

## О ДОКАЗАТЕЛЬСТВЕ ЯЗЫКОВОГО РОДСТВА

Основу сравнительного языкознания составляет понятие языкового родства, без которого невозможны генетическая классификация языков, реконструкция праязыков и почти любые исследования истории языков. Термин «языковое родство» при этом, как правило, не эксплицируется, а принимается как данность.

В большинстве теоретических работ так или иначе намечается определение языкового родства. Ср. у Лемана: “The comparison of various languages led to the assumption that some languages are related, that they developed from a common source” [Lehmann 1992: 6]; у Траска (на примере романских языков): “We say that the Romance languages are genetically related, which means that they all started out as nothing more than regional dialects of a single ancestral language” [Trask 1996: 179]; у Рулена: “genetic classifications... trace the genealogy of a language family from the parent language to the daughter languages” [Ruhlen 1987: 2] и т. д.

Из подобного определения естественным образом вытекает концепция генеалогического древа и языковой дивергенции как основной модели исторического развития языка. Наиболее принята при этом модель генеалогического древа как классического одновершинного графа, в котором два и более узлов могут иметь единого предка, но никакой узел не может иметь более чем одного предка. Хотя эта модель и оспаривалась в ряде работ, она, несомненно, является общепризнанной, и абсолютное большинство лингвистических классификаций построено именно таким образом. О ее правомерности мы еще поговорим ниже.

В недавнем курсе лингвистики А. Я. Шайкевича проводится различие между онтологическим подходом к языковому родству

(«родственные языки — это разные временные и пространственные варианты одной и той же непрерывной лингвистической традиции» [Шайкевич 1995: 198]) и гносеологическим подходом (т. е. доказательством языкового родства). Однако как в этой работе, так и практически во всех других, посвященных данной теме, не анализируется подробно процедура доказательства языкового родства.

Как правило, при доказательстве языкового родства апеллируют либо к интуитивной очевидности такового (ср. у Мейе: “Le hittite, le ‘tokharien’, le sanskrit, le perse, le grec, le latin, l’irlandais, le gotique, le lituanien, le vieux-slave, l’arménien présentent dans leur grammaire et leur vocabulaire des concordances évidentes...” [Meillet 1907: 13], либо вообще к неясному представлению о «сходстве» [Ruhlen 1987: 12–14]. В отличие от последнего автора, бóльшая часть исследователей все же пытается формализовать представление о языковом родстве, апеллируя при этом к основному достижению сравнительного языкознания — сравнительно-историческому методу. Считается, что родственными могут быть только такие языки, между системами которых устанавливаются регулярные соответствия. При этом, однако, хорошо известны примеры наличия регулярных соответствий, возникающих в результате массовых заимствований: ср. китайский и японский, латинские заимствования в европейских языках и т. п.

Остается констатировать, что в современных работах по компаративистике отсутствует как точное определение родства языков, так и процедура доказательства такого родства. Фактически проблема родства языковых семей в каждом отдельном случае решается голосованием: славянская семья, индоевропейская семья, картвельская семья и т. д. признаются существующими, поскольку существует согласие подавляющего большинства специалистов в этих областях. В существовании алтайской семьи высказываются сомнения, поскольку не все специалисты по тюркским, монгольским и тунгусо-маньчжурским языкам согласны с алтайской теорией. В то же время, например, гораздо слабее разработанная теория о единстве афразийских языков является фактически общепризнанной. В последнее время снова активизировались дебаты по поводу существования макросемей — прежде всего, так называемой ностратической семьи языков.

Заметим, что в подобного рода дискуссиях важную роль играет проблема глубины языкового родства — хотя это редко эксплицируется. Единство славянских, романских, тюркских языков очевидно для самих носителей этих языков, поэтому оно практически никогда не оспаривается в академических кругах. Связь между отдельными

ветвями индоевропейских языков гораздо отдаленнее. Она неочевидна для носителей, но убедительна для лингвистов, знакомых хотя бы с несколькими индоевропейскими языками и с предъявляемыми в индоевропеистике доказательствами. Наконец, для оценки родства внутри макросемей типа ностратической большинству специалистов просто не хватает знаний по отдельным семьям — что влечет за собой естественный скепсис и обвинения в «панкомпаративизме».

