

Лексикостатистика новоиндоарийских языков: взгляд полевого лингвиста¹

Данная публикация представляет собой некоторые соображения о статье А. И. Когана «Genealogical classification of New Indo-Aryan languages and lexicostatistics» в №14 «Вопросов языкового родства». Автор высказывает ряд предложений по исправлению списков Сводеша, представленных А. И. Коганом, на основании данных исторической фонетики и семантики индоарийских языков в целом, а также анализирует списки языков хинди, бенгали, ория и куллуи, сопоставляя их со своими полевыми данными и сведениями из словарей и корпусов. После внесения предложенных исправлений автор сопоставляет деревья, построенные программой Starling на основании первоначальных и исправленных списков. Изменения не очень значительны, однако повышают достоверность результатов, а кроме того, показывают, что дальнейшее исправление списков может сделать дерево значительно более точным.

Ключевые слова: лексикостатистика, индоарийские языки, классификация языков, глоттохронология, полевая лингвистика, корпусная лингвистика, историческая фонетика

В 2016 году в №14 «Вопросов языкового родства» вышла статья А. И. Когана «Genealogical classification of New Indo-Aryan languages and lexicostatistics». Автор строит генеалогическое древо индоарийских языков на основании лексикостатистики по стословным спискам Сводеша. Основные теоретические положения статьи были давно мне знакомы из докладов автора, однако замечательной особенностью данной публикации является весьма обширное приложение, содержащее сами стословные списки по 35 языкам, включая проставленные для каждого слова индексы этимологической когнации. Масштаб постановки задачи впечатляет — фактически речь идет о попытке окинуть индоарийские языки взглядом с высоты птичьего полёта, и, конечно, при реализации такой глобальной задачи неизбежны отдельные недочёты и неточности. Именно поэтому я считаю необходимым дополнить масштабную панораму А. И. Когана результатами, полученными на уровне полевой работы. Хочу выразить благодарность А. И. Когану за предоставление мне файла для программы Starling, в котором он составлял свои списки, Е. В. Коровиной за помощь в построении деревьев и работе с программой Starling, Ю. В. Мазуровой, Е. А. Ренковской и А. В. Дыбо за ценные редакторские замечания.

Задача, за которую взялся А. И. Коган, мне, как полемому лингвисту, представляется неохватной в рамках одной статьи. Поэтому данная статья задумана как начало серии статей о стословных списках индоарийских языков. 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. Любой список, полученный из единственного источника, требует дальнейшей обязательной обработки и проверки — на других носителях, на корпусе текстов, на других словарях и пр. Таким образом, тща-

¹ Работа выполнена частично при поддержке РФФИ, проект 16-34-01040 «Грамматическое описание и словарь индоарийского языка куллуи», 2016–2018 гг.

тельный анализ всех тридцати пяти списков А. И. Когана — задача на много лет, в данной же статье основное внимание будет направлено на следующие списки:

- 1) хинди — язык, изучавшийся мной в университете в течение шести лет, а ныне основной язык-посредник для полевых исследований нашей группы в Индии. Хотя у меня нет стословного списка, записанного непосредственно от информанта, я считаю возможным проверить список А. И. Когана на основании статистики корпуса *Hindi Web 2013 (hiTenTen13)*, содержащего около 351 млн слов;
- 2) бенгали — язык, стословный список которого с учётом контекстов и стимулов на хинди и английском был записан мной от одного информанта в 2016 году. При проверке стословника бенгали был использован корпус *Bengali Web (BengaliWaC)*, содержащий более 11 млн слов. Кроме того, понадобилась проверка по ряду толково-двуязычных словарей (<http://www.english-bangla.com/>, Лоскутов 1974, Ghosh 2011);
- 3) ория — язык, стословный список которого также был записан мной в 2016 году. Расхождения в списках проверены с помощью толково-двуязычных словарей (Tripathy 2015, Biswal 2015), по толково-четырёхязычному словарю (Praharaj 1931–1940), а также по корпусу Goldhahn 2012 из 30 тысяч предложений; однако ряд слов из списка А. И. Когана в корпусе просто отсутствовал, и их изучение потребовало дополнительной проверки. Отсутствие общедоступного обширного корпуса ория компенсируется тем, что этот язык обладает уникальной письменностью; это дало мне возможность дополнительно использовать статистику Google²;
- 4) куллуи — стословные списки собраны Е. М. Шуванниковой в 2013 году в Кулу от пяти информантов. В дальнейшем большая часть лексики была проверена нами в ходе полевой работы 2014, 2016 и 2017 гг.

Помимо этих списков, в рамках статьи хотелось бы обсудить некоторые моменты, общие для индоарийских языков в целом — степень их исследованности, особенности семантики и исторической фонетики.

Общие спорные моменты

Основной вопрос, по которому я не согласна с автором статьи, — это степень исследованности индоарийских языков в настоящий момент. Действительно, для «больших» статусных языков (хинди, бенгали, непали и т. п.) уже созданы обширные словари и

² Разумеется, возможности для использования интернет-поисковиков очень ограничены, подробнее об этом см., например, Беликов 2016. Однако приводимые в настоящей статье примеры из Google связаны с попытками получить гораздо более точную информацию, чем отличие в употребительности двух синонимичных слов одного и того же языка на порядок или несколько порядков. Причина, по которой Google не загружает все результаты поиска, как и тот факт, что число результатов поиска является лишь приближительной оценкой, неоднократно обсуждалась в различных публикациях в сети. Что касается эксперимента с Yandex, показывающего абсурдные результаты для запросов «в Украину», «на Украину» и «Украину», то при повторении эксперимента в Google результаты оказались непротиворечивыми. Первоначальная проверка стословных списков хинди, бенгали и ория проводилась нами на основании статистики Google, но его перепроверка на корпусах практически не изменила результат (наиболее значительные изменения претерпел результат проверки списка ория, что, вероятно, объясняется недостаточным объёмом корпуса). Это свидетельствует о том, что числа статистики Google всё же не полностью случайны, а отражают некоторую реальность. Поэтому, признавая аргументы В. И. Беликова о невозможности использовать поисковики в качестве тонкого инструментария, мы всё же считаем возможным в некоторых пределах опираться на них для получения данных о языках, имеющих характерную письменность, но не имеющих подготовленного лингвистами корпуса достаточного объёма.

грамматики, изучена историческая фонология. Однако автор распространяет это положение на все 35 языков, которые входят в его базу, что в корне неверно (Kogan 2016: 233). По большинству малых индоарийских языков в лучшем случае доступен небольшой очерк, и адекватность этого очерка часто сомнительна. Кроме того, даже качественное описание, если в задачу автора не входил сбор стословника, может невольно ввести в заблуждение тех исследователей, которые пользуются им как источником. Так, Б. П. Махапатра в предисловии к очерку куллуи М. Р. Ранганатхи говорит о том, что информация получена от единственного носителя, чьё место жительства не указано (Ranganatha 1980: 16). При сопоставлении с полевыми данными нашей исследовательской группы в этом очерке (помимо множества ошибок и упрощений) очевидно сильное влияние лексики и грамматики хинди. Это может объясняться особенностями носителя или же специфической направленностью исследования, связанного с переписью населения, с языковой политикой Индии в целом и с задачей укрепления престижа хинди в частности (упомянутое предисловие в значительной степени посвящено возрастающему влиянию хинди как фактора объединения нации). Гораздо более подробно и точно грамматика М. Р. Тхакура, носителя куллуи и выдающегося лингвиста, написанная на хинди (Thakur 1975). Но, во-первых, она гораздо менее доступна, во-вторых, не содержит словаря, поэтому извлечение базовой лексики из неё потребовало бы полной росписи её текста. Однако и в этой книге можно найти множество неточностей. Так, ещё со времён Дж. А. Грирсона к говорам куллуи относят внешний и внутренний сираджи, и эта же информация упоминается в книге Тхакура (Тхакур 1975: 129). Наше поверхностное полевое знакомство с этими идиомами говорит о достаточно сильном их отличии от куллуи по существенным старым изоглоссам. Так, например, информанты сираджи давали личное и притяжательное местоимения первого лица единственного числа *ham-*, *mhar-* (а не *as-*, *asər-*), 'поле' *khech* вм. *chet*, 'муха' *takh-* вм. *tach-*.³ В грамматике также обнаруживаются существенные расхождения, так что не совсем ясно, насколько правомерно считать их диалектами куллуи. По всей видимости, первоначальная таксономия давалась в значительной степени по административным границам, и к куллуи были отнесены все говоры княжества Куллу.

Та же проблема связана и с идиомом химачали, стословный список которого входит в базу А. И. Когана. Название «химачали» относится к различным индоарийским идиомам, распространённым в штате Химачал-Прадеш. Х. Хендриксен, книгой которого А. И. Коган пользовался для создания списков, характеризует химачали следующим образом: «Himachali embraces the following dialects counting from the south and the west: Sirmauri, Jaunsari, Baghatī, Kyoṅṭhli, Koṭgarhi, Koci, Maṇdealī and Kuḷui» (Hendriksen 1986: 3). М. Р. Тхакур включает в этот список также чамбеали и бхадравахи (Тхакур 2012: 12–20). Всё это — достаточно сильно отличающиеся друг от друга идиомы. В статье А. И. Когана в качестве источника стословника химачали указываются носители языка (Kogan 2016: 239), однако не дано социолингвистических сведений об этих носителях — мы не знаем, в каких населённых пунктах они родились и выросли, насколько хорошо знают язык, на каких языках преимущественно общаются в последние годы и т. п. Таким образом, мы не можем быть уверены даже в том, о каком идиоме или идиомах идёт речь.

Поскольку описания малых новоиндийских языков оставляют желать лучшего, говорить о точности их исторической фонетики также не приходится. Например, в этимологическом словаре Тёрнера (Turner 1969–1985) содержится 8 (восемь!) примеров из куллуи. Главное, что хотелось бы сказать в разделе «Замечания общего характера»: прежде чем подсчитывать лексикостатистику, надо понять, лексикостатистику чего

³ Подробнее о значении этих отличий см. Hendriksen 1986: 192.

