

А. И. Коган / Anton Kogan

(Институт востоковедения РАН, Москва)

Ответ на статью А. С. Крыловой

«Лексикостатистика новоиндоарийских языков: взгляд полевого лингвиста»

В своей недавней статье (Крылова 2017) А. С. Крылова предложила ряд поправок и дополнений к индоарийской лексикостатистической базе данных, приведенной в нашей работе (Коган 2016). Кроме того, ей были высказаны некоторые замечания, касающиеся самих принципов отбора лексики, включаемой в стословные списки. Все эти поправки, дополнения и замечания, безусловно, интересны, но в значительной своей части небесспорны, а потому требуют детального обсуждения. Прежде, однако, хотелось бы поблагодарить А. С. Крылову за пополнение базы материалами, собранными ей в ходе полевой работы. Речь идет, в первую очередь, о материалах по языкам куллуи и ория. Учет этих новых данных, вне всякого сомнения, позволит уточнить родословное древо и сделает его лучше отражающим реальную картину генетических отношений новоиндийских языков.

Разбор замечаний А. С. Крыловой представляется целесообразным производить, следуя порядку и рубрикации, предложенным в ее статье.

### Замечания к семантике

1. Индоарийские местоимения, выступающие в стословном значении ‘all’ и продолжающие др.-инд. *sarva*- ‘все’ (хинди-урду *sab*, пандж. *sabh*, неп. *sabai* и т. д.), объявляются этимологически не связанными с местоимениями, восходящими к прототипу *sāra*- ‘весь, целый’ (потх., хиндко *sārē*, лахнда *sārā* и т. д.<sup>1</sup>). Утверждается, что два данных этимона не связаны на праиндоевропейском уровне (Крылова 2017: 282). Между тем, связь между ними представляется вероятной, на что в свое время указывал еще Жюль Блок (Bloch 1919: 420). Оба они, скорее всего, восходят к производным и.-е. *\*solo-*, *\*sol(e)u-* ‘целый’<sup>2</sup>. Таким об-

разом, присвоение отражениям др.-инд. *sarva*- и *sāra*-<sup>3</sup> одного номера в базе данных представляется нам вполне оправданным. Включение же в стословные списки обоих слов в качестве синонимов, как это предлагает А. С. Крылова, учитывая сказанное выше, нецелесообразно.

5. Утверждается, что в стословном списке хинди-урду в значении ‘neck’ дано название горла, а не шеи (Крылова 2017: 283). Однако у хинди-урду *galā*, включенного нами в базу данных, имеется как значение ‘горло’, так и значение ‘шея’ (Бархударов и др. 1972: 452)<sup>4</sup>. Поскольку данное слово является исконным, предпочтение при составлении списка было отдано именно ему, а не персидскому заимствованию *gardan*.

### Замечания к этимологиям

1. А. С. Крылова отрицает возможность исконного происхождения названия красного цвета в большинстве новоиндийских языков (хинди-урду, пандж., лахнда, гудж. *lāl*, синдхи *lālu*, бенг., асс. *lal* и т. д.<sup>5</sup>), предлагая выводить его из перс. *la'l* ‘рубин’ (Крылова 2017: 283–284). Следует сказать, что связь индоарийских слов с персидским постулируется и нами (Коган 2005). Однако нам представляется более вероятным не предполагаемое А. С. Крыловой заимствование с последующим появлением у существительного ‘рубин’ дополнительной семантики ‘красный’, а контактирование данного существительного с исконным индоарийским прилагательным. Это последнее мы считаем отражением вероятного прототипа *\*lōhila-*, а влиянием со стороны семантически близкого персидского слова объясняем появле-

<sup>1</sup> Полный перечень отражений этих двух основ в привлеченных нами для рассмотрения новоиндийских языках см. в статье (Коган 2016: 239).

<sup>2</sup> Отражения данной основы в других индоевропейских языках см. в: Pokorny 1959: 979–980.