В настоящий момент дискуссии по вопросу т. н. «отдаленного родства» начинают играть все большую роль в компаративистике. Успешное развитие и применение сравнительно-исторического метода привели к тому, что подавляющее большинство таксономических единиц, не превышающих индоевропейскую семью по глубине, уже выявлено, и попытки углубления сравнений представляются вполне закономерными. Однако макрокомпаративистика имеет свои особенности. Сравнимые единицы объективно находятся на гораздо более далеком расстоянии друг от друга, чем в семьях типа славянской и индоевропейской. Сходство затемнено тысячелетиями независимого развития и не является очевидным даже для специалистов. Таким образом, вопрос о критериях языкового родства и способах его доказательства, ранее, возможно, не нуждавшийся в специальной экспликации, теперь выходит на первый план.

Будем исходить из следующего общего определения родства языков (онтологического, по А. Я. Шайкевичу):

*Два языка (или более) являются родственными, если они восходят к единому источнику (= праязыку).*

Таким образом, необходимым и достаточным условием онтологического доказательства родства языков является предъявление источника этого родства — праязыка.

К сожалению, ситуаций сохранения праязыка в письменных памятниках и четко прослеженной истории его развития в современный язык (языки) известно не так много. В большинстве случаев языковое наследование и родство нуждается в обосновании, а следовательно, необходимо процессуальное (гносеологическое, по А. Я. Шайкевичу) определение языкового родства.

В каком случае язык А может считаться праязыком для языка В? В идеальном случае хотелось бы, чтобы язык-потомок полностью выводился из своего предка, т. е., чтобы существовала система правил, при помощи которых:

- а) все фонемы языка В выводятся из фонем языка А;
- б) все морфемы языка В выводятся из морфем языка А;

в) все лексемы языка В выводятся из лексем языка А;

г) все грамматические и синтаксические правила языка В выводятся из соответствующих правил языка А.

На практике, однако, выполняется только первое (фонологическое условие). Второе и третье условия выполняются только на множестве морфем и лексем, сохранившихся при переходе от А к В. Часть морфем и лексем неизбежно теряется, а взамен часто возникают новые, например, в результате разнообразных заимствований. В случае с грамматическими и синтаксическими правилами могут наблюдаться различные случаи: от предельного консерватизма (например, в истории литовского или исландского языков) до радикальнейших перемен (например, в истории английского или болгарского языков). Учитывая всю важность грамматической структуры для языковой истории и языковой классификации, я все же не рискнул бы строить определение языкового родства на грамматических критериях — и прежде всего потому, что во многих вполне очевидных случаях языкового родства грамматические критерии просто отсутствуют. Современный китайский язык (его пекинский диалект), по-видимому, не сохранил ни одной грамматической морфемы древнекитайского языка. Можно констатировать, что за период непрерывного развития в три тысячи лет грамматическая система китайского языка полностью перестроилась (за исключением разве что основных принципов порядка слов). И сходных примеров можно привести множество — от современных новоиндийских языков до современного английского языка.

Мы будем поэтому считать, что язык А является предком языка В, если

а) все фонемы языка В выводятся из фонем языка А;

б) некоторая часть морфем (в том числе грамматических) и лексем языка В выводятся из морфем и лексем языка А.

Существенным здесь является вопрос о том, какова именно должна быть эта «некоторая часть». Ясно, что речь должна идти о так называемой «базисной» лексике языка в противовес культурной лексике, часто заимствуемой из языка в язык. Определение «базисной» лексики мы давать не будем, а зададим ее списком, совпадающим со стандартным 100-словным списком М. Сводеша [Сводеш 1960]<sup>1</sup>. В принципе можно было бы предлагать и другие списки, однако список Сводеша уже вполне хорошо зарекомендовал себя на практике. Легко

<sup>1</sup> Относительно состава стандартного 100-слового списка Сводеша см. [Старостин 1989а].

показать, что, скажем, любой романский, германский или славянский язык в своем 100-словном списке сохраняет около 90% слов, бытовавших в этом списке в прароманском (= латинском), прагерманском и праславянском языках соответственно. Любой современный индоевропейский язык в этом списке сохраняет порядка 50% слов, реконструируемых для гипотетического индоевропейского праязыка, и т. п.

Уже на последнем примере ясно, что точное количество морфем (лексем), сохранившихся при переходе от праязыка к языку-потомку, задано быть не может. Это величина, связанная функциональной зависимостью со временем, разделяющим моменты фиксации соответствующих языков. Существуют, таким образом, различные степени родства (что, впрочем, интуитивно очевидно). Способы измерения лингвистического времени — отдельная проблема, выходящая за рамки данной статьи (см. [Старостин 1989а]). Для нас сейчас важен сам факт высокой сохранности базисной лексики.