мы подсчитываем, оценить достоверность источника. В случае малых индоарийских языков эта задача требует прежде всего полевых исследований и заполнения лакун в описательной лингвистике.

Замечания к семантике

В данном разделе собраны замечания, относящиеся к особенностям семантики индоарийских языков. Лексикостатистика предполагает сопоставление слов с одинаковыми значениями, однако на практике в стословные списки зачастую попадают слова с разными значениями, сопоставлять которые неправомерно. На некоторые из таких «опасных мест», характерных для индоарийских языков, хотелось бы обратить внимание.

1. *Sab, s̄are* и их параллели в других языках возводятся к одному корню, этимология обозначена одним и тем же №1 (Kogan 2016: 239). В словаре Тёрнера это две разные этимологии *s̄arva* и *s̄ara*, которые отличаются и на праиндоевропейском уровне. В языке хинди (и во многих других языках) в действительности представлены и широко употребляются оба эти корня, с сохранением семантического различия *sab* ‘omnis’ и *s̄arā* ‘totus’. Но в каждый из списков А. И. Когана попал лишь один из корней. Поправка:⁴ добавляем в стословный список хинди слово *s̄arā*, присваиваем №581 ему и его параллелям в других языках (PTH *s̄are*, HNK *s̄are*, GJR *s̄arā*, LHD *s̄arā*, KOT *sarə*, HIM *s̄are*, PRY *sare*, BNJ *s̄ari*). Вообще говоря, правильно было бы присвоить отдельным значениям ‘omnis’ и ‘totus’ разные номера; последнее, однако, фактически предполагает расширение 100-словного списка, что выходит за рамки задач настоящей статьи.
2. В индоарийских языках обычно представлено два переходных глагола, соответствующих английскому *burn* (например, куллуи *dza:l̄ŋa* и *p^hikiŋa*). Один из них обозначает действие, целью которого является выделение света или тепла, другой — уничтожение объекта. В некоторых языках в списках А. И. Когана представлены оба эти глагола, в других — только один (Kogan 2016: 241). Как и в предыдущем случае, имело бы смысл учитывать оба значения под разными номерами, но расширение списка — отдельная задача.
3. Этимоны ‘heart’ и ‘liver’ в индоарийских языках довольно неустойчивы. В Индии распространено вегетарианство, поэтому названия внутренних органов знакомы почти исключительно врачам и, как правило, являются санскритскими заимствованиями. При попытке исключить заимствования велика вероятность получить слово, означающее сердце не как анатомический орган, а какместилище жизни или чувств.
4. ‘All’ vs ‘full’. В большом количестве языков значение ‘full’ представлено двумя словами с этимологиями №31 и №32 (Kogan 2016: 244). Это связано с наличием двух не всегда хорошо различимых значений ‘наполненный чем-либо (о сосуде, контейнере)’ и ‘целый, завершённый’, второе из которых сближается скорее с *all* ‘totus’, и не должно присутствовать в списке в значении ‘full’. Первое связано с корнем *bhar-*, второе — с корнем *pūr-*. Поправка: для всех списков, в которых представлены оба корня, удаляем слово, восходящее к *pr*-инд *n̄ira* (PTH *n̄irā*, HNK *n̄irā*, GIR *n̄iro*, DGR *n̄irā*, RAI *n̄ir*, GII *n̄irū*, ASS *nur*, AWD *nurhar*, KIM *nurro*, HND *n̄irā*, DKH *n̄irā*, PNI *n̄irā*, IHD *p̄irā*, SND *p̄iro*, MAR *purā*, BNG *pura*, KOT *purə*, HIM *p̄irā*, MND *p̄ira*, GRH *p̄iru*).

⁴ Здесь и далее поправки относятся к стословным спискам, приводимым в статье А. И. Когана.

5. В индоарийских языках, как правило, различаются слова для обозначения шеи и горла. При этом нередко на слово *neck* (шея) информант может дать слово, обозначающее горло. То, что в списке Сводеша содержится только слово *neck*, — случайность, но поскольку формула подсчёта разрабатывалась именно для этого списка, следует его придерживаться. Между тем в списке хинди для значения ‘neck’ дано *galā* ‘горло’, а не *gardan* ‘шея’, в то время как для бенгали ситуация обратная (Kogan 2016: 249). О большинстве языков, сведения по которым не так легко проверить, мы не можем с уверенностью сказать, какое из двух значений попало в список.

В разделе «Замечания к семантике», таким образом, мы вносим в списки 29 изменений, влияющих на подсчёты.

Замечания к этимологиям

Данный раздел содержит замечания, связанные с нашими текущими знаниями и представлениями об истории индоарийских языков в целом.

1. ‘Red’ *lāl* в хинди и его параллели в других языках в списках считается исконным словом и имеет этимологию №63 (Kogan 2016: 250). Подробнее об этой гипотезе происхождения *lāl* из др.-инд. **lohila* см. (Kogan 2005: 156). Однако ни **lohila*, ни его производные не зафиксированы в древнеиндийских и среднеиндийских текстах. В статье Тёрнера (Turner 1969–1985: № 11168 **lōhila*) перечисляются примеры только из дравидских и нурдистанских языков и одна (с вопросительным знаком) параллель из сингальского. При этом ни в одном из перечисленных языков корневого гласного *ā* нет. Слово *lāl* в индоарийских языках появляется только в средневековых новоиндийских текстах в тех же нескольких значениях, в которых оно употребляется в современном хинди: ‘рубин’, ‘красный’ и ‘любимый, дитя’ (Callewaert 2009: 1836). Следует отметить, что в средневековых новоиндийских диалектах уже распалась древнеиндийская падежная система, но в то же время ещё не была сформирована система послелогов, поэтому существительное в роли определения зачастую соединяется с определяемым существительным путём простого соположения. Ср., например, ранние рукописи Кабира:

rāma rasāina rasanā cāṣū (1)

Рама амрита вкус пробовать-1SgPrae
‘пробую вкус амриты Рамы’

ita mana mādira rahau nita coṣaī (2)

здесь разум храм жить-Imprv всегда хороший-Adv
‘здесь в храме разума живи всегда хорошо’

ṣaṭa dala kāvala nivāsiyā (3)

шесть лепесток лотос обитающий
‘обитающий в лотосе с шестью лепестками’

aṣṭa kavala dala bhītarā (4)

восемь лотос лепесток промежуток
‘между восемью лотосовыми лепестками’

kadalī kusama dala bhītarā (5)

лотос цветок лепесток промежуток
‘между лепестков цветка лотоса’

В этой ситуации затруднительно отличить существительное 'рубин' от прилагательного 'рубиновый, красный', и даже словосочетание от композита типа *tatpuruṣha*. По всей видимости, именно в этот период персизм *lāl* (из перс. *la'l* 'рубин') и получил распространение в индоарийских языках. Так как гортанная смычка не является характерной для них фонемой, она была утрачена. Можно заметить, что основные исконные цветообозначения хинди, как и большинство исконных базовых прилагательных, являются изменяемыми и заканчиваются в исходной форме на *ā*: ср. *kālā* 'чёрный', *pīlā* 'жёлтый', *harā* 'зелёный', *nīlā* 'синий', *baṛā* 'большой', *choṭā* 'маленький', *ṭhaṇḍā* 'холодный', *sukhā* 'сухой', *bharā* 'полный', *pūrā* 'полный, целый', *acchā* 'хороший', *naṇṇā* 'новый', *kharā* 'стоящий', *patlā* 'тонкий', но ср. *safed* 'белый' < перс., *garām* 'тёплый' < перс.⁵ Также существенным аргументом в пользу персидского происхождения *lāl* и его параллелей в других языках является то, что ни в одном из языков, в которых прошёл переход *-l > -l̥*, не зафиксирована форма с *-l̥*, однако встречаются формы с *-l*: ср. кулли *lāl* 'красный', но *phāl* 'плод', *gol* 'круглый'; ория *lāla* 'красный', но *phāla* 'плод', *golā* 'мяч' (*gol/gol* 'круглый', будучи абстрактным геометрическим понятием, видимо, имеет дублет-санскритизм). Вероятно, переход *-l > -l̥* прошёл до распространения персизма *lāl* в индоарийских языках. Форму *loyā* в думаки, вероятно, следует считать исконной и связывать с др.-инд. *lohita* (Turner 1969–1985: №11165 *lōhita*). Поправка: присваиваем *lāl* и его параллелям (HND *lāl*, DKN *lāl*, PNJ *lāl*, DGR *lāl*, SND *lālu*, GUJ *lāl*, BNG *lal*, HIM *lāl*, KUL *lāl*, MND *lāl*, ORY *lāl*, AWD *lāl*, KUM *lāl*, ROM *lolo*, BRJ *lāl*, GRH *lāl*, PRY *lal*, MAI *lāl*, MEW *lāl*, WGD *lāl*, HNK *lāl*, GJR *lāl*, LHD *lāl*, RAJ *lāl*, ASS *lal*, KOT *lāl*) отрицательные номера.

2. 'Nail' GUJ *nakh*, MAR *nakh*, BNG *nokh*, ASS *nokh*, HIN *nokh*, ORY *nakha*, а возможно, и ряд их параллелей в менее изученных языках, которым присвоен №14 (Kogan 2016: 241), следует признать санскритизмами, а не потомками др.-инд. *nakhá*. Интервокальное *kh* вместе с рядом других взрывных согласных в среднеиндийский период ослабляется в *h*, затем вовсе исчезая в части языков, таких как бенгали и гуджарати (Masica 1991: 180–181, 204). Хотя такое развитие взрывных согласных характерно не для всех слов, для *nakhá* оно хорошо зафиксировано (см., например, его рефлекс в стословных списках пенджаби, дакхини, браджа, а также хинди форму *nah* 'ноготь'). Сохранение интервокального *kh* возможно в качестве результата геминации, но в этом случае новоиндийский рефлекс должен содержать либо геминату, либо её следы в виде заместительного удлинения гласного (Masica 1991: 181, 187): *nakha > **nakkha > **nākha*. Ср. также рефлекс *múkha* 'лицо, рот' (этимология № 54). Поправка: для всех параллелей с *kh* (GUJ *nakh*, MAR *nakh*, BNG *nakh*, ASS *nakh*, HIM *nakh*, ORY *nakha*, WGD *nakh*, BNJ *nakh*) заменим №14 на №-14.