<sup>3</sup> Не исключено, что данная основа, вопреки предположению Р. Тернера (Turner 1966: 770), не является этимологически тождественной др.-инд. *sāra*- ‘сердцевина; суть, сущность’.

<sup>4</sup> Ср., например, такие контексты употребления данного слова, как *galē mē hāth dālnā* ‘обнять за шею’, *galē milnā* ‘обниматься’.

<sup>5</sup> Полный список новоиндийских форм см. в Kogan 2016: 250.

ние в новоиндийских формах долгого *ā* на месте ожидаемого *ō*<sup>6</sup>. А. С. Крылова ставит под сомнение саму возможность реконструкции данного прототипа, а между тем, предполагавшаяся Р. Тернером (Turner 1966: 650) праформа *\*lōhila-*, вероятнее всего, представляет собой региональное отражение др.-инд. *rudhira-* ‘красный’. Утверждение А. С. Крыловой о том, что в средневековых новоиндийских диалектах существительное ‘рубин’ затруднительно отличить от прилагательного ‘рубиновый, красный’ ввиду несформированности системы послелогов, вызывает немало сомнений и при этом едва ли может считаться весомым аргументом. Примеры из языка Кабира, приведенные в (Крылова 2017: 283), демонстрируют отсутствие формальных средств маркировки генитивных определений, однако мы не знаем несомненных примеров функционирования последних в качестве цветообозначений в индоарийских языках. Напротив, известные нам названия цветов, этимологически связанные с существительными, характеризуются морфологическим оформлением, отличным от такового у генитивных определений<sup>7</sup>. Кроме того, если принять точку зрения А. С. Крыловой, остается неясным, почему прилагательное ‘красный’ и существительное ‘рубин’ различаются в современном индоарийском, где на синхронном уровне они, несомненно, представляют собой два разных слова, что в некоторых языках фиксируется в орфографии (ср., например, разное написание этих слов в урду: لال ‘красный’ и لؤلؤ ‘рубин’). Весьма сомнительным представляется и тезис о том, что «основные исконные цветообозначения хинди, как и большинство исконных базовых прилагательных, являются изменяемыми и заканчиваются в исходной форме

на *ā*» (Крылова 2017: 284). В лучшем случае это гипотеза, требующая проверки. Осуществление такой проверки, по всей видимости, предполагает расписку больших словарей, а простое перечисление прилагательных из стословного списка, которым ограничивается А. С. Крылова, не может служить надежным обоснованием. Прилагательные хинди-урду *gōl* ‘круглый’ и *bhāri* ‘тяжелый’ представляют собой показательные контрпримеры, причем, вопреки утверждению А. С. Крыловой, их фонетический облик отнюдь не дает оснований «трактовать их как санскритские заимствования»<sup>8</sup>. Отсутствие форм типа *\*lāl* в новоиндийских языках, где произошел переход *l > !* (Крылова 2017: 284), может иметь синхронно-фонетическое объяснение: подобные формы попросту трудны для артикуляции.

5. Новоиндийские названия семени, этимологически связанные с др.-инд. *bija-* и не утратившие интервокальную аффрикату *j* (хинди-урду *bīj*, синдхи *b'iju*, каччи *bijj*, куллуи *bejja*, бенг. *bij*, асс. *biz* и др.<sup>9</sup>), должны, по мнению А. С. Крыловой, быть признаны санскритизмами (Крылова 2017: 285). Данное утверждение, однако, по всей видимости неверно. Подобные формы могут рассматриваться как рефлексы засвидетельствованной в древнеиндийском основы *bijya-*, на что указывал еще Р. Тернер (Turner 1966: 523). Конечная гемината в каччи, а также краткий гласный в каччи и синдхи однозначно указывают на более ранний (среднеиндийский) прототип *\*bijja-*, вполне закономерно продолжающий др.-инд. *bījya-*.
6. Утверждается, что продолжения др.-инд. *sūrya-/sūriya-* ‘солнце’ «утратили в первом случае *r*, во втором *y*» (Крылова 2017: 285). Если первая часть данного тезиса несомненна, то вторая представ-