Возможна ли ситуация, при которой базисная лексика языка за короткий период заменяется на другую? Такие случаи известны, и чаще всего классифицируются как смена языка. Несомненно, что базисная кельтская лексика в первые века нашей эры сменилась латинской на территории современных Франции и Испании. Характерно, однако, что мы при этом называем французский и испанский языки потомками латыни, но не местных кельтских языков. Дело в том, что если уж начинает активно заимствоваться базисная лексика, то остальная лексика языка, как правило, проникается заимствованиями в еще большей мере. Этот процесс сопровождается и активным заимствованием грамматической системы, в результате чего язык фактически меняет свою генетическую принадлежность.

Случаи отдельных — и даже довольно многочисленных — заимствований в базисной лексике также известны. Ясно, однако, что число заимствований не должно превышать число исконно сохранившихся лексем (иначе, как сказано выше, фактически происходит смена генетической принадлежности языка). Предположим, однако, такую гипотетическую ситуацию:

1) существуют два засвидетельствованных (или реконструированных) древних языка, А и В;

2) существует современный язык С, в базисной лексике которого насчитывается около 50% слов из каждого из упомянутых выше языков.

Очевидно, что один из древних языков является предком, а другой — донором для языка С. Но поскольку как в случае родства, так и в случае массовых заимствований наблюдаются регулярные фонетические

соответствия, то мы не в состоянии выяснить, какой из языков А и В является предком языка С, а какой — донором.

Решение этой дилеммы было предложено С. Е. Яхонтовым, и оно основывается на факте разной степени устойчивости базисной лексики (к сожалению, сама эта работа до сих пор не опубликована; опыт использования этой методики см., например, в [АПиПЯЯ]). С. Е. Яхонтов предложил разбиение 100-словного списка на максимально устойчивую 35-словную часть и менее устойчивую 65-словную<sup>2</sup>. Предполагается, что в случае исконного родства процент совпадений в 35-словной части списка должен быть выше, чем процент совпадений в 65-словной части списка. В приведенном выше случае (заметим, сугубо гипотетическом — мне неизвестны языки с 50% заимствований в базисной лексике), таким образом, предком окажется тот язык, чья лексика в большей мере представлена в 35-словном списке языка С.

Отсюда, в частности, следует вывод о невозможности «смешанных» языков в рамках классической модели генеалогического древа. Любой язык может — в том, что касается его базисных компонентов — иметь только одного предка.

Итак, дадим полное гносеологическое определение языкового наследования:

Язык А является предком языка В, если:

- а) все фонемы языка В выводятся из фонем языка А;
- б) эти фонетические правила действуют на множестве базисной лексики языка В;
- в) доля сохранившейся лексики возрастает, если берется выборка из более устойчивой лексики.

Из этого определения естественным образом вытекает определение языкового родства. Если имеются языки В и С, восходящие к языку А, то в силу условия а) все фонемы языков В и С будут выводиться из фонем языка А, т. е. соответствовать друг другу. Эти соответствия будут

<sup>2</sup> С. Е. Яхонтов использует модифицированный список Сводеша, из которого исключены значения *all, burn, bark, bite, claw (nail), feather, lie, seed, warm, we*, вместо которых используются: *far, heavy, near, salt, short, snake, thin, wind, worm, year*. В его 35-словный список включены следующие значения: *blood, bone, die, dog, ear, egg, eye, fire, fish, full, give, hand, horn, I, know, louse, moon, name, new, nose, one, salt, stone, sun, tail, this, thou, tongue, tooth, two, water, what, who, wind, year*.

В реальной практике (в частности, компьютерной) приходится считаться как с классическим списком (который уже обработан для очень большого числа языков), так и со списком Яхонтова. Поэтому, например, моя компьютерная система STARLING требует для каждого языка заполнения 110-словной анкеты.

