeva, Vera S. & Dzhoj. I. Edelman. 2000–. *Etimologičeskij slovar iranskikh jazykov*. Moscow: Nauka.
- Russel, James A. 1987. *Zoroastrianism in Armenia*. Harvard University Press.
- Rix, Helmut et al. (eds.). 2001. *Lexicon der indogermanischen Verben*. 2nd ed. Wiesbaden: Reichert.
- Smoczyński, Wojciech. 2007. *Słownik etymologiczny języka litewskiego*. Vilnius: Vilniaus Universitetas.
- Taracha, Piotr. 2000. *Ersetzen und Entsühnen. Das mittelhethitische Ersatzritual für den Grosskönig Tuthalija (CTH *448.4) und verwandte texte*. Leiden: Brill.
- Toporov, Vladimir N. 1974. Slavianskije komentarii k neskolkim latinskim arkhajzjam. *Etimologija* 1972: 3–19. Moscow: Nauka.
- Trubachev, Oleg N. (ed.). 1974–. *Etimologičeskij slovar slavianskikh jazykov*. 32 vol. Moscow: Nauka.
- Vaillant, André. 1974. *Grammaire comparée des langues slaves*. Vol. 4. *La formation des noms*. Paris: Klincksieck.
- Vasmer, Max. 1953–1958. *Russisches etymologisches Wörterbuch*. 3 vols. Heidelberg: Winter.
- Waal, Willemijn. 2014. Hittite *gulš-/GUL-š-*, ^ᵛ*gulšeš/ᵛGUL-šeš*, Cuneiform Luwian *gulzā(i)-/GUL-zā(i)-*, Hieroglyphic Luwian *REL-za-* and the *Kuwanšeš*-deities. In P. Taracha and M. Kapelus (eds.). *Proceedings of the Eighth International Congress of Hittitology, Warsaw, September 2011*: 1016–1033. Warsaw: Agade.
- Yakubovich, Ilya. 2013. The Luwian Deity Kwanza. *Aramazd (Armenian Journal of Near Eastern Studies)* 8 (1–2): 282–297.

Ilya S. Yakubovich. The Slavic draughtsman: etymology of Slav. *čbrtv ‘devil’.

The paper addresses the origin of the Slavic word for ‘demon, devil’, which can be reconstructed as *čbrtv but does not have obvious cognates elsewhere in Indo-European. Using typological parallels from other Indo-European religious traditions, notably Anatolian and Armenian, I aim to demonstrate that Proto-Slavic *čbrtv can be interpreted as a supernatural

Draughtsman in charge of determining human fate. The original god of destiny probably came to be perceived as the bringer of death and then syncretised with the embodiment of all evil in the Christian tradition.

Keywords: etymology, devil, Slavic, Anatolian, Armenian.

The origin of Khanty retroflex nasal*

Proto-Khanty is traditionally reconstructed with a retroflex nasal phoneme *ŋ, whose origin remains disputed. According to one theory, it is directly inherited from Proto-Uralic. The other theory holds that Proto-Uralic *n, usually preserved as *n in Khanty, sporadically yielded *ŋ. We argue that Proto-Khanty *ŋ results from a regularly conditioned sound change.

Keywords: Ugric languages, historical phonology, retroflex consonants, Proto-Uralic.

1. Introduction

Khanty is unique among branches of the Uralic family in that it has a retroflex nasal phoneme /ŋ/ (often called “cacuminal” in Uralistic literature), distinct from both alveolar /n/ and palatal(ized) /ń/. The Proto-Khanty opposition of alveolar vs. retroflex vs. palatal(ized) nasals is reconstructed on the basis of the following correspondences:

PKh	V	Vj.	Vart.	Likr.	Mj.	Trj.	J	Irt.	Ni.	Š	Kaz.	Sy.	O
*n	n	n	n	n	n	n	n	n	n	n	n	n	n
*ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	n	n	n	n	ŋ	ŋ	n
*ń	ń	ń	ń	ń	ń	ń	ń	ń	ń	ń	ń	ń	ń

Proto-Khanty also had a similar triple opposition in the lateral series: alveolar *l, retroflex *ɭ and palatal(ized) *ł. Together with the retroflex affricate *č̣ (usually transcribed as *č̣), inherited from Proto-Uralic, *ŋ and *ɭ form part of a tightly integrated consonant system of Proto-Khanty:

	labial	alveolar	retroflex	palatal(ized)	velar
stops	p	t			k
affricates			č̣	č̣	
sibilant fricative		s			
lateral fricative		ɬ			
nasals	m	n	ŋ	ń	ŋ
lateral approximants		l	ɭ	ł	
trill		r			
glides	w			j	ɣ

Each of the five places of articulation could also be used to form homorganic clusters of nasals and stops (or affricates): *mp, *nt, *ŋč̣, *ńč̣, *ŋk (Honti 1999: 105). Unlike most other clus-

* I am grateful to Anna Dybo for comments that have helped to improve this paper. Any remaining mistakes are my own responsibility.

ters, these homorganic clusters could never be broken up by epenthetic schwa (Nikolaeva 2000). Proto-Khanty had a morphophonological rule whereby **ŋ* plus **t* yielded **ŋč* (Honti 1999: 98). Also, proto-Khanty **t* and **n*, on the one hand, and **č* and **ŋ*, on the other, could not co-occur within the same stem (Helimski 2002).

The existence of a retroflex nasal in Proto-Khanty and the absence of anything similar in other Uralic languages raises the issue of the origin of this phoneme. There is one group of cases where this origin is immediately clear: Proto-Uralic **nč* always yields Khanty **ŋč* (note that **č* is a retroflex affricate both in Khanty and Proto-Uralic), while Proto-Uralic **nt* is regularly preserved as Khanty **nt*. However, there are many instances of inherited words with Khanty **ŋ* in other positions as well, so the question of its origin remains unsolved.