<sup>6</sup> В нашей монографии «Дардские языки: генетическая характеристика» данная гипотеза формулируется следующим образом: «Наличие нерегулярного *ā* в современных индоарийских формах, возможно, объясняется аналогией с бытующим во многих языках севера Индостана заимствованным из персидского существительным *lāl* ‘рубин’» (Коган 2005: 157). Неясно, почему описанный сценарий историко-фонетического развития совершенно не упоминается А. С. Крыловой, указавшей в числе аргументов против нашей этимологии на отсутствие долгого *ā* в отражениях праформы *\*lōhila-*, приведенных в словаре Р. Тернера.

<sup>7</sup> Ср., например, некоторые цветообозначения современного хинди-урду: *baīganī* ‘фиолетовый’ от *baīgan* ‘баклажан’ при генитивном образовании от последнего слова *baīgan kā; sunahrā, sunahlā* ‘золотистый’, несомненно связанное с *ṣōṇā* ‘золото’, но вовсе не идентичное послеложному сочетанию *ṣōṇē kā*, функционирующему как генитивное определение.

<sup>8</sup> Относительно этих двух слов и их трактовки в статье А. С. Крыловой необходимо высказать несколько дополнительных замечаний. Существительное хинди-урду *gōl* ‘круг, шар’ является фонетически регулярным отражением др.-инд. *gōla-* ‘шар’. Адъективное значение ‘круглый’ у этого слова зафиксировано уже в раннесреднеиндийскую эпоху (ср. пали *gōla-* ‘круглый; шар’), поэтому в данном случае едва ли можно вслед за автором статьи предполагать процесс, параллельный образованию *lāl* ‘красный’ от *lāl* ‘рубин’. Прилагательное же *bhāri* ‘тяжелый’ в принципе не может являться санскритизмом: в древнеиндийском в соответствующем значении используется основа *guru-*, а слово, фонетически и семантически тождественное хинди-урду *bhāri* попросту не засвидетельствовано. Традиционно прототипом для этого прилагательного считается др.-инд. *bhārika-* ‘forming a load, heavy’ (Turner 1966: 539).

<sup>9</sup> Полный список форм см. в Kogan 2016: 252.

ляется спорной. В упомянутой А. С. Крыловой работе К. Масики отмечена возможность перехода интервокального *-y-* > *-jj-* в среднеиндийских диалектах (Masica 1991: 169). В новоиндийскую эпоху гемината регулярным образом упрощалась и закономерное отражение др.-инд. *sūriya-* должно было принять вид *sūrVj*, что и обнаруживается в реально засвидетельствованных новых индоарийских формах (ср. дакхини *sūrīj*, хинди-урду, пандж., гудж., кум. *sūraj* и т. д.<sup>10</sup>). Объявлять подобные формы «частично адаптированными санскритизмами», как это делает А. С. Крылова, мы не видим веских оснований.

### Поправки к списку хинди

4. Вопреки утверждению А. С. Крыловой (Крылова 2017: 286), хинди *mās* ‘мясо’ едва ли может рассматриваться как основной синоним, который следует включать в стословный список. Слово это относится к книжному стилю, характерно только для литературного хинди и неупотребительно в урду. Его незаимствованное происхождение представляется маловероятным. По всей видимости, исконным отражением др.-инд. *māṃsa-* ‘мясо’ в хинди-урду является устаревшее и практически вышедшее ныне из употребления *mās*<sup>11</sup>. Включенное же в нашу лексикостатистическую базу персидское заимствование *gōšt* стилистически нейтрально, и необходимости заменять его каким-либо другим словом мы по этой причине не видим.
5. Существительное *mēh* ‘дождь’ имеется во всех доступных нам словарях хинди и урду. Едва ли есть основания считать это слово малоизвестным, как утверждает А. С. Крылова (Крылова 2017: 286). Нет веских причин и объявлять его диалектным заимствованием. Во всяком случае, в пользу этого не свидетельствуют никакие факты исторической фонетики. Поэтому мы считаем вполне оправданным включение в базу именно этого слова, а не его заимствованного синонима *bāriṣ* (< кл.-перс. *bāriṣ* ‘осадки’).