соблюдаться на множестве базисной лексики языков В и С (восходящей к базисной лексике языка А), и доля общей лексики будет возрастать по мере возрастания устойчивости сравниваемой лексической выборки. Итак, языки В и С родственны друг другу, если:

а) все фонемы языков В и С регулярно соответствуют друг другу<sup>3</sup>;

<sup>3</sup> Вопрос о непогрешимости фонетических соответствий давно обсуждается в сравнительном языкознании. Допустимы ли исключения из фонетических правил, и если да, то не подрывает ли это сформулированного выше пункта а)? На практике всякому компаративисту известны случаи отклонений от регулярных соответствий — объясняемых аналогиями, эвфемистическими, табуистическими или оноματοпоэтическими причинами, а зачастую никак не объясняемых (причина нерегулярности вполне может быть не установлена). Имеет смысл поэтому говорить не о полностью регулярных, но о систематических соответствиях, т. е. о соответствиях, регулярно связывающих каждую фонему языка А с одной или несколькими фонемами языка В. Система соответствий должна оцениваться не по принципу абсолютного отсутствия исключений (хотя такое, конечно, чрезвычайно желательно), но по принципу улучшения: хороша такая система, которую можно улучшить, не меняя множества лексики, на котором эта система работает. Классический пример такого улучшения — закон Вернера, который показал, что казавшееся немотивированным расщепление индоевропейских глухих согласных в германском однозначно связано с местом ударения в древнеиндийском и древнегреческом, и, таким образом, уменьшил степень неопределенности в системе индоевропейских фонетических соответствий.

Особый вопрос — объективность фонетических соответствий, их независимость от пристрастий конкретного исследователя. Пока не предложена идеальная формальная процедура установления фонетических соответствий. Тем не менее некоторые алгоритмы все же существуют. Один из них реализован в моей компьютерной системе STARLING, которая позволяет довольно успешно устанавливать основной набор фонетических соответствий между любыми двумя родственными языками на материале переводных списков. Суть алгоритма в следующем:

а) вычисляется частота каждой фонемы в списке каждого из сравниваемых языков; б) в языке А берется подмножество слов N, содержащих фонему x; в) в языке В берется подмножество слов N', являющихся переводом слов подмножества N языка А; г) вычисляется частота каждой фонемы языка В в подмножестве N' и сравнивается со стандартной частотой данной фонемы во всем списке слов языка В; д) фонема x', частота которой в подмножестве N' существенным образом (вычисляемым, например, по правилу «трех сигм») превышает стандартную частоту данной фонемы, считается соответствующей фонеме x языка А.

Заметим, что при такой процедуре одной фонеме x может в принципе соответствовать несколько фонем в языке В, что не противоречит реальной компаративистической практике.

Эта процедура позволяет устанавливать около 80% правильных фонетических соответствий и почти не дает ложных соответствий. Точность соответствий естественно возрастает при увеличении исходного массива данных.

Ценность этого компьютерного алгоритма состоит не столько в компьютеризации работы компаративиста (этот алгоритм не может определить дополнительных

б) эти фонетические соответствия действуют на множестве базисной лексики языков В и С;

в) доля общей лексики<sup>4</sup> возрастает, если берется выборка из более устойчивой лексики<sup>5</sup>.

Заметим, что такое определение родства не требует непосредственного предъявления исходного языка А. Реально в подавляющем большинстве случаев предок А оказывается незасвидетельствованным.

распределений, не срабатывает в случае выпадения фонем, т. е. соответствия некоторой фонемы в одном языке нулю в другом, не позволяет установить многих редких, но существенных, соответствий), сколько в доказательстве объективности факта фонетических соответствий вообще и подтверждения выводов, эмпирически полученных исследователем.

<sup>4</sup> Когда я говорю об «общей» лексике или «сохранившейся» лексике, я имею в виду лексику, удовлетворяющую семантическим критериям стандартной лексикостатистики, т. е. слова, имеющие общую (соответствующую друг другу) форму и единое значение. Любая мена значения автоматически выводит слово из сравнения. Подчеркну, что это вопрос чисто процессуальный: ясно, что слова могут менять значение, но также ясно, что если при определении родства мы допустим неточные семантические сравнения, вероятность ошибки и погрешности чрезвычайно возрастает — поскольку очень трудно формально отделить допустимый семантический сдвиг от маловероятного или вовсе невероятного. При установлении фонетических соответствий и при составлении этимологических словарей семантически неоднозначные параллели, конечно, допустимы (и могут оцениваться или оспариваться специалистами); но при тестировании родства на множестве базисной лексики следует считать только случаи одно-однозначного семантического соответствия. Существенно здесь не то, что пол. *brzuch* имеет хорошее соответствие в рус. *брюхо*, но то, что 'живот' по-польски передается как *brzuch*, а по-русски — как *живот*, т. е. словом другого происхождения.