From the early 20th century, there were two principal answers to this question in the Uralistic literature. The first was formulated already by Kustaa Fredrik Karjalainen, who had discovered the triple opposition of coronal nasals during his fieldwork on Khanty. According to Karjalainen, this opposition goes back to at least Proto-Finno-Ugric (Karjalainen 1913–1918: 6). He shows that when Khanty has **ŋ* and all other Finno-Ugric languages, including Mansi, have *n*, Hungarian has *ny*. Thus there are not two, but three series of regular correspondences:

Khanty	Hungarian	other Uralic
n	n	n
ŋ	ny	n
ń	ny	ń

Karjalainen also tried to find traces of original **ŋ* in other Uralic branches, notably Permian, but in order to do this, he needed to assume that in certain cases original **ŋ* yielded Khanty *n* for no apparent reason (Karjalainen 1913–1918: 28–30).

In his “Comparative Grammar of the Uralic Languages” (1960), Björn Collinder, following Karjalainen’s theory, reconstructed both **n* and **ŋ* (**ñ* in his notation) for Proto-Uralic. However, for some reason, he failed to mention that Hungarian has distinct reflexes of these consonants (Collinder 1960: 73, 133–134), thus losing the most powerful argument in favor of his reconstruction. Karjalainen’s theory was also adopted by V. M. Illich-Svitych, who projected the reconstruction of the retroflex nasal back to Proto-Nostratic (Illich-Svitych 1967: 323; 1971: 150), and by E. Helimski (1985: 75). All the abovementioned scholars were following the Neogrammarian paradigm, prohibiting unmotivated splits in historical phonology.

The alternative theory of the origin of the retroflex nasal in Khanty stems from scholars working outside the Neogrammarian paradigm. Formulated already by Erkki Itkonen (1957), it was most succinctly summarized by László Honti:

The cacuminals (**ŋ* **l*) are the result of a secondary development within pO: they originally occurred as allophonic variants when adjacent to the (non-distinctively) cacuminal affricate **č*, then spread, beginning with affective and descriptive vocabulary, to other positions, where they became phonemic. (Honti 1998: 337)

The secondary origin of Khanty **ŋ* was accepted by P. Sammallahti, although his wording is more cautious and does not necessarily suggest sporadic sound change:

An additional change in the consonantal system was caused by the ‘split-genesis’ of retroflex /l/ and /ŋ/. At least /ŋ/ was originally a contextual variant of /n/ before /č/, but for /lč/ no reliable etymologies can be found. (Sammallahti 1988: 512)

So far, all scholars either projected the Khanty retroflex nasal back to Proto-Uralic, or were willing to accept an unconditioned phonemic split in the prehistory of Proto-Khanty. As far as we know, nobody has tried to explain the origin of Khanty *ŋ in terms of a conditioned split of reflexes. Below, we will try to show that no sporadic sound changes need to be invoked in order to solve this problem.

2. Rules accounting for Proto-Khanty *ŋ

The following rules can be postulated for the development of Proto-Khanty *ŋ from Proto-Uralic *n.

Rule 1: PU *nč > PKh *ŋč.

- We will not list any examples for this rule here, since it is self-evident and uncontroversial.

Rule 2: PU *kVnV > PKh *kVŋ

- PU *kana- ‘to dig’ > PKh *k̄ŋ- > V Vj. VK Likr. Trj. k̄ŋ-, Mj. J k̄ŋ-, DN Kam. KoP Kr. Ts. χen-, Fil. χeń-, Kaz. Sy. χəŋ-, O χan- ‘to dig; to scoop’ (DEWOS: 508; UEW: 125, Sammallahti 1988: 545). Cf. PMs *kūn- > TJ kōn-, KU χūn-, P k̄ūn-, So. χūn- ‘to scoop’; Hung hány- ‘to throw’. The word is ultimately borrowed from Proto-Indo-Iranian *k^han- ‘to dig’.
- PU *kunV ‘belly’ > PKh *kuŋ > V Vj. VK Vart, Likr. Trj. kōŋ, Mj. J kōn, DN-Sal. Fil. KoP Kr. Sog. Ts. Ni. Š O χōn, Kaz. Sy. χōŋ ‘belly’ (DEWOS: 509–510; UEW: 208).
- PU *kaloni ‘on one’s back’ > PKh *kuŋ-čāγ > V Vj. kōŋčay, Trj. kōŋγi, J kōnyi, DN χōnča, KoP χōnčá, Kr. χānča, -á, Ni. Š χōnša, Kaz. Sy. χōŋša, O χōnsá ‘on one’s back’ (DEWOS: 514; UEW: 179). Cf. PMs *kān- > TJ TČ kanāw, KU χonī, P kōnəγ, So. χōni ‘on one’s back’; Hung hanyatt ‘on one’s back’. One might think that in this word a trivial development *nč > *ŋč has taken place, but in reality what we are dealing with here is the Proto-Khanty morphophonological rule (already mentioned above), according to which, *ŋ plus *t on a morpheme boundary yields *ŋč. The adverb *kuŋ-čāγ ‘on one’s back’ contains the adverbial suffix *-tāγ found also in the following words:
PU *kuma- ‘face down; to turn over’ > PKh *kom-tāγ > V Vj. kōmtay, Trj. J kāməγi, DN χōmta, DT χāmta, KoP χāmta, χōmta, Kr. χāmta, -á, Ni. Š Kaz. χōmta, O χāmta ‘prone, facedown’ (DEWOS: 502; UEW 201; Sammallahti 1988: 537);
PU *perä ‘behind’ > PKh *pir-tāγ > V Vj. pərtäγ, Trj. J pəryi, DN KoP Kr. -pərtá, Kaz. Sy. pərta, O pərtá ‘backwards’ (DEWOS: 1220; UEW 373; Sammallahti 1988: 553).
- PU *künä- ‘elbow’ > PKh *küŋč- > Ni. Š künš-ətəŋ, Kaz. küŋš-ələŋ, O kuns-aləŋ ‘elbow’; PKh *küŋč-ŋāj > V Vj. Likr. kōŋŋi, VK Vart. Mj. J kōŋŋi, Trj. kōŋŋi, DN Fil. Koš. kōŋŋaj, Tš. kōŋŋaj, KoP Kr. Ts. kōŋŋaj ‘elbow’ (DEWOS: 647; UEW: 158; Sammallahti 1988: 544). Cf. PMs *kün- > KU kōānyal, So. konl-əwl ‘elbow’; Hung kőnyök ‘elbow’. Although not everything in this etymology is clear (the Proto-Saamic reflex *kerñélē is especially hard to explain), at least the Mansi form shows the reflex of *-n-. Khanty *č may well result from the aforementioned morphophonological change *t > *č; but the identity and function of the suffix remain unknown.
- PUg *känäk ‘light’ (adj.) > PKh *kāŋəγ > Mj. kōŋəγ, J kōŋəγ, DN KoP Kr. Ts. Ni. Š kena, Kaz. keŋ, Sy. keŋ, O kon ‘light’ (adj.) (DEWOS: 648; UEW: 862). Cf. PMs *kinγā > KO kiyna, LU kinna ‘light’ (adj.); Hung könnyű ‘light’ (adj.).