<sup>10</sup> Полный список форм см. в Kogan 2016: 253. Данное историко-фонетическое развитие обнаруживается и в других примерах. Ср., например, хинди-урду *dhīraj* ‘самообладание, стойкость, терпение’ < др.-инд. *dhīrya-/ dhīriya-*.

<sup>11</sup> Слово засвидетельствовано, например, в словаре Platts 1884. Ср. родственные формы, восходящие к древнеиндийской незаализованной основе *mās*, в наиболее генетически близких к хинди-урду индоарийских языках: дакхини, пандж. *mās-*, авадхи, кум. *māsu*.

### Поправки к списку бенгали

3. Предполагаемое А. С. Крыловой (Крылова 2017: 287) исконное происхождение бенг. *agun* ‘огонь’, хотя и допускается Р. Тернером в качестве одной из возможных альтернатив (Turner 1966: 821), представляется маловероятным, поскольку закономерным отражением др.-инд. *agni-* в бенгали является устаревшее бенг. *agi* ‘огонь’. Скорее в данном случае можно говорить об имевшей место в прошлом контаминации старого адаптированного санскритизма<sup>12</sup> с исконным словом, приведшей к появлению в современной бенгальской форме начального *a* (< *ā*) вместо ожидаемого *ɔ*.

\*\*\*

Отдельно хотелось бы остановиться на предлагаемом в работе А. С. Крыловой методе отбора лексики для включения в лексикостатистическую базу. В начале статьи автор заявляет: «Сразу отмечу, что мои представления о методике научного исследования не предполагают ни возможности получить «правильный» стословный список, пользуясь только словарём, ни возможности получить его от какого-то одного, пусть идеального, информанта» (Крылова 2017: 279). В чем состоят эти самые «представления о методике научного исследования», остается во многом неясным, поскольку нигде в дальнейшем тексте А. С. Крылова их полностью и эксплицитно не излагает. Лишь однажды, говоря о трудностях, возникающих при составлении стословного списка хинди, она кратко объясняет причину своего скептического отношения к словарям и нашему методу работы с ними: «... как в двуязычные словари ..., так и в толковые или синонимические ... входит лексика самых разных диалектов, зачастую без поясняющих помет. Почти каждое слово стословника имеет 5-10 синонимов... В этой ситуации тактика А. И. Когана избегать включения в список заимствований при наличии любого синонима, не противоречащего исторической фонетике, может привести к включению в стословный список слов, принадлежащих другим идиомам, расположенным в зоне хинди, и употребляющимся только на ограниченной территории» (Крылова 2017: 285).

Прежде всего, следует сказать, что утверждение об отсутствии поясняющих помет далеко не всегда соответствует действительности. Так, в использо-

<sup>12</sup> Такую трактовку допускает и Р. Тернер (Turner 1966: 821).

вавшемся нами и цитируемом А. С. Крыловой словаре Бархударов и др. 1972 диалектизмы снабжены специальной пометой *диал.* (диалектное). Разумеется, нельзя исключить, что какие-то диалектные элементы остались неотмеченными, но каждый раз, когда исследователь предполагает подобное, ему необходимо обосновывать свое этимологическое решение, т. е. показывать высокую вероятность заимствования того или иного слова из определенного диалекта. Сделать это без обращения к фактам исторической фонетики, как правило, попросту невозможно. В любом случае, едва ли оправдано судить обо всех словарях, исходя из «презюмции виновности», а *ргіогі* приписывая их авторам неспособность или нежелание решать проблему выявления диалектизмов<sup>13</sup>. В связи с этой проблемой А. С. Крылова утверждает также следующее: «Самым простым способом отсечь такие диалектизмы будет приблизительный подсчёт сравнительной употребительности синонимов в корпусе текстов» (Крылова 2017: 285). К сожалению, не было пояснено, в чем секрет эффективности такого способа. Совершенно не понятно, каким образом результаты подсчёта употребительности слова могут указывать на его этимологию.