<sup>5</sup> Ясно, что пункты б) и в) можно переформулировать более общим образом — поскольку 35-словный список является более устойчивым подмножеством 100-словного, а 100-словный — более устойчивым подмножеством всей лексики языка. В принципе можно сказать:

б) эти фонетические соответствия действуют на некотором множестве лексики языков В и С;

в) доля общей лексики возрастает, если берется выборка из более устойчивой лексики.

Чрезвычайно существенно, однако, оперировать с некоторым фиксированным списком (списками). Ясно, что для любой пары языков, обнаруживающих сходства, можно набрать свой список, доля сходжений в котором будет чрезвычайно велика, и все критерии родства окажутся в таком случае размыты. Поэтому попытки представить различные списки базисной лексики для различных, к примеру, географических ареалов не представляются плодотворными. Список Сводеша уже хорошо себя зарекомендовал на материале самых различных языковых семей, и кажется целесообразным сохранить его в качестве стандарта тестирования языкового родства.



Случаи типа латинского и его потомков — романских языков — весьма показательны для теории, но редки в компаративистической практике, из-за общей скудости древних письменных источников.

Сформулированное выше определение языкового родства в принципе не зависит от числа сравниваемых языков и от глубины родства. Ясно, тем не менее, что при очень малых долях совпадений (т. е. при очень дальнем родстве) возможность установления регулярных соответствий (пункт а) становится все более сомнительной. В литературе часто фигурирует цифра 8–10 тысяч лет как абсолютный предел возможности установления языкового родства (см., например, [Trask 1996: 377]), хотя обычно не говорится, откуда эта цифра берется. На самом деле она восходит к стандартной глоттохронологической формуле Сводеша, согласно которой за 10 000 лет в двух родственных языках должно сохраниться всего 5–6% общей лексики — ситуация, при которой сравнение и реконструкция уже не могут дать положительных результатов, и нельзя отличить исконно родственные морфемы от случайных совпадений (при расхождении порядка 16 000 лет по формуле Сводеша языки вообще должны утратить абсолютно все сходство).

Это простое соображение реально лежит в основе всей критики существующих теорий дальнего языкового родства и, на первый взгляд, действительно, кажется непреодолимым. Из него вытекают активно предлагаемые в настоящее время альтернативные модели генетических классификаций (см., например, [Nichols 1992]), выдвигающие на первый план не генетические, а типологические критерии. Некоторые авторы, вообще разочаровавшиеся в лексикостатистике и сравнительно-историческом методе, предлагают новые модели исторического развития языков (такова, например, теория *punctuated equilibrium* Диксона, который считает, что язык может вообще не развиваться в течение долгих промежутков времени, а затем скачкообразно начать развиваться и разделяться, см. [Dixon 1997])<sup>6</sup>.

Решение задачи и на этот раз предоставляет нам классический сравнительно-исторический метод. Наличие регулярных соответствий

---

<sup>6</sup> Последняя теория представляется совершенно абсурдной с точки зрения классического сравнительно-исторического метода. Нет ни одного засвидетельствованного в памятниках языка, который не претерпел бы существенных изменений хотя бы за 1000 лет своего развития. Необходимость языковых изменений с очевидностью вытекает из самой сущности языка как средства коммуникации: информационный код не может не искажаться при передаче из поколения в поколение, так же, как не может не искажаться сигнал в телефонной линии или генетический код. Но подробное обсуждение этой темы выходит за рамки настоящей статьи.

между языками позволяет интерпретировать каждый ряд таких соответствий как отражение некоторой фонемы праязыка. В результате оказывается возможным реконструировать систему фонем праязыка, а следовательно, и набор морфем (лексем), состоявших из этих фонем. Здесь не место описывать подробно процедуру реконструкции, опробованную сейчас уже на десятках (если не сотнях) языковых семей (см., например, [Birnbauм 1978] с библиографией). Существенно, однако, что почти в каждом случае реконструкции мы в состоянии довольно полно реконструировать список базисной лексики соответствующего праязыка. Как для близких (типа славянской), так и для более далеких (типа индоевропейской) семей удастся восстановить практически полный 100-словный список Сводеша. Здесь и лежит очевидный выход из, казалось бы, тупиковой ситуации: можно сравнивать не только современные языки, но и реконструированные праязыки, которые — в случае родства — должны обнаруживать между собой гораздо больше совпадений, чем их современные потомки. К праязыкам при этом предъявляются ровно те же требования, что и к современным языкам.