3. 'Mouth' WGD *muṇḍo* BNJ *muṇḍo* (Kogan 2016: 249) присвоен №54 (др.-инд. *múkha*). О фонетике вагди и банджари мы знаем мало, однако более вероятной представляется связь этих слов с *mūrdhán* 'голова' (Turner 1969–1985: №10247 *mūrdhán*)⁶. Ср. также кумаони *muṇḍo*, пенджаби *muṇḍ* 'голова'. Поправка: присваиваем словам вагди и банджари №554.

4. 'Mouth' BNG *mukh*, ASS *mukh* NEP *mukh*, KUM *mukh* (Kogan 2016: 249) следует признать санскритизмами по причине, изложенной в п. 2 для *nakh*. Поправка: для всех параллелей с *kh* заменим №54 на №-54.

⁵ Возможными исключениями являются *bhārī* 'тяжёлый' и *gol* 'круглый'. Однако фонетический облик обоих этих слов позволяет трактовать их как санскритские заимствования. Заметим, что второе из них неотличимо от существительного *gol* 'круг', от которого и произведено. Вероятно, таким же способом произведено *lāl* 'красный' от *lāl* 'рубин'.

⁶ Здесь и далее отсылки к Turner 1969–1985 даются по номерам словарных статей онлайн-версии, а не страниц книги).

5. Новоиндоарийские слова, восходящие к др.-инд. *bīja* 'seed' (Turner 1969–1985: №9250 *bīja*) утратили *j* в интервокальной позиции. Слова из списков А. И. Когана, сохранившие *j* (Kogan 2016: 252), следует признать санскритизмами. Поправка: присваиваем словам с сохранением *j* (HND *bīj*, DKH *bīj*, GJR *bīj*, LHD *bīj*, SND *b'iju*, RAJ *bīj*, BNG *bij*, ASS *biz*, KOT *bīdz*, HIM *bīj*, KUL *bejja*, ORY *bija*, BRJ *bīj*, GRH *bīj*, MAI *bīj*, KCH *bijj*, BNJ *bijā*) отрицательный номер.

6. Новоиндоарийские слова, восходящие к др.-инд. *sūrua/sūriua* утратили в первом случае *r*, во втором *u* (Turner 1969–1985: № 13574 *sūra*. Masica 1991: 169). Следовательно, все слова, сохранившие *r* и *u* (как правило, перешедшее в *i*), следует признать частично адаптированными санскритизмами (Kogan 2016: 253). Поправка: присваиваем HND *sūraj*, DKH *sūrīj*, PNJ *sūraj*, DGR *sūraj*, RAJ *sūraj*, GUJ *sūraj*, HIM *sūraj*, KUL *surjā*, MND *sūraj*, AWD *suruj*, KUM *sūraj*, BRJ *sūraj*, MAI *surīj*, KCH *sūraj*, MEW *sūraj*, ASS *xuruz*, GRH *sūraj*, BNJ *surjyā* отрицательные номера.

Итого в данном разделе вносится 75 влияющих на подсчёты изменений в списки.

Поправки к списку хинди

Хинди — официальный язык Республики Индия в целом, а также многих штатов, входящих в её состав. Так называемая «зона хинди» в действительности представляет собой огромный континуум различных идиомов, каждый из которых следовало бы вносить в «идеальную лексикостатистическую базу» отдельно. Лишь немногие из них уже внесены в базу А. И. Когана благодаря их сравнительно высокому статусу и хорошей изученности. Прочие имеют статус местных диалектов хинди. Однако первоначальная диалектная база хинди — это околodelийские диалекты хариани и кхари боли. При этом как в двуязычные словари (Platts 1884, Бархударов 1972), так и в толковые или синонимические (Kumar 1997, ряд интернет-словарей) входит лексика самых разных диалектов, зачастую без поясняющих помет. Почти каждое слово стословника имеет 5-10 синонимов. В силу низкой нормированности хинди, такие слова могут легко встретиться в песнях и кинофильмах, художественной литературе и публицистике. 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. Самым простым способом отсеять такие диалектизмы будет приблизительный подсчёт сравнительной употребительности синонимов в корпусе текстов. Запросы вводились письмом деванагари. Опытным путём установлено, что более частотны написания без диакритики «нукта» и с заменой диакритики «чандрабинду» на диакритику «анусвара», поэтому в этих пунктах мы отходим от словарной орфографии. В случае большого количества синонимов, как правило, можно выделить один или несколько наиболее употребительных, тогда как у остальных частотность ниже в 10-20 и более раз. Среди этих наиболее употребительных слов можно выбрать исконные, если они представлены. Таким образом нами были проверены несколько вызвавших у нас сомнение пунктов из списка А. И. Когана.

1. 'Blood' *lahū* (Kogan 2016: 240) даёт 3105 вхождений в Hindi Web 2013. Основное бытовое слово — персизм *khūn* (31537 вхождений), основное медицинское — санскритизм *rakt* (16461 вхождение). Исконных слов сопоставимой частотности нет. Поправка: вычёркиваем *lahū*, добавляем два заимствования.

2. ‘Earth’, как правило, имеет много переводов и семантических тонкостей. Не совсем понятно, почему из списка хинди исключено слово *miṭṭī*, параллели которого представлены в других языках (Kogan 2016: 243). Hindi Web 2013 для *dhartī* и *miṭṭī* даёт сопоставимое число вхождений (33698 и 24256 соответственно). При этом семантика этих корней существенно различается (не только в хинди, но и в других языках) Если обратиться к такому источнику языковой информации, как Википедия на языке хинди⁷, то мы видим, что запрос *dhartī* переадресует нас на статью «планета Земля», а запрос *miṭṭī* — на статью «почва» (обе статьи озаглавлены санскритскими заимствованиями, за которыми идут перечисления синонимов). Действительно, семантика этих слов несколько отличается, и адекватное решение этой проблемы требует расширения списка, что не входит в наши задачи (см. раздел «Замечания к семантике»). Поправка: включаем *miṭṭī* в список хинди.

3. ‘Man’ в списке хинди (Kogan 2016: 248) представлено словом *mard* (8398 вхождений в Hindi Web 2013). Более последовательным было бы включение прежде всего эквивалента *ādmī* (125042 вхождения). Результаты подсчёта это не изменит, так как оба эти слова являются заимствованиями. Поправка: заменяем *mard* на *ādmī*.

4. ‘Meat’ представлено только словом *goṣṭ* (Kogan 2016: 248). Однако запрос *mās* даёт 8614 вхождений в Hindi Web 2013, *goṣṭ* — 1127. *goṣṭ* — несомненное заимствование; *mās* может быть как санскритским заимствованием, так и исконным словом. В других языках его параллели широко представлены и отмечены как исконные (№124). Поправка: заменяем *goṣṭ* на *mās* №124 в списке хинди.

5. ‘Rain’ в списке хинди — *mēh* (Kogan 2016: 250). Это слово малоизвестное, в словарях встречается, но, вероятно, попало туда из какого-то диалекта (например, раджастан). Hindi Web 2013 даёт для *mēh* 73 вхождения, что для слова из стословника хинди удивительно мало. Самым распространённым словом для обозначения дождя можно считать персизм *bāriṣ* (30702 вхождения); конкурентов он, вероятно, не имеет. Статья в словаре-тезаурусе (Kumar 1997) озаглавлена санскритизмом *varṣā* (14739 вхождений); среди перечисленных в статье синонимов из исконных слов самыми употребительными следует считать *jal* (63936 вхождений, основное значение ‘вода’), *pānī* (166879, основное значение ‘вода’) и *barsāt* (11389, основное значение ‘сезон дождей’), однако в этом случае отсеять другие значения слов мы не можем. Поправка: убираем *mēh*, вписываем *bāriṣ* и *varṣā* с отрицательными номерами.

6. ‘White’ в списке хинди *ujlā*, *ciṭṭā* (Kogan 2016: 256). Запрос в Hindi Web 2013 даёт на них 495 и 148 вхождений соответственно. Кроме того, *ujlā* имеет основное значение ‘ясный, сияющий, чистый’. Наиболее употребительным и семантически точным из слов, означающих ‘белый’, является персизм *safed* (15588 вхождений). Сопоставимых по употребительности с *safed* слов нет. Поправка: убираем *ujlā*, *ciṭṭā*, вписываем *safed* с отрицательным номером.

7. ‘Woman’ в списке хинди *aurat* (Kogan 2016: 256) — 27553 вхождения. Логично было бы добавить также санскритизмы *mahilā* (106114 вхождений) и *strī* (56179 вхождений). Из нашего полевого опыта скорее следует, что *aurat* более частотно, чем *strī*, в устном употреблении, поэтому можно предположить, что *strī* несколько чаще встречается в Hindi Web 2013 за счёт возможного преобладания в корпусе официального стиля. Поправка: добавляем *mahilā*, что не влияет на результаты подсчётов, так как все перечисленные слова являются заимствованиями.

Таким образом, в разделе «Поправки к списку хинди» мы постулируем 8 изменений, влияющих на подсчёты.

⁷ Разумеется, речь идёт не об источнике научной информации, а о Википедии на языке хинди как о ресурсе, который редактируется преимущественно носителями языка хинди.