Подсчётам по корпусам А. С. Крылова вообще отводит, как представляется, незаслуженно важную роль. Предлагаемый ей алгоритм отбора слов для стословника, по-видимому, можно кратко сформулировать так: отбирать слова, обнаруживающие максимальное число вхождений в веб-корпусах. Адекватность данного алгоритма далеко не бесспорна. Необходимым условием включения в лексикостатистическую базу той или иной лексической единицы является, как известно, стилистическая немаркированность последней. Можно ли считать наибольшую встречаемость в текстовом корпусе свидетельством принадлежности к нейтральному стилю? Ясно, что ответ на этот вопрос неоднозначен и зависит как от характера текстов, подобранных для создания корпуса, так и от специфики конкретного языкового материала. Так, в языке хинди стилистические различия в лексике

<sup>13</sup> В данной связи нужно также отметить, что хотя развитая синонимия, действительно, характерна для словарного состава хинди и урду, главным фактором, породившим такую ситуацию, является вовсе не диалектная неоднородность, а влияние более престижных в разные эпохи литературных языков. Это видно, в частности, из того факта, что основная часть синонимов приходится на санскритские, персидские, арабские и английские заимствования. Выделение подобной лексики обычно не сопряжено с какими-либо трудностями.

проявляются прежде всего в использовании в разных стилях заимствований из разных источников. При наличии двух синонимов, один из которых усвоен из санскрита, а другой — из персидского или (через посредство последнего) арабского языка, санскритизм практически всегда принадлежит к книжному или официальному стилю, персизм и арабизм же стилистически нейтральны. Если создатели корпуса отдавали предпочтение текстам, написанным на нормированном литературном языке, встречаемость санскритизмов может превысить встречаемость персизмов и арабизмов. По-видимому, с подобной ситуацией и столкнулась А. С. Крылова, получившая большее число вхождений для таких санскритских заимствований, как *mās* ‘мясо’ и *mahilā* ‘женщина’ в сравнении с их синонимами персидского происхождения *gōšt* и *aurat* соответственно.

Описанный случай представляет собой яркую (и притом, вероятнее всего, не единственную) иллюстрацию того факта, что результаты подсчётов по корпусам иногда могут скорее ввести в заблуждение, нежели прояснить реальную картину. Следует признать, что полученные А. С. Крыловой цифры едва ли могут рассматриваться в качестве надежных указаний на стилистическую характеристику слова. Это в свою очередь заставляет нас задаться вопросом о том, в какой мере эти цифры вообще показательны для лексикостатистических штудий.

Весьма странными представляются нам отдельные решения А. С. Крыловой, касающиеся замены конкретных слов в конкретных списках. Речь идет, прежде всего, о предложениях заменить одни заимствования на другие (поправка №3 к списку хинди, поправка №2 к списку бенгали) или включить в базу два синонима с отрицательными номерами (поправка №7 к списку хинди, поправки №6<sup>14</sup> и 10 к списку бенгали). Влияние таких поправок на результаты лексикостатистических подсчетов, очевидным образом, равно нулю, и оправданность их поэтому представляется по меньшей мере сомнительной.

Учитывая все вышесказанное, мы не считаем необходимым вносить в стословный список хинди

<sup>14</sup> Поправка №6 к списку бенгали вызывает особенно сильное недоумение. Предлагается внести в стословник англицизм *libhar* ‘печень’ (<англ. liver), употребляемый в бытовом языке «преимущественно в кулинарии» (Крылова 2017: 287). Лексику с подобной сферой употребления, несомненно, никоим образом нельзя считать «стословной». Ср. диагностические контексты для слова «печень», предложенные в статье (Kassian et al. 2010: 67): «Он вырезал печень из туши», «Человечья печень больше собачьей».