- PKh **känəγ⁻¹* > Vj. *könəγ-*, *köñəγ-*, Trj. *könəγ-*, *köñəγ-*, J *könəγ-*, KoP Kr. *kənəj-*, Kaz. *keñi-*, Sy. *keñi-* ‘to growl (of bear, dog)’ (DEWOS: 648; UEW: 856). Cf. PMs **kīnəγ-* > KM *kenγ-*, So. *kēnγ-* ‘to growl (of bear)’. Comparison with Hung *könyörög-* ‘to pray’, accepted in UEW, seems improbable. The word can be onomatopoeic. Still, it shows the expected development of **n*.
- PU **kaŋa-ila* ‘armpit’ > PKh **kVŋəŋ* > V *kunəŋ-pətə*, VT *kunəŋ-pətə*, Vj. *kunəŋ-pətä*, Trj. *kunəŋ-pətə*, J *kunəŋ-pətə*, DN *χönəŋ-pət*, DT KoP Kr. *χänəŋ-pət*, Ni. *χunəŋ-pät*, Š *χunəŋ-ītpə*, Mul. *χunəŋ-päti*, Kaz. *χonəŋ-päti*, *χonəŋ-päti*, Sy. *χönəŋ-pät*, O *χönəm-pät* ‘armpit’ (DEWOS: 515–516; UEW: 178; Sammallahti 1988: 543). Cf. PMs **kanəl* > TJ TČ *kalnā*, KU *χānəl*, P *kanəl*, So. *χanəl* ‘armpit’; Hung *hón* (dial. *hóny*, *hany*, *hóln*) ‘armpit’. Here the Proto-Uralic form apparently had **-ŋ-*, but forms in daughter languages suffered various assimilations and metatheses. At least the Khanty word goes back to something like **kanVŋV-*. The Mansi word reflects **kanVIV*, while the Hungarian form goes back to **kalnV*.

Exceptions:

- PKh **känəŋ* > V Vj. *kanəŋ*, Trj. J *känəŋ*, DN KoP Kr. *χonəŋ*, Ni. Š Kaz. *χənəŋ* ‘bank (of a river); edge (of a forest, shawl etc.)’ (DEWOS: 514–515; UEW: 124–125). This word is compared in UEW to PMs **kant-* > KU *-χant*, KM *-kant* ‘near, close to’, Hung (dial.) *hany* ‘swamp’ and Komi-Permyak (dial.) *kan* ‘side’. These words are not especially close semantically or morphologically. The dubious nature of the comparison is recognized in UEW, where all parts of it are supplied with question marks. The most obvious connection of the Khanty word is with Proto-Selkup **kanək* ‘bank (of a river)’ (Alatalo 2004: 289). The Selkup word is thought to be a Khanty loan (UEW: 124–125), but the only reason for this is the supposed Finno-Ugric ancestry of the Khanty form. If the Finno-Ugric etymology is erroneous, the direction of the loan can be reversed. The Selkup word, in its turn, has been compared to similar words in Yeniseian languages: Kott *hanəŋ* ‘shore’ and Pumpokol *kónnoŋ* ‘mountain’ (Helimski 1982: 249). According to S. Starostin, “Pump. *kónnoŋ* (despite Helimski KC 249) should be distinguished from Kott. *hanəŋ* ‘shore’, which — as rightly pointed out by the author — is a Uralism (Selk. *qaniŋ*, Khant. *χonəŋ* ‘shore’)” (Starostin 2005). The Pumpokol form is compared by Starostin to Ket *qanñeŋ* — plural of *qañj* ‘mountain (wooded)’ < Proto-Yeniseian **qāñj*. It seems that the only reason to treat the Kott form as a Uralism is once again the supposed inherited nature of the Khanty word. If this premise is erroneous, we have two available options. One is to compare Pumpokol *kónnoŋ* ‘mountain’ with Kott *hanəŋ* ‘shore’ and abandon the idea of a connection between the former and the Ket plural. Then we can reconstruct Proto-Yeniseian **kanəŋ*² (consonant correspondences are regular, Starostin 1982: 148, 160, 162; as for vowels, we would expect Pumpokol ⟨a⟩, but it is not quite clear to what extent we can rely on transcription of vowels in 18th century sources). Another option is to accept Starostin’s etymology of the Pumpokol word and take the Kott word to be a loan from Pumpokol (Kott initial *h-* regularly goes back to **k-*, Starostin 1982: 160). In both cases, the Selkup word can be considered a Yeniseian loan that was further transferred to Khanty. Either scenario seems preferable to accepting the traditional Uralic etymology of the Khanty word.
- PKh **kjin-* > Trj. J *kän-*, DN DT KoP Kr. Ni. Kaz. Sy. O *χän-* ‘to stick (to), adhere (to) (intr.); to touch, move’; PKh **kan-t-* > V *kont-*, Vj. *kənt-*, Trj. *könt-*, Ni. *χunt-*, Kaz. *χənt-* ‘to stick