Приведем пример из области весьма далеко родственных друг с другом языков. Учитывая сведения из исторической фонетики индоевропейских и алтайских языков, а также систему соответствий между ностратическими языками, предложенную В. М. Иллич-Свитычем, проанализируем совпадения в области базисной лексики между русским и эвенкийским языками<sup>7</sup>.

В 35-словном списке С. Е. Яхонтова обнаруживается пять совпадений — исключительно местоимения и числительные:

Русский	Эвенкийский
я (< слав. *azъ < ПИЕ *eǵho-m, где исторически собственно местоименный корень — *-m, ср. косв. основу *me-)	bi < ТМ *bi < алт. *bi
э-тот (< слав. *e- < ПИЕ *e- 'этот, тот')	e-r < ТМ *e- < алт. *e

<sup>7</sup> Здесь мы используем традиционную индоевропейскую реконструкцию, а для праалтайского — реконструкцию, принятую коллективом Алтайского этимологического словаря (автор данной статьи, А. В. Дыбо и О. А. Мудрак), в настоящее время готовящегося к печати.

Большинство приводимых примеров присутствует уже в [ОСНЯ]. Мною добавлен лишь ряд примеров, обнаруживающих неучтенное В. М. Иллич-Свитычем, но довольно часто встречающееся соответствие алтайских глухих придыхательных \*k', \*t' индоевропейским звонким \*g<sup>w</sup>, \*d(w) (ср. примеры 'гора', 'звезда', 'давать', 'тяжелый').

Русский	Эвенкийский
два (< слав. *dъva < ПИЕ *d̥uō-u)	žū-r < ТМ *žube- < алт. *t̥ubu
что (< слав. *čъ-to < ПИЕ *k̥ <sup>w</sup> i-, *k̥ <sup>w</sup> e-)	ê-kūn < ТМ *xia- (*xaj-) < алт. *k'a(j)
кто (< слав. *kъ-to < ПИЕ *k̥ <sup>w</sup> e-)	ê-kūn < ТМ *xia- (*xaj-) < алт. *k'a(j)

В 65-словной части списка находим еще следующие 6 совпадений:

Русский	Эвенкийский
холодный (< слав. *xoldъ-нъ < ПИЕ *g(h)elǎ- / *kelǎ-, возм. *(s)ghelǎ-)	gil-di < ТМ *gil- < алт. *gilo
гора (< слав. *gora < ПИЕ *g <sup>w</sup> er-)	ure < ТМ *xurē < алт. *k'uri
звезда (< слав. *gъvѣzda < ПИЕ *g <sup>w</sup> eizdh-)	ōsī-кта < ТМ *xōsī-кта < алт. *k'ūčV
тот (< слав. *tъ < ПИЕ *to-)	ta-r < ТМ *ta- < алт. *t'a (*t'e)
мы (< слав. *my < ПИЕ *mo-)	mi-t < ТМ *mü-n < алт. *bū-na
пепел (< слав. *perelъ < ПИЕ *pel-)	xule-pten < ТМ *pulñe < алт. *p'olñe

Итак, совпадают 14% 35-слового списка и всего лишь 9% 65-слового (во всем списке С. Е. Яхонтова налицо 11% совпадений).

При переходе от современных языков к реконструированным праязыкам получаем следующий список совпадений:

В 35-словной части — все перечисленные выше корни плюс еще:

Значение	Индоевропейский	Алтайский
‘кровь’	*esH-r	*sēga
‘рыба’	*dhǵhū-	*d̥iōjgi
‘давать’	*dō-	*t'uja
‘имя’	*enot-η (*elom-η)	*Ц̣ǎm(η)o
‘солнце’	*seH <sup>w</sup> -n- (/ *seH <sup>w</sup> -l-)	*ṣjagu
‘огонь’	*peH <sup>w</sup> -r-	*p'ore
‘ты’	*tū	*t'i / *t'a (корень сохранился только в монг., но представляет собой явный архаизм)

В 65-словной части — все перечисленные выше корни плюс еще:

Значение	Индоевропейский	Алтайский
‘черный’	*kr̥-sno-	*k'aru (~ k-)
‘есть’	*ed-	*itV (корень сохранился только в монг., но представляет собой явный архаизм)