Поправки к списку бенгали

Список бенгали был получен мной в 2016 году от одной информантки. Параллельно фиксировались слова разговорного бенгали и «шадху бхаша», санскритизированного литературного языка, что позволило отсеять часть синонимов. Расхождения нашего списка со списком А. И. Когана потребовали более тщательного изучения по словарям, корпусу Bengali Web и ряду источников в сети Интернет. Предлагаются следующие поправки к списку А. И. Когана.

1. 'All'. Помимо *śob* 'omnis' (Kogan 2016: 239) в нашем списке представлены два синонима для 'totus' — *śara* и *puro*. Первый, по словам информантки, употребляется в основном с временными периодами. Оба достаточно хорошо представлены в Bengali Web (3317 и 2680 вхождений соответственно). Поправка: включаем *śara* №581 и *puro* №32 в список.

2. 'Cloud' BNG *barid* №-2 (Kogan 2016: 241) при поиске в Bengali Web дал 0 вхождений⁸. В то же время *megh*, полученное от информантки, дало 372. Поправка: заменяем *barid* на *megh* (№-1). В подсчётах эта замена роли не играет, так как *megh* — санскритизм (см. п. 2, 4 раздела «Замечания к этимологиям»).

3. 'Fire' BNG *agun* №-4, в списке А. И. Когана обозначенное как заимствование (Kogan 2016: 244), вслед за Тёрнером считаем исконным словом (Turner 1969–1985: №14198 *agni*). Санскритизмы в бенгальском языке обычно сохраняют санскритскую орфографию, но произносятся по бенгальским правилам. Так, в качестве слова «шадху бхаша» наша информантка дала *agni*. Форма *agun* не вполне ясна, но качество гласного говорит о его удлинении, что не позволяет считать слово санскритизмом. Поправка: присваиваем *agun* №27.

4. 'Full' BNG *bhōra* №31 (Kogan 2016: 244) является словарной формой глагола 'наполнять', заменим его на полученное нами прилагательное *bhōrti*, что, впрочем, не должно отразиться на подсчётах.

5. 'Hair' BNG *keś* №185 (Kogan 2016: 245) даёт 25 вхождений в Bengali Web. Моя информантка вместо этого слова выдала синоним *chul* (551 вхождение). Поправка: удаляем *keś*, заменяем на *chul* №308 (Turner 1969–1985: № 4883 *cūḍa*).

6. 'Liver' BNG *ḷkṛit* (Kogan 2016: 247) информанткой было маркировано как «шадху бхаша», бытовое наименование — англицизм *libhar* (преимущественно в кулинарии). В Bengali Web *ḷkṛit* и *libhar* дают 22 и 40 вхождений, в Google 53700 и 556000 соответственно. Поправка: добавляем *libhar* с отрицательным номером в список. На подсчёты это изменение не влияет, оба слова являются заимствованиями.

7. 'Many' BNG *bōhi* №51 (Kogan 2016: 248). Наша информантка дала бытовое *mek*, «шадху бхаша» — *procur*. *bōhi* даёт 3560 вхождений в Bengali Web, *mek* — 23201 вхождение. *bōhi* присутствует в бенгальском языке в том числе как первая часть санскритских композитов-санскритизмов, однако различить санскритизм и исконное слово по фонетическому облику в этом случае не представляется возможным, поэтому *bōhi* как самостоятельное слово следует считать исконным. Форма *mek*, напротив, — очевидный санскритизм. В связи с высокой встречаемостью мы всё же включим его в список с присвоением №-1, хотя на подсчёты это не влияет.

8. 'New' BNG *nōbō* №57 (Kogan 2016: 249) было маркировано информанткой как «шадху бхаша», и это не случайно. Фонетический облик этого слова не даёт оснований считать его исконным. Форма *nōbō*, несомненно, происходит из др.-инд. *nāva*, однако по-

⁸ Хотя словари и подтверждают слово *barid* в этом значении, при поиске в Google соотношение между *barid* и *megh* составило 98 к 2130000.

томком др.-инд. *nāva*, удовлетворяющим требованиям исторической фонетики, в действительности является бенгальское *no* ‘младший, четвёртый (о ребёнке в семье)’ (Turner 1969–1985: №6983 *nāva*, Ghosh 2011, Лоскутов 1974). Ср. аналогичный пример: Turner 1969–1985: №10431 *yāva* ‘barley’ = бенгальское *ja* (в современных словарях литературного бенгальского языка это слово представлено в форме *ji* ‘овёс’, см. Ghosh 2011, Лоскутов 1974). Исходя из этих соображений, следовало бы приписать *nɔbo* отрицательный номер; но в силу его низкой частотности (241 вхождение в Bengali Web) мы предлагаем просто не включать его в список. В качестве разговорного синонима *nɔbo* наша информантка дала *notun* (13913 вхождений в Bengali Web) — слово, по всей видимости, связанное с др.-инд. *nūtana*⁹, прошедшее не вполне понятные фонетические изменения; вероятно, его следует считать частично адаптированным санскритизмом¹⁰. Поправка: приписываем *nɔbo* №-57, добавляем *notun* с №-1.

9. ‘Rain’ BNG *badɔl* №127 (Kogan 2016: 250). В нашем списке в качестве бытовых слов даны *badɔl* (101 вхождение в Bengali Web) и *ɟɔl* (893 вхождения). Основное значение *ɟɔl* — ‘вода’, поэтому проверить частоту его употребления в значении ‘дождь’ затруднительно, однако это значение присутствует в большинстве словарей. В связи с малой частотностью *badɔl* вероятно, следует добавить *ɟɔl* в список в качестве второго слова с №319.

10. Вместо ‘road’ BNG *poth* №-8 (Kogan 2016: 251) в нашем списке значится персизм *rasta*. Для *poth* обнаружено 8008 вхождений в Bengali Web, для *rasta* — 2390 вхождений. Разница незначительна, и мы можем добавить *rasta* в качестве второго слова с отрицательным номером, что, впрочем, не должно повлиять на подсчёты.

11. ‘Root’ BNG *šikɔɾ* №313 (Kogan 2016: 251) даёт 12 вхождений в Bengali Web. В нашем списке его место занимает *mul* (6143 вхождения). Поправка: заменяем *šikɔɾ* на *mul* №259.

12. ‘Tail’ BNG *langul* №317 (Kogan 2016: 254) даёт 0 вхождений в Bengali Web, а в Google — около 2000. В нашем списке значится *lej* (141 вхождение в Bengali Web, около 536000 в Google). Малая встречаемость *langul* позволяет его исключить, однако в онлайн-словаре мы находим форму *lenguɾ*, которая также приводится в Turner 1969–1985: №11009 *lāṅgūlá*. Эта форма не встречается в Bengali Web, но в Google имеет около 50-80 тысяч вхождений. Форма *lej* возводится к Turner 1969–1985: №10915 *lañja*. По всей видимости, *lañja* и *lāṅgūlá* связаны на праиндоевропейском уровне с глаголом *leṅg* ‘качаться’, который приведён в статье Тёрнера. Поправка: заменяем *langul* на *lenguɾ* №317, добавляем синоним *lej* №317.

13. ‘White’ BNG *dhɔbɔl* №267 (Kogan 2016: 256) в нашем списке отсутствует. Следует отметить, что *dhɔbɔl* по своему фонетическому облику — санскритизм. Действительным фонетическим соответствием др.-инд. *dhavala* ‘белый, чистый’ является бенгальское *dhɔla* ‘белой масти, светлокожий’ (см. Turner 1969–1985: №6767 *dhavalá*, а также пункт 8 данного раздела о ‘new’ BNG *nɔbo*). Учитывая это, *dhɔbɔl* справедливо было бы присвоить отрицательный номер. Однако то, что в Bengali Web это слово встречается 1 раз, в то время как *śada* из нашего списка (*śada* также является заголовком статьи в бенгальской Википедии о белом цвете) встречается 869 раз, позволяет его вовсе исключить, заменив на персизм *śada* с отрицательным номером.

14. ‘Woman’ BNG *meɟe* №321 (Kogan 2016: 256): действительно, довольно распространённое слово (4302 вхождения в Bengali Web), но основное его значение — ‘girl’, что подтверждает, например, статья в бенгальской Википедии. Наша информантка в качестве основного бытового слова дала *mɔhila* (2248 вхождений в Bengali Web), статья в бенгаль-

⁹ См. Turner 1931: <http://dsal.srv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.1:1:5036.turner.1329326>.

¹⁰ Закономерный рефлекс др.-инд. *nūtana* должен был бы выглядеть как ***nūan*.

ской Википедии о человеке женского пола озаглавлена словом *nari* (8533 вхождения в Bengali Web), однако оба этих слова следует считать заимствованиями. Таким образом, исконного слова для понятия 'woman' в бенгали нет. Поправка: заменяем *meue* на *mohila* и *nari* с отрицательными номерами.

15. 'Yellow' BNG *hólode* (Kogan 2016: 257): присутствует в нашем списке в форме *hólud*, это слово является и заголовком статьи о жёлтом цвете в бенгальской Википедии. Количество вхождений *hólode* в Bengali Web — 12, в Google — около 167 тысяч, *hólud* в Bengali Web — 263, в Google — около 1 миллиона 830 тысяч. Однако *hólud* обозначает также куркуму, распространённую индийскую пряность. Поэтому выяснить относительную употребительность этих двух форм затруднительно. Возможная поправка: заменяем *hólode* на *hólud* без замены номера.

Итого в разделе «Поправки к списку бенгали» насчитываем 9 изменений, влияющих на подсчёты.