¹ Here and below we do not provide PU/PUG reconstructions for exclusively Ob-Ugric words.

² Reconstructions **kənəŋ*, **kanəŋ* and **kənəŋ* are also possible.

(to), to glue (tr.)’ (DEWOS: 504–505). Cf. PMs **kan-* > TČ *kan-*, KU *χân-*, P *kan-*, So. *χan-* ‘to touch; to hang (intr.); to stick (to) (intr.)’; PM **kan-t-* > KM *kânt-*, P *kant-* ‘to hang (tr.)’. This is a genuine exception, but the root is exclusively Ob-Ugric. If (as we argue below in section 4) the shift **n* > **ŋ* has taken place in Proto-Ugric times, a word that was borrowed from an unknown source into Proto-Ob-Ugric need not be subject to this sound law.

Rule 3: PU **...kVn(V)* > PKh **...γVŋ*

- PU **ikin* > PKh **äγəŋ* > V VK *äγəŋ*, Vart. Likr. Trj. *äγəŋ*, Mj. J *äγəŋ*, DN Fil. KoP Kr. *äŋəŋ*, Ni. Š *aŋəŋ*, Kaz. Sy. Pit. *aŋəŋ*, O *äŋəŋ* ‘chin; lower jaw’ (DEWOS: 43–44; Sammallahti 1988: 541). Cf. PMs **iγəŋ* > TČ *in*, KU *iγəŋ*, P *jēn*, So. *ēŋəŋ* ‘chin’; Hung *íny* ‘gum; palate’.
- PKh **kiγəŋ* > V Vj. Trj. *kəγəŋ*, VT J *kəγəŋ*, DN Kr. Ts. *kəŋəŋ*, Ni. Š Mul. *kīŋəŋ*, Kaz. *kīwəŋ*, Sy. *kījəŋ*, O *kijəŋ* ‘laces, strings (on clothes, shoes); button’ (DEWOS: 605–606). Cf. PMs **kīγəŋ* > TJ TČ *kīŋ*, KU *kiγəŋ*, P *kīn*, So. *kēŋəŋ* ‘button; string on clothes’.
- PKh **čoγəŋ* > V Vj. VK Vart. Likr. *fäγəŋ*, Mj. *fäγəŋ*, J *tāwŋ*, KoP *təŋəŋ*, Kaz. *šūŋəŋ* ‘fist; knuckles’ (DEWOS: 1503–1504). Cf. PMs **caknī* > TČ *čaχnē*, KU *śaχəŋ*, P *śaχəŋ*, So. *śaχni* ‘fist’. In the last two cases a PU reconstruction is not possible, since the words are limited to Ob-Ugric. Nevertheless, the environment is the same as in the first case.

Rule 4: PU **nVkkV* (POU **nVkkV*) > PKh **ŋVk*

- PU **nikkā-* ‘to stick in, push’ > PKh **nik-* > VK *nək-*, Likr. *ŋək-*, Trj. J *nək-* ‘to push (smb) lightly’; PKh **ŋikəj-* > Kaz. *ŋāki-* ‘to push, nudge’ (DEWOS: 984; UEW: 304–305). Cf. PMs **nāk-* > KM *nāk-* ‘to push’.
- PKh **ŋēkī* > V *ŋiγ*, VT *niγ*, Vj. *ŋiki*, Trj. J *niki*, DN Kam. KoP Kr. Ts. *nekə*, Ni. *nāk*, Kaz. *ŋeki* ‘trigger (of a trap), peg (for strings in a musical instrument), etc.’ (DEWOS: 986). Cf. PMs **nāŋk* > KM *nōχ*, P *naχ* (pl. *naŋkət*) ‘pintle, pivot’. If the P plural form *naŋkət* was created by analogy with nouns that have the alternation *-χ* / *-ŋk-*, the PMs form can be reconstructed as **nāk*, which would agree better with PKh **ŋēk*.
- PKh **ŋōk-* > V *ŋoγ-*, Vj. *ŋoγ-* (*ŋōka*), VK Vart. *ŋoγ-*, Likr. *naγə-*, *ŋakə-*, Mj. *ŋok*, *ŋokə-*, Trj. *ŋok-*, J *nok-*, DN Fil. KoP Kr. Tš. *noχ-*, Ni. Š *nuχ-*, Kaz. *ŋoχ-*, Sy. *ŋuχ-*, O *noχ-* ‘to peck’ (DEWOS: 987–988). Cf. PMs **ŋōk-* > TJ TČ *nāk-*, KU *nāχ-*, P *nēkə-*, *nēk-*, So. *nāχə-* ‘to peck’.
- PKh **ŋāk-* > Ni. *noχ-*, Kaz. Sy. *ŋoχ-*, O *naχ-* ‘to limp’ (DEWOS: 988). Cf. PMs **nāk-* > KU *nōχ-*, P *nōk-*, So. *nōχ-* ‘to limp’.
- PKh **ŋik* > Ni. *nāk*, Kaz. *ŋāk* ‘joint’ (Ni.), ‘spell (of weather)’ (Kaz.) (DEWOS: 983–984). Cf. PMs **nāk* > TJ TČ KU P *nāχ*, So. *nak* ‘joint’. Judging by its distribution in Khanty, the word can be a Mansi loan; but the correspondences are still regular and we list it here for the sake of completeness.