поправки, предложенные в статье (Крылова 2017). Поправки к спискам бенгали (исключая рассмотренные и отвергнутые выше), ория и куллуи принимаются нами, причина чему — более глубокое в сравнении с нашим знакомство А. С. Крыловой с материалом этих языков. Результаты лексикостатистических подсчетов по исправленной базе данных позволяют построить родословное древо, изображенное на рис. 1.

В заключение хотелось бы отметить, что даже с учетом всех наших возражений работа А. С. Крыловой дала чрезвычайно интересные результаты. Многие высказанные в ней соображения в силу своей дискуссионности, несомненно, будут способствовать более активному обсуждению как вопросов генетической классификации индоарийских языков, так и некоторых методологических проблем лексикостатистики и глоттохронологии.

#### Сокращения

асс. — ассамский  
 бенг. — бенгали  
 гудж. — гуджарати  
 др.-инд. — древнеиндийский  
 и.-е. — индоевропейский  
 кл.-перс. — классический персидский  
 кум. — кумауни  
 неп. — непали  
 пандж. — панджаби  
 перс. — персидский  
 потх. — потхохари

#### Сокращения названий языков на родословном древе

ASS — ассамский; AWD — авадхи; BNG — бенгали; BNJ — банджари; BRJ — брадж; DGR — догри; DKH — дакхини; DUM — думаки; GJR — годжри; GRH — гархвали; GUJ — гуджарати; HIM — химачали; HND — хинди-урду; HNK — хиндко; KCH — каччи; KNK — конкани; KUL — куллуи;

KUM — кумауни; LHD — лахда (мултани); MAI — майтхили; MAL — мальдивский; MAR — маратхи; MEW — мевати; MND — мандеали; NEP — непали; ORY — ория; PNJ — панджаби; PTH — потхохари; PRY — парья; RAJ — раджастхани (марвари); ROM — цыганский; SND — синдхи; SNG — сингальский; WGD — вагди; WPH — котгархи

#### Литература

Бархударов, А. С., В. М. Бескровный, Г. А. Зограф, В. П. Липеровский (под редакцией В. М. Бескровного). 1972. *Хинди-русский словарь в двух томах*. М.: Советская энциклопедия.  
 Коган, А. И. 2005. *Дардские языки. Генетическая характеристика*. Москва: Восточная литература.  
 Крылова, А. С. 2017. Лексикостатистика новоиндоарийских языков: взгляд полевого лингвиста. *Вопросы языкового родства* 15(3-4): 279–298.

#### References

Barkhudarov, A. S., V. M. Beskrovnyj, G. A. Zograf, V. P. Liperovskij. 1972. *Hindi-russkij slovar' v dvoukh tomakh*. Moskva: Sovetskaja enciklopedija.  
 Bloch, Jules. 1920. *La formation de la langue marathe*. Paris: É. Champion.  
 Kassian, Alexei, George Starostin, Anna Dybo, Vasilij Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.  
 Kogan, Anton. 2016. Genealogical classification of New Indo-Aryan languages and lexicostatistics. *Journal of Language Relationship* 14(3-4): 227–258.  
 Kogan, A. I. 2005. *Dardskije jazyki. Geneticheskaia kharakteristika*. Moskva: Vostochnaja literatura.  
 Krylova, A. S. 2017. Leksikostatistika novoindoarijskikh jazykov: vzgl'ad polevogo lingvista. *Journal of Language Relationship* 15 (3-4): 279–298.  
 Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.  
 Platts, John Tompson. 1884. *A dictionary of Urdu, classical Hindi, and English*. London: W. H. Allen & Co.  
 Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern / München: Francke Verlag.  
 Turner, R. L. 1966. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