Поправки к списку ория

Слословный список языка ория был получен мной в 2016 году от одного информанта. К сожалению, объём корпуса ория из Leipzig Corpora Collection, который мне удалось найти, составляет всего 30 тысяч предложений, и многие слова из списка А. И. Когана в нём не представлены, а слова из моего списка представлены слабо. Но благодаря тому, что язык ория обладает уникальной письменностью, результаты запросов в Google с очень высокой вероятностью будут относиться именно к языку ория (в отличие от хинди, в котором базовая лексика может графически совпадать с лексикой непали, маратхи или санскрита, а также бенгальского, базовая лексика которого может графически совпасть с ассамской). В силу этих обстоятельств мы приволим статистику и по Leipzig Corpora Collection, и по Google для тех пунктов списков, в которых имеются существенные расхождения. В списке А. И. Когана слова ория приводятся в системе транслитерации Яляу слова в транскрипции, так же, как это сделано в его списке бенгали. Таким образом, то, что А. И. Коган передаёт как *a* и *ā*, я передаю как *ɔ* и *a* соответственно.

1. 'Ashes' ORY *chāra* №140 (Kogan 2016: 239): словарями не подтверждается, как и нашим списком (в отличие от *pāuñśa* №567). В корпусе слово не представлено. В Google слово представлено слабо (около 900 ссылок), вероятно, в основном, как вариант написания *chār* 'ничтожный'. Приписывание ему значения 'пепел', по-видимому, связано с каким-то из сочетаний, обозначающих уничтожение, в том числе сожжение до состояния пепла. Поправка: убираем *chāra* из списка.

2. 'Bark' ORY *bakkala* №328 (Kogan 2016: 239): согласно словарям, следует транслитерировать как *bakala*.

3. 'Breast' ORY *buku* №297 (Kogan 2016: 241): по словам нашего информанта, относится к классическому ория, в отличие от *chāti* №12, которое подтверждается нашим списком. В корпусе *buku* не представлено, тогда как полученное нами от информанта *chāti* имеет 40 вхождений в корпус. В Google *buku* встречается около 600 раз, *chāti* — около 19 тысяч раз. Поправка: убираем *buku* из списка.

4. 'Cold' ORY *thanda* №458 (Kogan 2016: 242): в наш список не попало. Информант дал санскритизм *fitɔ*. По количеству вхождений в корпус они вполне сопоставимы (39 и 94), поправки в список А. И. Когана не требуются.

5. 'Come' ORY *ail-* №181 (Kogan 2016: 242): перфектная основа, в которую входит стандартный перфектный суффикс *-il-*. Для выяснения вопроса о происхождении основ

ās- и *a-* требуется отдельное исследование, выходящее за рамки целей данной статьи. Полной парадигмы от основы *a-* не образуется. В словаре Прахараджа лишь упоминается о возможности употребления форм типа *āila* или *aila* (и прочих) вместо *āsila* (и прочих) в прошедшем времени (Praharaj 1931–1940: <http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.0:110.praharaj>)¹¹. Однако очевидно, что перфектный суффикс следует исключить. Поправка: заменяем *ail-* на *a-*. На подсчёт эта замена не влияет.

6) ‘feather’ ORY *pālak* №306 (Kogan 2016: 244): согласно словарю Biswal 2015: 870, значит ‘nourishing, rearing up’. Указание на значение ‘перо’ находим в словаре Прахараджа, однако здесь говорится лишь о диалектном слове, распространённом в сообществах ория в Миднапуре в Западной Бенгалии, в окружении бенгалоязычного населения. В той же словарной статье Прахарадж толкует эту форму через слово *pṛṛṛ* (Praharaj 1931–1940: <http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.28:5159.praharaj.3504923>). Очевидно, речь о заимствовании бенгальского *pālṅk* ‘перо’. Слово из нашего списка, персизм *pṛṛṛ*, подтверждается словарями ория (Biswal 2015: 818, Tripathy 2015: 326). Распространённость его оценить затруднительно, так как оно имеет (вероятно, более распространённый) омоним — местоимение ‘другой’. В словаре Прахараджа оно присутствует только в толкованиях, но не в заголовке словарной статьи, из чего можно предположить, что это слово является самым общеизвестным в данном значении. Поправка: заменяем *pālak* на *pṛṛṛ* (в транслитерации А. И. Когана — *para*) №-1.

7. ‘Foot’ ORY *pā* №30 (Kogan 2016: 244): подтверждается словарём Прахараджа, но не встречается в корпусе, а в Google находится только 121 вхождение. *pādṅ* из нашего списка даёт 121 вхождение в корпус и 35 тысяч — в Google. *pādṅ* — санскритское заимствование, но исконные слова сопоставимой частотности отсутствуют. Поправка: заменяем *pā* на *pādṅ* №-1.

8. ‘Full’ ORY *bhari* №31 (Kogan 2016: 244): согласно словарям ория (Biswal 2015: 1204, Tripathy 2015: 372, 437), имеет значение ‘heavy’, но не ‘full’. Представленное в нашем списке *pūrā* подтверждается словарями и встречается в корпусе 18 раз. Поправка: убираем *bhari* из списка.

9. ‘Hair’ ORY *cuṭi* №308 (Kogan 2016: 245): согласно словарю (Biswal 2015: 448), обозначает исключительно вершину горы. На той же странице см статью о *cuṭṅ* ‘hair of the head’, имеющее также значения ‘crest of cock’, ‘top of a temple’, ‘hump of a bull’. В корпусе *cuṭṅ* встречается 2 раза. В нашем списке представлены два слова: *bāl* и *cuṭi*, при этом второе отмечено информантом как слово прибрежного (околостолычного) диалекта, а первое — как слово западное или племенное. Однако словарь (Biswal 2015: 445) переводит *cuṭi* как ‘a tuft of hair’ (пучок волос, причёска), а словарь (Tripathy 2015: 422) толкует *hair* в основном через слово *bāl*, вовсе не упоминая *cuṭi*. В корпусе *cuṭi* также встречается 2 раза, по сравнению со 119 для *bāl*. Поправка: заменяем *cuṭi* на *bāl* №36.

10. ‘Heart’ ORY *hrudaya* №-5 (Kogan 2016: 246): слово, согласно словарям и свидетельству информанта связанное скорее с духовным понятием. Сердце в анатомическом смысле — *hrutpiṅḍṅ*. Оба они являются санскритизмами. Поправка: заменяем *hrudaya* на *hrutpiṅḍṅ* без замены номера.

11. ‘Meat’ ORY *māesa* №124 (Kogan 2016: 248): вероятно, опечатка. По словарям и свидетельству информанта — *māsa*. Поправка: заменяем *māesa* на *māsa*, на подсчёт замена не влияет.

12. ‘Moon’ ORY *candramā* №-8 (Kogan 2016: 248): малоупотребительный санскритизм (2 вхождения в корпусе). В нашем списке значится *jnhṅ*, которое также является заго-

¹¹ Здесь и далее отсылки к Praharaj 1931–1940 даются гиперссылками на онлайн-версию.

ловком статьи о Луне в Википедии на ория и даёт 14 вхождений в корпус. Это слово отсутствует в статье (Turner 1969–1985: №5301 *ḷyótsnā*), но очевидно, что это этимологическая параллель к *ḷyótsnā*. Поправка: заменяем *candramā* на *ḷṇhṇ* №340.

13. ‘Round’ ORY *gola* №66 (Kogan 2016: 251): вероятно, опечатка. Прилагательное ‘круглый’ в словаре (Biswal 2015: 370) приводится в двух вариантах написаний: *gol/gol*, то есть существует и в форме санскритизма, и в форме исконного слова. Поправка: заменяем *gola* на *gol* без замены номера.

14. ‘Say’ ORY *bol-* №70 (Kogan 2016: 251): при поиске в словарной форме *bolibā* даёт 1 вхождение в корпус и около 300 в Google, что несопоставимо с 30 вхождениями в корпус и 42 тысячами в Google для представленного в нашем списке *kṣh-* № 69 в той же форме. Поправка: убираем *bol-* из списка.

15. ‘Seed’ ORY *bīja* №72 (Kogan 2016: 252): по его фонетическому облику следует считать санскритизмом (см. Masica 1991: 180–181, также пункт 5 из раздела данной статьи «Замечания к этимологиям»). В корпусе это слово даёт 7 вхождений. В нашем списке значится *māii* (52 вхождения), заимствование из дравидийских языков (Burrow 1984: #4639 Ga. (Oll.) *mājik*). Поправка: помимо замены номера *bīja* на отрицательный, добавляем *māji* с отрицательным номером.

16. ‘Sun’ ORY *surya* №-11 (Kogan 2016: 253): вероятно, результат ошибочной транслитерации. В действительности варианты произношения, приведённые в словаре Прахараджа — *sūrja*, *sūruj*, *suraj*, *sūrjya*. В нашем списке также значится *surjṣ*. Поправка: заменяем *surya* на *surjṣ* без замены номера — всё это следует считать различным образом адаптированными санскритизмами (см. пункт 6 раздела «Замечания к этимологиям»).

17. ‘Tail’ ORY *languḷa* №317 (Kogan 2016: 254): и в словарях, и в корпусе, и в Google находится только в форме *lānguḷa* (12 вхождений в корпус, около 100 вхождений в Google). В нашем списке значится *lanjṣ* (25 вхождений в корпус, около 1000 вхождений в Google). Поправка: заменяем *languḷa* на *lānguḷa*, добавляем *lanjṣ* №317.

18. ‘Walk (go)’ ORY *gal-* №93 (Kogan 2016: 255): отсутствует в словарях и, по видимому, является результатом ошибочной транслитерации специфического редкого начертания слога *cā* (𑌕) в письменности ория. Обычное его написание — 𑌖, но в этом редком начертании он сходен со слогом *ga* (𑌗). В действительности речь о глаголе *cāl-*, зафиксированном и в нашем списке. Заметим, что этот глагольный корень представлен, например, и в хинди. Но употребление корней *ja-* и *cal-* в хинди (как и, судя по словарным толкованиям, *ji-* и *cāl-* в ория) сильно различаются. Если первый скорее можно соотнести с английским *go* (‘идти, уходить’), то второй — с английским *walk* (‘ходить, передвигаться пешком’). Поэтому сопоставлять их между собой было бы неправомерно. Поскольку семантические соответствия *cāl-* отсутствуют в списках А. И. Когана, не стоит включать этот глагол в список ория, а значение ‘walk (go)’ следует уточнить. Поправка: удаляем *gal-* из списка, меняем значение ‘walk’ на ‘go’ для всех языков.