Rule 5: POU **...nVγ* > PKh **...ŋVγ*

Here, in the only example that has a Uralic etymology, final **-γ* is a suffix, added in (Ob-)Ugric times. So it makes little sense to formulate the input of the rule in terms of Proto-Uralic reconstruction.

- PU **šēni* ‘tinder, bracket fungus’ > PKh **sāŋəγ* > V *-sānəγ*, V Vj. VK *-sāŋəγ*, Trj. *sāŋəγ*, J *sānəγ*, DN KoP Koš. Kr. Sog. Ts. *sānə*, KoP *sānək*, Fil. *sānə*, Ni. *sanə*, Kaz. Sy. *saŋ*, O *sān* ‘bracket fungus’ (DEWOS: 1345; UEW: 494–495; Sammallahti 1988: 548). Cf. PMs **šīnəγ* > TJ TČ *šīnəw*, KU *šēnī*, P *šēniγ*, So. *sēniγ* ‘bracket fungus’.

- PKh **āṅəγ* > V Vj. *āṅəγ*, Likr. Mj. Trj. *āṅəγ*, J *āṅəγ*, DN KoP Kr. *āṅə*, Ni. Š *anə*, Kaz. Sy. Pit. *aṅ*, O *ān* ‘cup; plate’ (DEWOS: 114–115). Cf. PMs **ānī* > TJ TČ *ānī*, KU *ānə*, P *āna*, So. *āni* ‘cup, plate, vessel’.
- PKh **sāṅəγ* > V Vj. VK *seṅəγ*, Likr. Mj. Trj. *sāṅəγ*, J *sāṅəγ*, DN KoP Kr. Ts. Ni. *senə*, Kaz. *seṅ*, Sy. *seṅ*, O *sen* ‘nit; crab louse’ (DEWOS: 1345–1346). Cf. PMs **šāṅəγ* > TJ TČ *šāṅəw*, KU *šānī*, P *šāniγ*, So. *sāniγ* ‘nit’.

Rule 6: PU **mVnV* > PKh **mVn*

- PU **muna* ‘egg; testicle’ > PKh **maṅ* > V Vj. *mṅṅ*, Trj. *mōn*, J *mōn*, DN Fil. KoP Kr. *mun*, DT *mān*, Ni. *mun*, Kaz. *mṅṅ*, O *mon* ‘penis; testicle’ (DEWOS: 935; UEW 285–286; Sammallahti 1988: 538). Cf. PMs **man* > TJ *man*, KU *mān*, P *mon*, LO *mon* ‘testicle’; Hung (obs., dial.) *mony* ‘egg; testicle; penis’.
- PUg **minV-* ‘to tear; to dislocate’ > PKh **māṅəṃ-* > Ni. Š O *menəṃ-*, Kaz. *meṅəṃ-* ‘to tear off’ (DEWOS: 935–936; UEW: 870–871). Cf. PMs **māṅəṃt-* > TJ *māṅəṃt-*, KU *māṅəṃt-*, P *māṅəṃt-*, So. *manəṃt-* ‘to tear’; Hung (dial.) *ki-mēnyül-* ‘to be dislocated (of joint)’, *ki-mēnyít-* ‘to dislocate (of joint)’.

Exception:

- PU **meni-* ‘to go’ > PKh **min-* > V Vj. VK Sur. Irt. *mən-*, Ni. Š Kaz. Sy. *mān-*, O *mān-* ‘to go’ (DEWOS: 931–932; UEW: 272; Sammallahti 1988: 538). Cf. PMs **min-* > TJ *miń-*, KU P So. *min-* ‘to go’; Hung *mən-* ‘to go’. See section 3 below on the possible cause of this exception.

Forms with unexpected PKh **ŋ*³

- PU **enā* > PKh **āṅā* > V Vj. VK *eṅə*, Vart. Likr. Trj. *āṅə*, Mj. *āṅə*, *āṅə*, J *āṅə*, DN KoP Kr. *enə* ‘thick; big’ (DEWOS: 109–110; UEW: 74–75; Sammallahti 1988: 541). Cf. PMs **jāṅəγ* > TJ *jinəw*, KU *jāni*, P *jāniγ*, So. *janiγ* ‘big’. Cf., however, another derivative from the same root: PKh **āṅəṃ-* > V Vj. VK *enəṃ-*, Likr. Mj. Trj. J *āṅəṃ-*, Irt. Ni. Š *enəṃ-*, Kaz. *enəṃ-*, Sy. *enəṃ-*, (rarely) *eṅəṃ-* ‘to grow’. It is possible that the Mansi word with *-əγ* directly reflects the Proto-Ob-Ugric form, while in Khanty *-əγ* was secondarily replaced with another suffix. If so, both the presence of **ŋ* in the adjective and its absence in the verb can be explained by Rule 5.
- PUg **pəṅə* ‘to fart’ > PKh **pṅṅ* > DN *pān*, Kr. *pān-*, Ni. *pōn*, Kaz. *pōṅ*, O *pān* ‘a fart’; PKh **paṅəγ-* > V Vj. *poṅəγ-*, Trj. *pōṅəγ-*, J *pōṅəγ-*, DN KoP Kr. *pānə-*, Ni. *punij-*, Kaz. *poṅi-* ‘to fart’ (DEWOS: 1169; UEW: 413). Cf. PMs **pəṅəγ* > TJ *poṅχ*, KM *pānəγ*, P *ponəγ*, LO *ponəγ* ‘a fart’; Hung *fiṅ* ‘to fart’. The verb has a regular **ŋ* (Rule 5), whereas the noun apparently acquired the retroflex nasal under the analogical influence of the verb.
- PU **niwa-* ‘to remove hair from skin’ > PKh **ṅaw-* > Kaz. Sy. *ṅəw-*, O *naw-* ‘to remove hair from reindeer hide’ (DEWOS: 1024; UEW: 306; Sammallahti 1988: 546). A genuine exception.