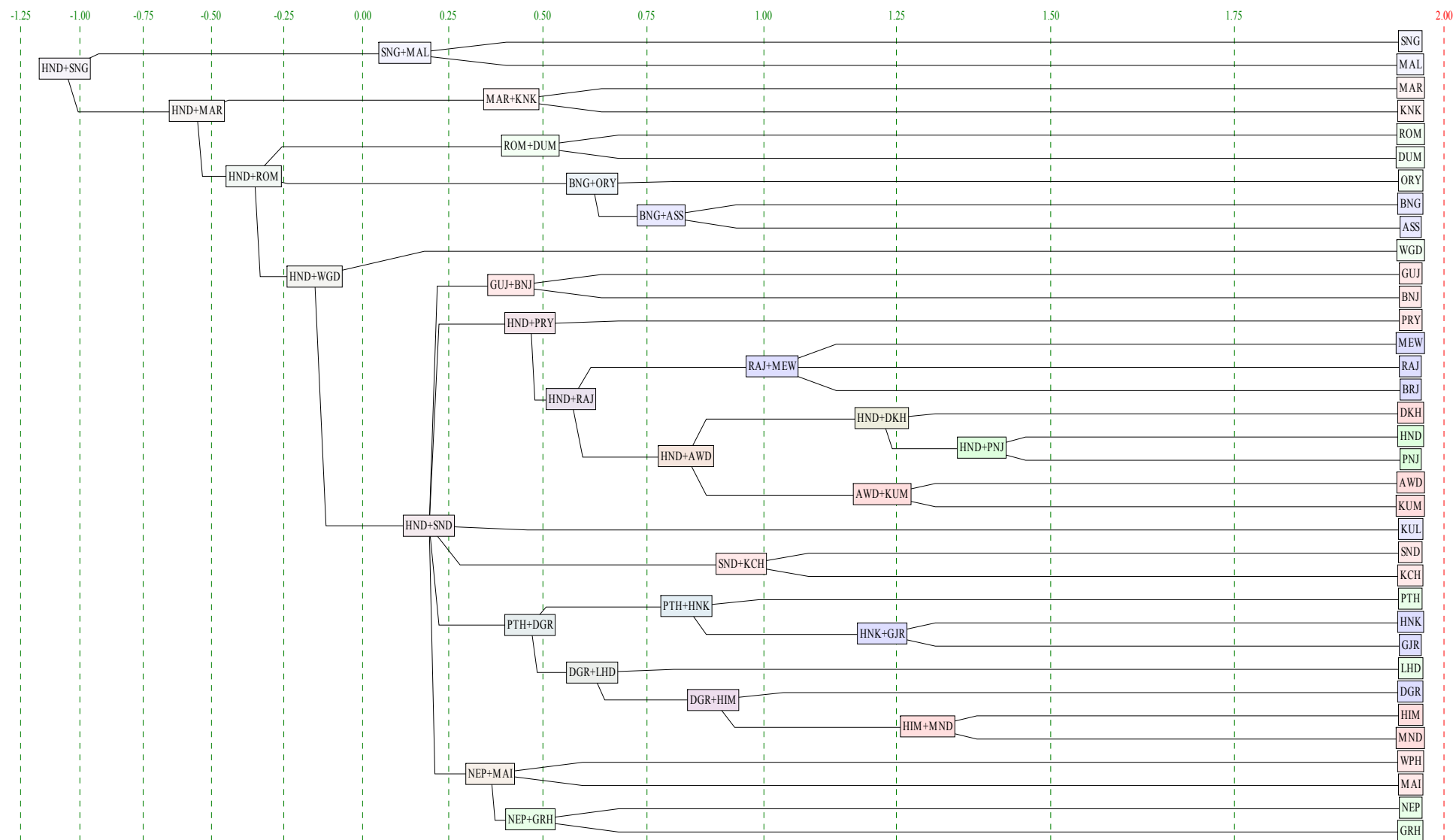


Рис. 1. Родословное дерево индоирийских языков, построенное с учетом принятых нами замечаний А. С. Крыловой