Таким образом, в разделе «Поправки к списку ория» 9 изменений, влияющих на подсчёты.

Поправки к списку куллуи

Список языка куллуи составлен А. И. Коганом с помощью грамматического очерка М. Р. Ранганатхи, содержащего наиболее полный словник куллуи на момент написания статьи. Гораздо более подробная и достоверная грамматика М. Р. Тхакура содержит значительное количество лексики, но эта лексика не объединена в словарь, а разбросана по

разным разделам грамматики, часто без перевода на хинди. В настоящее время более полный, периодически обновляемый словарь куллуи можно найти на сайте www.pahari-languages.ru. Он включает данные из словаря Ранганатхи и постоянно пополняется лексикой из грамматики Тхакура и нашими полевыми данными. Согласно сложившейся в нашей научно-исследовательской группе традиции, транскрипция слов куллуи даётся не в индологической транскрипции, а в МФА. Стословный список языка куллуи был собран в 2013 году Е. Князевой (Шуванниковой) от пяти информантов из деревни Наггар; позже семантика многих из этих слов была уточнена нами в ходе работ по сбору словаря и корпуса устных текстов. Разберём те случаи, где встречаются расхождения со списком А. И. Когана.

1. 'Ashes' в списке А. И. Когана не представлено. В нашем списке значится $b^h\text{ɔ:s}$. Поправка: добавляем $b^h\text{ɔ:s}$ №489.

2. 'Bark' KUL *khol* №552 (Kogan 2016: 239): в нашем списке отсутствует. Трое из пяти информантов назвали слово *lepəɾa*. Возможная его этимология связана с (Turner 1969–1985: 11114 *lēpya* 'plaster'). Поправка: заменяем *khol* на *lepəɾa* №490.

3. 'Bite' KUL *kāṭ-* №8 (Kogan 2016: 240): представлено в наших списках от двух информантов. Также двое информантов представили корень $k^h\bar{a}$ -, один — *ʃok-* и один — *tse|a-*. Вносим корень $k^h\bar{a}$ - №490, значение которого было позже подтверждено информантами в ходе дальнейшей работы.

4. Помимо 'black' KUL *kāla* №9 (Kogan 2016: 240), двое информантов также дали прилагательное *ʃitta*. Поправка: добавляем *ʃitta* №450.

5. 'Dry' в списке куллуи А. И. Когана не представлено. В нашем списке значится *fukka*. Поправка: добавляем *fukka* №20.

6. Кроме 'egg' KUL *aṇḍa* №24 (Kogan 2016: 243), двое из наших информантов дали также эквивалент *ḍanna* — слово с не вполне ясной этимологией; видимо, следует считать, что оно относится к тому же корню, но с метатезой. В этом случае его внесение в список не повлечёт за собой изменений при подсчётах, а при дальнейшей работе со списками, возможно, будет найдена более удовлетворительная этимология. Поправка: вносим *ḍanna* №24.

7. 'Eye' KUL *hək^{hi}* №25 (Kogan 2016: 243): в нашем списке не представлено. Вместо него все информанты выдавали *ɕt^{hi}*. Поправка: заменяем *hək^{hi}* на *ɕt^{hi}* без замены номера.

8. 'Fat' в списке куллуи А. И. Когана отсутствует. В нашем списке персизм *tsərvī*. Поправка: вносим *tsərvī* с отрицательным номером.

9. 'Fish' KUL *machli* №28 (Kogan 2016: 244): в нашем списке не представлено. Все информанты дали *mṭsi*. Это же слово находим и у М. Р. Тхакура. Поправка: заменяем *machli* на *mṭsi* без замены номера.

10. 'Foot' KUL *per* (Kogan 2016: 244): в нашем списке не представлено. Три информанта дали *dzō:ng* (Тёрнер 5082 *jānḡhā*) и три — *ʃā:ŋ* (Turner 1969–1985: №5428 *ʃāṅka*). Поправка: заменяем *per* на *dzō:ng* №130 и *ʃā:ŋ* №131.

11. 'Good' KUL *bhala* №184 и *khəra* №554 (Kogan 2016: 245): в нашем списке не представлены. Четыре информанта дали *ʃob^hla* (вероятно, Turner 1969–1985: №12532 *śubha*) и два — *bā:ka* (допустимы несколько гипотез, Turner 1969–1985: 11345 *varṇya*, 9145 *bandhura* и 11191 *vaṅka*). Однако в ходе полевой работы последующих лет стало ясно, что *bā:ka* можно скорее перевести как 'beautiful'. Поправка: заменяем *bhala* и *khəra* на *ʃob^hla* №588.

12. 'Head' KUL *sir* №38 (Kogan 2016: 245): в наших списках не представлено. Следует заметить, что если бы такое слово действительно входило в стословник куллуи, его следовало бы считать заимствованием из хинди, так как правильным соответствием др.-инд.

ś- в куллуи является *f*-. Следовательно, у слова *sir* (Turner 1969–1985: №12452 *śiras*) должен был бы быть отрицательный номер. Однако четверо информантов дали *munḍi*. Поправка: заменяем *sir* на *munḍi* №338.

13. ‘Heart’ KUL *dil* №4 (Kogan 2016: 246) в наших списках не представлено. Трое информантов дали *kokṛi* (этимология неясна). Поправка: заменяем *dil* на *kokṛi* №666.

14. ‘Horn’ KUL *sīngḥ* №40 (Kogan 2016: 246): в наших списках не представлено. Если бы оно входило в стословный список куллуи, ему следовало бы присвоить отрицательный номер (см. пункт 12 ‘head’ KUL *sir*). Однако все наши информанты дали *fi:ng*. Поправка: заменяем *sīngḥ* на *fi:ng* без замены номера.

15. ‘Kill’ KUL *mār-* №42 (Kogan 2016: 246) дал один из наших информантов. Четыре человека дали эквивалент *māka:-* (по-видимому, Turner 1969–1985: №10263 **mṛkṣa* ‘damaged’). Поправка: добавляем *māka:-* №444.

17. ‘Leaf’ KUL *ṛottrə* №45 (Kogan 2016: 247): в наших списках отсутствует. Все информанты выдали форму *ṛotṣa*. Поправка: заменяем *ṛottrə* на *ṛotṣa* без замены номера.

18. ‘Meat’ KUL *mās* №124 (Kogan 2016: 248): дано одним из информантов. Четыре информанта дали форму *fik^ha*. По всей видимости, это персизм (из перс. *sekār* ‘охота, дичь’), но фонетические изменения не совсем понятны. Поэтому лучше внести его в список на случай дальнейшей работы с ним, если всё же будет найдена исконная этимология. Поправка: добавляем *fik^ha* №-1.

19. ‘Moon’ KUL *joth* №340 (Kogan 2016: 248): слово, полученное от двух информантов. Однако от двух других получено *ḥa:nd* (Turner 1969–1985: №4661 *candrā*). Кроме того, согласно описанию М. Р. Тхакура и нашим полевым данным, в языке куллуи наблюдается противопоставление двух рядов аффрикат, и слово из списка А. И. Когана точнее было бы передавать как *dzot^h*. Дальнейшее изучение вопроса о происхождении двух рядов аффрикат в куллуи, возможно, приведёт к тому, что слово *ḥa:nd* мы сочтём заимствованием из хинди, но на данный момент мы не можем с уверенностью это утверждать. Поправка: заменяем *joth* на *dzot^h*, добавляем *ḥa:nd* №52.

20. ‘Mountain’ KUL *pahāḍ* №53 (Kogan 2016: 248): в нашем списке отсутствует. От трёх информантов было получено слово *d^ho:g* (Turner 1969–1985: №5603 **ḍhōkka*). Однако в качестве стимула выступало не *mountain*, а *hill*. По нашим данным, для понятия ‘mountain’ чаще используется слово *dzot* (Turner 1969–1985: №5362 *jhāṭa* ‘forest’). Поправка: заменяем *pahāḍ* на *dzot* №632.

21. ‘Neck’ KUL *kyāḍi* №228 (Kogan 2016: 249): не подтверждается нашими списками. Все информанты дали слово *mut^hu*, этимология которого неясна. *gḍḷa* неоднократно встречается в форме *gḍḷ*, значит в том числе и ‘горло’ (см. пункт 4 из раздела «Замечания к семантике»); но чтобы исключить значение ‘шея’, требуется более подробное исследование семантики. Поправка: заменяем *kyāḍi* на *mut^hu* №777.

22. ‘Sand’ KUL *ret* №67 (Kogan 2016: 251): действительно, было подтверждено несколькими информантами. Однако у двоих также встретилось *ballu*. Поправка: вносим *ballu* №68.

23. ‘Small’ KUL *choṭṭa* №77 (Kogan 2016: 252): не подтверждается нашими списками. Все информанты дали *ots^ha* (Turner 1969–1985: №2540 **ōccha*). Поправка: заменяем *choṭṭa* на *ots^ha* №888.

24. ‘Tail’ KUL *phunjiṭ* №84 (Kogan 2016: 254): не подтверждается нашими списками. Все информанты дали *li:ngḍa* или *li:ngḍi* (Turner 1969–1985: №11009 *lāṅgūlā*). Поправка: заменяем *phunjiṭ* на *li:ngḍa* №317.

Итого в разделе «Поправки к списку куллуи» 19 изменений, влияющих на подсчёты.