3. The problem of secondary *ny* in Hungarian

Hungarian parallels to Khanty words listed above demonstrate the validity of the correspondence between Khanty *ŋ* and Hungarian *ny*, discovered already by Karjalainen (1913–1918:

³ We list here only those words that have parallels in Mansi or other Uralic languages. Actually, many other Khanty words without external parallels also follow the rules formulated above.

24–25). Below we will separately list all reliable etymologies with Hungarian *ny* going back to PU **n*. These etymologies are not especially numerous, and it would be hard to formulate the conditions of the change, were it not for the fact that the same conditions are valid for the Khanty change **n > *ŋ*. Rules 2, 3, 4 and 6, formulated above for Khanty, work for Hungarian as well. Rule 1 (PU **nč > PKh *ŋč*) has no counterpart in Hungarian, since Hungarian has simplified all clusters of nasals and homorganic obstruents. The Hungarian reflex of PU **nč* is *r* (Aikio in press).

We do not have relevant examples for a possible Hungarian counterpart of Rule 5 (POU **...nVɣ > PKh *...ŋVɣ*). The remaining rules have their counterparts in Hungarian.

Rule 2: PU **kVnV > PKh *kVŋ*, Hung *h/kVny*.

- PU **kana-* ‘to dig’ > Hung *hány-* ‘to throw’ (UEW: 125, Sammallahti 1988: 545). Cf. PKh **k̄ŋ-* ‘to dig; to scoop’, PMs **k̄ŋ-* ‘to scoop’.
- PU **ka/oni* ‘on one’s back’ > Hung *hanyatt* ‘id.’ (UEW: 179). Cf. PKh **kuŋ-čāɣ* ‘id.’, PMs **k̄ān-* ‘id.’.
- PU **künä-* ‘elbow’ > Hung *könyök* ‘id.’ (UEW: 158; Sammallahti 1988: 544). Cf. PKh **küŋč-* ‘id.’, PMs **k̄ŋ-* ‘id.’.
- PUg **känöɣ* ‘light’ (adj.) > Hung *könnyű* ‘id.’ (UEW: 862). Cf. PKh **käŋəɣ* ‘id.’, PMs **kinɣā* ‘id.’.

Rule 3: PU **...kVn(V) > PKh *...ɣVŋ*, Hung *...ny*. We have only one example for this rule in Hungarian, so it could hardly be formulated without the Khanty parallel.

- PU **ikin* ‘gums’ > Hung *íny* ‘gum; palate’ (UEW: 80–81; Sammallahti 1988: 541). Cf. PKh **āɣəŋ* ‘chin; lower jaw’, PMs **īɣəŋ* ‘chin’.

Rule 4: PU **nVkkV > PKh *ŋVk*, Hung *nyVk*. Here the Hungarian part of the rule can also be illustrated by one example only.

- PU **nokki* ‘nape of the neck’ > Hung *nyak* ‘neck’, PSelk **nuku* ‘nape of the neck’ (UEW: 328–329; Alatalo 2004, #1385). This etymology, rejected by Janhunen (1981) and Sammallahti (1988), can be rehabilitated if we compare it to two other cases with the same vowel correspondence:

PU **soski-* ‘to chew’ > PSelk **tutu-* ‘id.’ (UEW: 448–449; Alatalo 2004, #1068);

PU **totki* ‘tench’ > Hung *tat hal* ‘id.’, PSelk **tutu* ‘crucian carp’ (UEW: 532; Alatalo 2004, #1066).

While the details of development, especially the origin of PSelk **u* in the second syllable, are not clear, the regularity of the correspondence is not in doubt.

Rule 6: PU **mVnV > PKh *mVŋ*, Hung *mVny*.

- PU **muna* ‘egg; testicle’ > Hung (obs., dial.) *mony* ‘egg; testicle; penis’ (UEW 285–286; Sammallahti 1988: 538). Cf. PKh **maŋ* ‘penis; testicle’, PMs **man* ‘testicle’.
- PUg **minV-* ‘to tear; to dislocate’ > Hung (dial.) *ki-ményül-* ‘to be dislocated (of joint)’, *ki-ményít-* ‘to dislocate (of joint)’ (UEW: 870–871). Cf. PKh **mäŋəm-* ‘to tear off’, PMs **mänəmt-* ‘to tear’.