Выводы

Данные, представленные в статье А. И. Когана, безусловно, являются важнейшей базой для построения новой классификации индоарийских языков. Недочеты, перечисленные нами, лишь указывают, в каком направлении её следует продолжать. Это, прежде всего, уточнение семантики и исторической фонетики (главным образом, с целью выявления санскритизмов), что актуально для всех языков. Для больших языков также необходима проверка лексики по корпусам и словарям, а для малых — сбор достоверных полевых данных. В следующей статье мы надеемся сопоставить списки А. И. Когана с полевыми списками кумаони, марвари и чхаттисгархи, собранными Е. А. Ренковской, а также проанализировать список непали. Легко заметить, что количество влияющих на подсчёты изменений, полученных методом анализа по корпусам хорошо изученных языков, не очень велико — в каждом списке около восьми. Однако для малоизученного куллуи эта цифра достигла 19; и наиболее массовыми оказались изменения, связанные с общим соответствием базисной лексики тому, что мы знаем об исторической фонетике индоарийских языков. Общее число значимых для подсчёта изменений — 149. Посмотрим, какой эффект они дали.

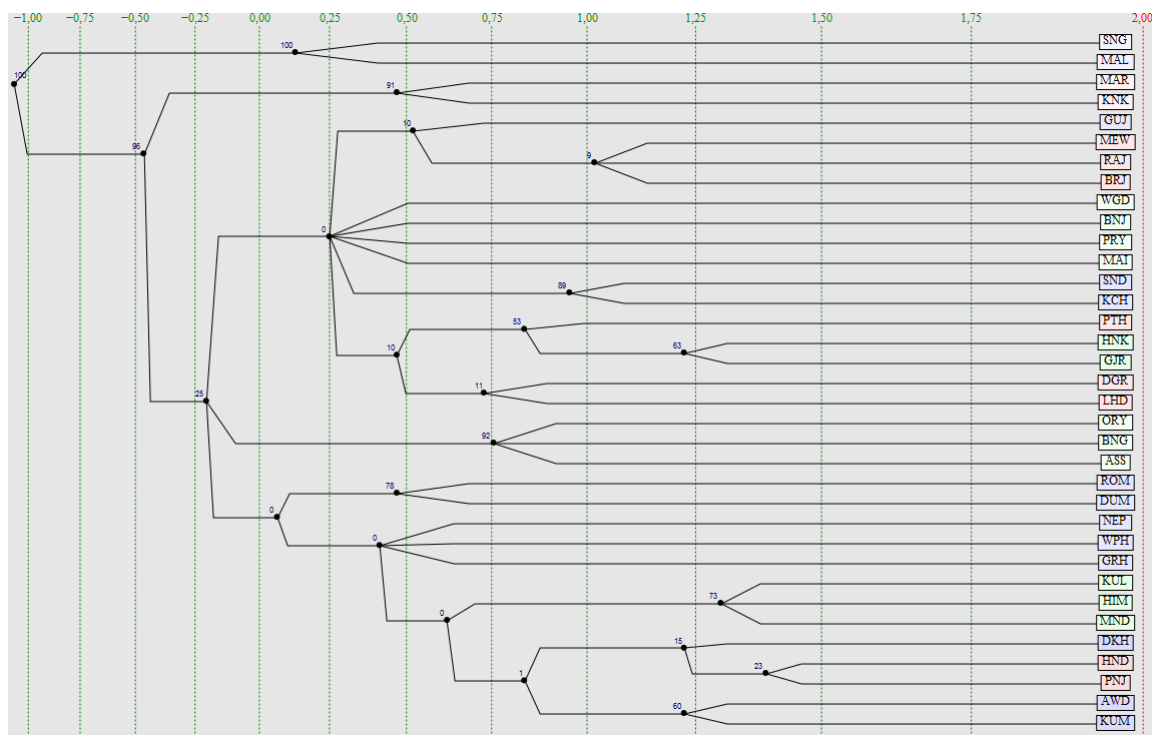


Рис. 1. Дерево, построенное по спискам А. И. Когана в Starling (бутстреп 200 раз)¹².

¹² WPH здесь и далее обозначает то же, что в статье А. И. Когана — KOT, отличие связано с более старой версией базы.

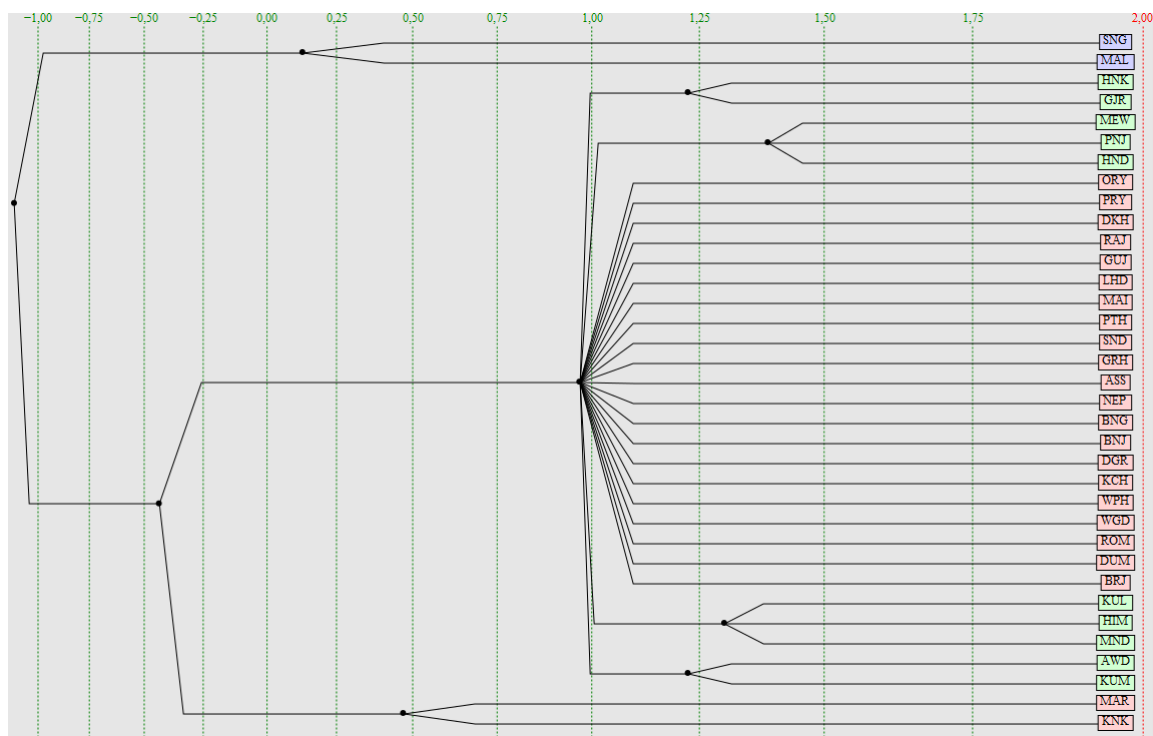


Рис. 2. Дерево, построенное по спискам А.И. Когана при сведении узлов, разница между которыми составляет менее 200 лет (при настройках по умолчанию сводятся узлы, разница между которыми составляет менее 150 лет).

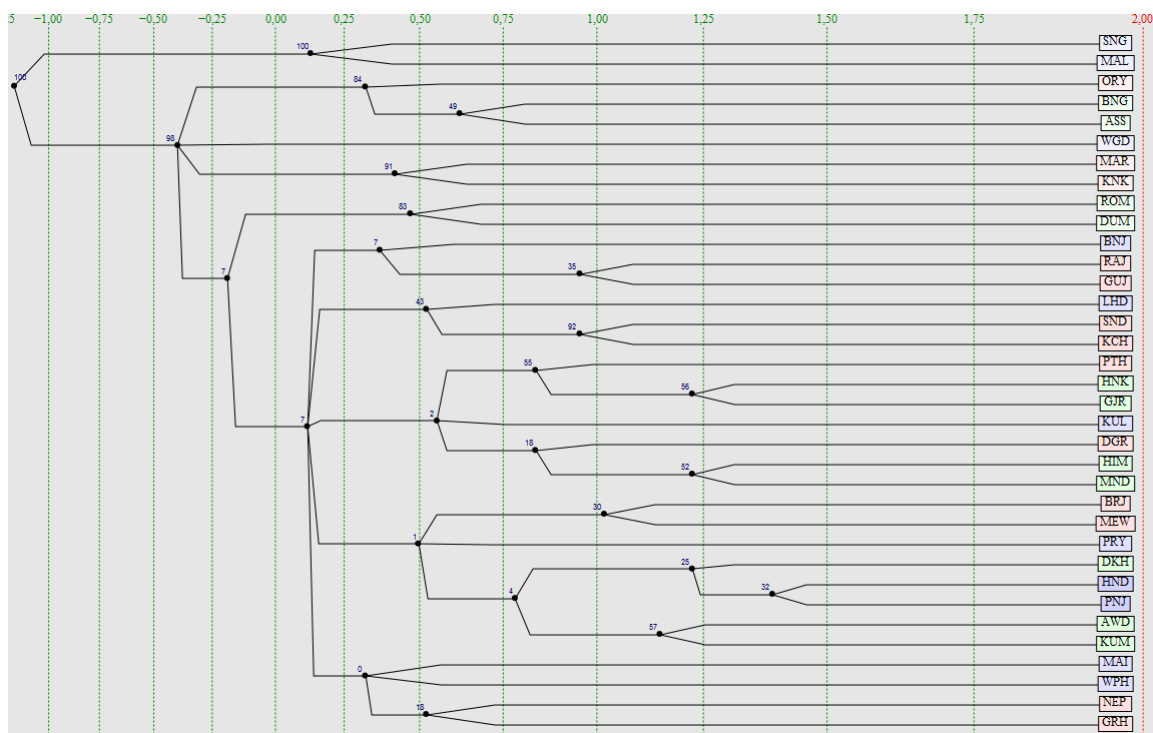


Рис. 3. Дерево, построенное по исправленным спискам (бутстреп 200 раз).

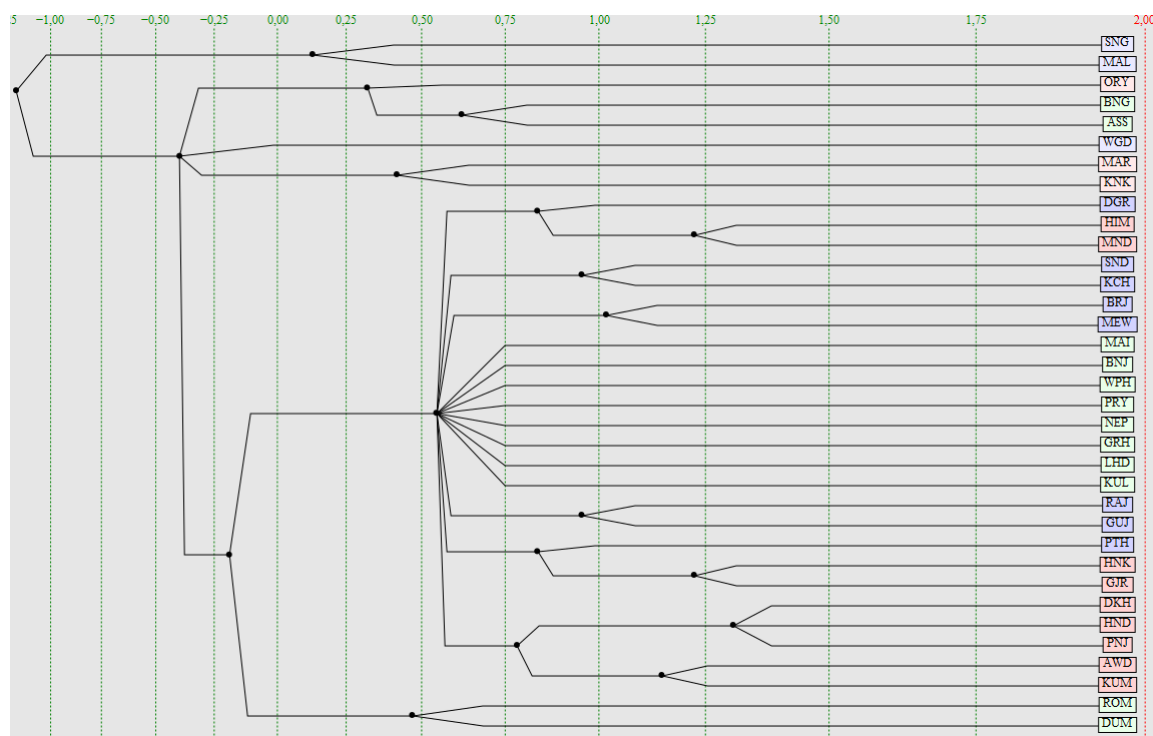


Рис. 4. Дерево, построенное по исправленным спискам при сведении узлов, разница между которыми составляет менее 200 лет.

Мы видим, что дерево стало более склонным к бинарному ветвлению; увеличилось расстояние между узлами, почти не осталось узлов с оценкой менее 1. В целом дерево стало больше соответствовать географическому разделению. Первый распад и в первоначальном, и в исправленном дереве делит языки на островную и континентальную ветви. В дереве А. И. Когана (Рис. 1) второй уровень — отделение южной группы маратхи и конкани, а затем следует распад на три ветви, из которых одна — восточная группа (ассамский, бенгальский, ория), а две других географически друг от друга неотделимы. При слиянии близко расположенных узлов (Рис. 2) этот последний узел, однако, сливается с более поздними, образуя нераздельный континуум из центральных и восточных языков. В обновлённом дереве (Рис. 3) на втором уровне разделяются южная, восточная и центральная группы. Раннее отделение вагди на том же уровне, видимо, связано с недостаточной изученностью этого идиома. На следующем уровне отделяются языки диаспор — цыганский и думаки, а затем уже следует недифференцированная географически центральная группа. Очевидно положительным изменением выглядит распад восточной ветви на ория и ассамско-бенгальскую. При слиянии близко расположенных узлов (Рис. 4) сохраняются не только восточная и южная ветви, но и ветвь «цыганский-думаки».

Следует понимать, что предложенное дерево ни в коем случае не может считаться окончательным результатом, так как проверка списков далека от завершения. Это лишь иллюстрация тех изменений, которые может повлечь за собой такая проверка.

Очень важно, что А. И. Коган впервые предложил лексикостатистическую основу для классификации индоарийских языков, внося свежую струю в индологическую традицию. Вероятно, при дальнейшем уточнении списков, более точным будет становиться и дерево, и однажды, благодаря лексикостатистическому подходу, дополненному полевыми и корпусными исследованиями, а также уточнением исторической фонетики, в давнем споре о генеалогической классификации индоарийских языков будет наконец поставлена точка.

Сокращения

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 — кутчи; KNK — конкани; KUL — куллуи; KUM — кумаони; LHD — лахнда (мультиани); MAI — майтхили; MAL — мальдивский (дживехи); MAR — маратхи; MEW — мевати; MND — мандеали; NEP — непали; ORY — ория; PNJ — пенджаби; PTH — поттохари; PRY — парья; RAJ — раджа-стави (марвари); ROM — романи; SND — синдхи; SNG — сингальский; WGD — вади; WPH — котгархи; др.-инд. — древнеиндийский

Литература

- Бархударов А. С., Бескровный В. М., Зограф Г. А., Липеровский В. П. (под редакцией Бескровного В. М.). 1972. *Хинди-русский словарь в двух томах*. М.: Советская энциклопедия.
- Беликов В. И. Что и как может получить лингвист из оцифрованных текстов // *Сибирский филологический журнал*. — 2016. — № 3. — С. 17–34.
- Лоскутов В. Н., Валвева Р. В., Полянский Б. Г. (под редакцией Быковой Е. М. и Бхоумика Н.). 1974. *Словарь бенгальского языка с русскими эквивалентами*. М.: Советская энциклопедия.
- Коган А. И. 2005. *Дардские языки. Генетическая характеристика*. М.: Восточная литература.
- Biswal K. 2015. *Saraswata Odia Bhasakosha (Odia-Odia-Insarai)*. Cuttak: Satvanaravan Book Store.
- Burrow T., Emeneau M. B. 1984. *A Dravidian etymological dictionary*. 2nd ed. Oxford [Oxfordshire]: Clarendon Press (<http://dsal.uchicago.edu/dictionaries/burrow/>).
- Callewaert W. M. 2009. *Dictionary of Bhakti. North-Indian Bhakti Texts into Kharī Bolī Hindī and English*. 3 v. New Delhi: D. K. Printworld (P) Ltd.
- Ghosh G., Ray S. (revised by Dasgupta B. M.) 2011. *Samsad Student's Bengali-English Dictionary*. Kolkata: Sahitya samsad.
- Goldhahn G., Eckart T., Ouasthoff U.: Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: *Proceedings of the 8th International Language Resources and Evaluation (LREC'12)*, 2012 (<http://wortschatz.uni-leipzig.de/en/download/>)
- Hendriksen H. 1986. *Himachali Studies. III. Grammar*. København: Munksgaard.
- Kogan A. 2016. *Genealogical classification of New Indo-Aryan languages and lexicostatistics*. In: *Journal of language relationship* 14/3-4: 227-258.
- Kumar A., Kumar K. 1997. *Samantar kosha (Hindi Thesaurus)*. 2 v. New Delhi: National Book Store.
- 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 C.P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Platts, J.T. 1884. *A dictionary of Urdu, classical Hindi, and English*. L.: W. H. Allen & Co.
- Praharaj, G.C. 1931–1940. *Purnnacandra odia bhashakosha*. Cuttack: Utkal Sahitya Press. (<http://dsal.uchicago.edu/dictionaries/praharaj/>)
- Ranganatha M. R. 1980. *Survey of Mandeali and Kului in Himachal Pradesh*. Census of India. 1971. Language monograph No. 7. New Delhi: Office of the Registrar General, India, Language Division.
- Thakur M.R. 1975. *Pahari bhasha kului ke vishesh sandarbh men*. Dehli: Sanmarg Prakashan. [Язык пахари с приложением описания куллуи].
- Thakur M.R. 2012. *Himachali*. New Delhi. Sahitva Academv.
- Tripathy B.K., Patnaik K.M. 2015. *Oxford English-English-Odia dictionary*. New Delhi, Oxford University Press.
- Turner, R.L. 1931. *A comparative and etymological dictionary of the Nepali language*. London: K. Paul, Trench, Trübner (<http://dsal.uchicago.edu/dictionaries/turner/>)
- Turner, R. L. 1962–1985. *A comparative dictionary of Indo-Aryan languages*. L.: Oxford University Press (<http://dsal.uchicago.edu/dictionaries/soas/>)
- <http://www.english-banola.com/>
- <https://the.sketchengine.co.uk/> – Hindi Web 2013 (hiTenTen13), Bengali Web (BengaliWaC)
- <http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl>

Anastasiya Krylova. Lexicostatistics and the New Indo-Aryan languages: a field linguist's perspective

In this paper, I present certain comments, objections, and addenda to Anton Kogan's article «Genealogical classification of New Indo-Aryan languages and lexicostatistics», published last year in the *Journal of Language Relationship* (14/4: 227–258). In particular, I introduce several tentative corrections to the Swadesh lists that were compiled by Kogan, based on the data of historical phonetics and semantics of the Indo-Aryan languages in general. I also analyze in detail the lists for four languages (Hindi, Odia, Bengali, and Kullui), comparing them with my own fieldwork data as well as data from other dictionaries and text corpora. Upon correcting the lists, I compare the resulting genealogical trees (generated by StarLing software) based on original and corrected lexicostatistical matrices. Although the differences are not highly significant, they nevertheless improve upon the validity of the results and demonstrate that further correction of the lists can increase the resulting tree's degree of accuracy.

Keywords: lexicostatistics, Indo-Aryan languages, language classification, glottochronology, field linguistics, corpus linguistics, historical phonetics