Exception:

- PU **meni-* ‘to go’ > Hung *mén-* ‘id.’ (UEW: 272; Sammallahti 1988: 538). Cf. PKh **min-* ‘id.’, PMs **min-* ‘id.’. The exceptional behaviour of this verb may have something to do with its morphophonological peculiarity in Ugric. Helimski (1990: 64–66) has shown that Hungarian reflexes of Proto-Uralic verbs **meni-* ‘to go’ and **woli-* ‘to be’ build their present stems

with the same suffix **-ś-* that forms present stems of Hungarian monosyllabic verbs of the shape CV-: *mëgy-* ‘to go’ < **men-ś-*, *vagy-* ‘to be’ < **wol-ś-*, *ësz-* ‘to eat’ < **se-ś-* < **sewi-ś-*, *vësz-* ‘to take’ < **we-ś-* < **weyi-ś-* etc. We can suppose that these two verbs became (or remained) monosyllabic already in common Ugric times. This can explain the exceptional preservation of **n*: the final consonant of a monosyllabic stem was protected from change by the immediately following initial consonants of affixes, after which the unchanged **n* was generalized to prevocalic position. It is especially important for the chronology of the change in question that this exception is common to Hungarian and Khanty.

One more rule can be tentatively formulated for Hungarian only (Khanty has **n* in the words in question):

Rule 7: pre-Hungarian **nVl/r* > Hung *nyVl/r*.

- PU **nüüdi* ‘handle’ > Hung *nyél* (acc. *nyelet*) ‘id.’. Cf. PKh **nül* ‘id.’, PMs **näl* ‘id.’ (DEWOS: 997–998; UEW: 304; Sammallahti 1988: 538).
- PUG **närkz* ‘saddle’ > Hung *nyerëg* (acc. *nyerget*) ‘id.’. Cf. PKh **nãγar* ‘id.’, PMs **nãγrã* ‘id.’ (DEWOS: 996; UEW: 874). Judging by irregular vowel correspondences, the word for ‘saddle’, together with other horse-related terms, was borrowed separately by different Ugric languages in Common Ugric times from an unknown source. However, this does not explain the discrepancy in initial consonants. Since the position before *-r-* in this word is reminiscent of the position before *-l-* in *nyél*, and these are the only reliable etymologies where secondary *ny* in Hungarian corresponds to Khanty **n*, we prefer to postulate a separate rule for these cases.

The rule can be confirmed by one more etymology:

- PU **nara* or **nora* ‘spring’ > Hung *nyár* (acc. *nyarat*) ‘summer’, PS **nãrã* ‘snow crust; spring (season)’ (see Janhunen 1977: 98 for Samoyed data; the comparison of Hungarian and Samoyed words was suggested by A. Dybo 2007: 170). Two etymologies of the Hungarian word are discussed in EWUng: 1) an inner-Hungarian semantic development *nyár* ‘swamp’ > *nyár* ‘summer’ and 2) borrowing from Proto-Turkic **jãr²* ‘spring, summer’ (more precisely, from pre-Proto-Turkic **ñãr²*). The first etymology involves an improbable semantic shift: parallel cases adduced in EWUng demonstrate developments like ‘summer’ > ‘melted water’, ‘a place free from snow’, not vice versa. Borrowing from pre-Proto-Turkic is also highly dubious, since no certain traces of such a layer of loanwords are known (see additional arguments against this etymology in Dybo 2007: 169–170). In the earliest layer of Hungarian loanwords from Turkic languages, Turkic **j-* is rendered by Hungarian *gy-* (Dybo 2007: 29–30).

4. Conclusion

We can see that Hungarian shares with Khanty Rules 2, 3, 4 and 6 together with a common exception from Rule 6 — the reflex of PU **meni-*. Moreover, despite the fact that most of the rules formulated above involve presence of a velar consonant, PU **nVkV* yields **nV(γ/w)* both in Khanty and Hungarian: PU **näki-* ‘to see’ > PKh **nũ(w)-*, Hung *nëz-*.

Of the three possible explanations — completely independent parallel development, diffusion of rules between closely related languages and common inheritance from an intermediate parent language — the first one can be definitely excluded. The rules formulated above are

typologically uncommon, and the existence of a common exception betrays a historical connection between Khanty and Hungarian developments. “Diffusionist” account remains a possibility, but were it true, we would expect more differences between Hungarian and Khanty rules (cf. subtle differences between effects of *RUKI*-rule in Indo-European languages). The remaining possibility involves a common intermediate node on the Uralic tree. As far as we know, no one has ever suggested a Uralic subgroup that would include Khanty and Hungarian, but exclude Mansi. The development of nasals is hardly sufficient for the postulation of such a subgroup. The only reasonable version of a “genetic” account would involve acceptance of the Ugric node. Then we could formulate Rules 2, 3, 4 and 6 as having taken place between the Proto-Uralic and Proto-Ugric stages. Under this account, Proto-Ugric *ŋ that resulted from these rules became (marginally) phonemic as a result of analogical leveling (PKh **min-*, Hung *mën-* ‘to go’) and, perhaps, common borrowings from unknown languages (PKh **kjn-*, PMs **kan-* ‘to stick’). Later, Proto-Ugric *ŋ was preserved in Proto-Khanty, merged with *n in Proto-Mansi (cf. the fate of *ŋ in a number of Khanty varieties) and yielded *ny* in Hungarian.

Abbreviations for languages and dialects⁴

Hung — Hungarian	PMs — Proto-Mansi	PS — Proto-Samoyed	PU — Proto-Uralic
PKh — Proto-Khanty	POU — Proto-Ob-Ugric	PSelk — Proto-Selkup	PUG — Proto-Ugric

References

- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). In press. Notes on the development of some consonant clusters in Hungarian. In: Sampsa Holopainen, Janne Saarikivi (eds.). *Peri orthotētos etymōn. Uusiutuva uralilainen etymologia* (Uralica Helsingiensia).
- Alatalo, Jarmo 2004. *Sölkupisches Wörterbuch aus Aufzeichnungen von Kai Donner, U. T. Sirelius und Jarmo Alatalo*. Zusammengestellt und herausgegeben von Jarmo Alatalo (Lexica Societatis Fenno-Ugricae XXX). Helsinki: Finnisch-Ugrische Gesellschaft.
- Collinder, Björn. 1960. *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.
- DEWOS = Steinitz, Wolfgang. 1966–1993. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache*. Lfg. 1–15. Berlin: Akademie Verlag.
- Dybo, A. V. 2007. *Lingvisticheskie kontakty rannikh tyurkov: leksicheskiy fond: pratyurkskiy period*. Moskva: Vostochnaya literatura.
- EWUng = Benkő, Loránd (ed.). 1992–1997. *Etymologisches Wörterbuch des Ungarischen*. Budapest: Akadémiai Kiadó.
- Helimski, Eugen. 1982. Keto-Uralica. In *Ketskiy sbornik. Antropologiya, etnografiya, mifologiya, lingvistika*: 238–251. Leningrad: Nauka.
- Helimski, Eugen. 1985. Rez.: L. Honti, Geschichte des obugrischen Vokalismus der ersten Silbe. *Sovetskoe finno-ugrovedenie* 21(1): 63–76.
- Helimski, Eugen. 1990. Ugrica: etimologii s istoriko-foneticheskim podtekstom. *Nyelvtudományi Közlemények* 91: 63–68.
- Helimski, Eugen. 2002. Eine Regel der Konsonantenkompatibilität im Ostjakischen. In: E. Helimski und A. Widmer (Hg.). *Wüša wüša — Sei gegrüßt! : Beiträge zur Finnougristik zu Ehren von Gert Sauer dargebracht zu seinem siebzigsten Geburtstag* (Veröffentlichungen der Societas Uralo-Altaica 57). Wiesbaden: Harrassowitz: 111–116.
- Honti, László. 1998. ObUgrian. In: D. Abondolo (ed.). *The Uralic Languages*. London / New York: Routledge.
- Honti, László. 1999. *Az Obi-Ugor konzonantizmus története* (Studia uralo-altaica: Supplementum 9). Szeged: Universitas Szegediensis de Attila József Nominata.

⁴ We use standard abbreviations for Khanty and Mansi varieties, see DEWOS for Khanty and Steinitz 1955 for Mansi.

- Illich-Svitych, V. M. 1967. Materialy k sravnitel'nomu slovaryu nostraticheskikh yazykov (indoeuropeyskiy, altayskiy, ural'skiy, dravidskiy, kartvel'skiy, semitokhamitskiy). In *Etimologiya 1965 (Materialy i issledovaniya po indoeuropeyskim i drugim yazykam)*: 321–373. Moskva: Nauka.
- Illich-Svitych, V. M. 1971. *Opyt sravneniya nostraticheskikh yazykov (semitokhamitskiy, kartvel'skiy, indoeuropeyskiy, ural'skiy, dravidiyanskiy, altayskiy). Vvedenie. Sravnitel'nyy slovar' (b-Қ)*. Pod redakciey i s vstupitel'noy stat'ey V. A. Dybo. Moskva: Nauka.
- Itkonen, Erkki. 1957. Suomalais-ugrilaisen kantakielen äänne- ja muotorakenteesta. *Virittäjä* 61/1: 1–22.
- Janhunen, Juha. 1977. *Samojedischer Wortschatz. Gemeinsamojedische Etymologien*. (Castrenianumin toimitteita 17.) Helsinki.
- Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Suomalais-Ugrilaisen Seuran Aikakauskirja / Journal de la Société Finno-Ougrienne* 77: 219–274.
- Karjalainen, K. F. 1913–1918. Beiträge zur geschichte der finnisch-ugrischen dentalen nasale. *Suomalais-Ugrilaisen Seuran Aikakauskirja / Journal de la Société Finno-Ougrienne* 30(24).
- Nikolaeva, Irina. 2000. The status of schwa in the phonological description of Northern Ostyak. *Finnisch-ugrische Mitteilungen* 23: 121–148.
- Sammallahti, Pekka. 1988. Historical Phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In: D. Sinor (ed.). *The Uralic Languages. Description, history and foreign influences*. Leiden / New York: Brill: 478–554.
- Starostin, Sergei. 1982. Praeniseyskaya rekonstrukciya i vneshnie svyazi eniseyskikh yazykov. In *Ketskiy sbornik. Antropologiya, etnografiya, mifologiya, lingvistika*: 144–237. Leningrad: Nauka.
- Starostin, Sergei. 2005. *Yenisseian etymology*. Available: <http://starling.rinet.ru/> [Accessed 27.11.2016].
- Steinitz, Wolfgang. 1955. *Geschichte des wogulischen Vokalismus* (Finnisch-ugrische Studien II). Berlin: Akademie Verlag.
- UEW = Rédei, Károly. 1988–1991. *Uralisches Etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.

М. А. Живлов. Происхождение хантыйского ретрофлексного носового.

Для прахантыйского традиционно восстанавливается ретрофлексный носовой согласный *ŋ, происхождение которого остаётся спорным. Согласно одной теории, он непосредственно унаследован из прауральского. По другой теории, прауральская фонема *n, обычно сохраняющаяся в хантыйском как *n, спорадически давала *ŋ. Мы приводим доводы в пользу того, что прахантыйский ретрофлексный носовой *ŋ является результатом регулярных позиционно обусловленных фонетических изменений.

Ключевые слова: угорские языки, историческая фонология, ретрофлексные согласные, прауральский язык.