Значение	Индоевропейский	Алтайский
‘лист’	* <i>lep-</i> , * <i>leup-</i>	* <i>liap</i> [e]
‘мужчина’	* <i>uī-ro-</i>	* <i>iōre</i>
‘много’	* <i>menə(gh)-</i>	* <i>mana</i>
‘шея’	* <i>mono-</i>	* <i>mōjŋo</i>
‘не’	* <i>ne</i> , * <i>ŋ</i>	* <i>āna</i>
‘тяжелый’	* <i>g<sup>w</sup>era-</i>	* <i>k’iuru</i>
‘червь’	* <i>k<sup>w</sup>ŋ-mi-</i>	* <i>k’iāru</i> (~ * <i>k’iūra</i> )

Итак, в 35-словной части списка получаем 34% совпадений, а в 65-словной части — 23% (всего между ПИЕ и праалтайским в 100-словном списке С. Е. Яхонтова имеем 27% совпадений). Налицо наглядное увеличение процента совпадений при переходе от современных языков к реконструированным праязыкам, которые обнаруживают между собой приблизительно тот же процент совпадений, который выявляется при сравнении современных индоевропейских или алтайских языков. В то же время стабильно сохраняется преобладание совпадений внутри наиболее устойчивой части базисной лексики — как при сравнении современных ностратических языков, так и при сравнении их реконструированных предков.

Данная процедура позволяет отвести еще несколько претензий, предъявляемых критиками дальнего родства:

а) Соображения типа: «при дальнем сравнении невозможно или затруднительно отличить исконные совпадения от случайностей».

Заметим, что если доля совпадений при сравнении праязыков увеличивается по сравнению с долей совпадений между современными языками, это уже явное свидетельство родства: ведь если совпадения случайны, то их должно быть приблизительно одно и то же количество в любых двух сравниваемых списках.

б) Соображения типа: «реконструкции неточны, и поэтому при сравнении праязыков происходит лишь нагромождение неточностей друг на друга».

Фонетическая точность реконструкции — вещь существенная, но не абсолютная (как и, например, фонетическая точность записи современного языка). Важно иметь соответствия между реконструируемыми сущностями, и это, пожалуй, даже важнее, чем знать точную артикуляцию этих сущностей в праязыках. Между тем, вероятность ошибки в установлении соответствий уменьшается прямо пропорционально

числу сравниваемых языков. Если мы, к примеру, устанавливаем соответствие между ПИЕ  $*k$  и ПК  $*k$ , то это соответствие можно оспаривать, считая соответствующие примеры случайными (особенно если имеется еще некоторое число соответствий между ПИЕ  $*k$  и ПК  $*k$ , а также ПИЕ  $*k$  и ПК  $*j$ ). Однако если к этому добавляется еще соответствие ПК  $*k$  : ПА  $*k'$  и соответствие ПИЕ  $*k$  : ПА  $*k'$ , а также хотя бы несколько случаев тройного соответствия ПИЕ  $*k$  : ПК  $*k$  : ПА  $*k'$ , вероятность случайности существенно уменьшается. При этом можно продолжать спорить о фонетической сущности ПИЕ  $*k$  и ПА  $*k'$  — реальность соответствий от этого не уменьшается.

в) Соображения типа: «чем глубже реконструкция, тем меньше в нашем распоряжении лексики, а следовательно, тем меньше материала для сравнения».

Этот часто встречающийся аргумент попросту неверен. Легко показать, что число корней (лексем), восстанавливаемых для праязыка, зависит не только (и не столько) от его глубины, а от числа сравниваемых языков-потомков и от способа их классификации. Если в семье всего два языка, то для праязыка могут быть формально восстановлены только корни, сохранившиеся в обоих языках. При увеличении числа языков возрастает и число реконструируемых корней, поскольку шанс быть возведенным к праязыку имеет любой корень, отраженный хотя бы в паре родственных языков<sup>8</sup>.

---

<sup>8</sup> На самом деле, любой корень любого из сравниваемых языков — если нельзя продемонстрировать, что он заимствован из некоторого другого языка — имеет шанс быть возведенным к праязыку. Этот шанс вполне может быть оценен статистически: при общем числе равноудаленных языков  $n$  и при степени сохранности корня (т. е. числе языков, сохранивших его рефлекс)  $x$  вероятность возвести произвольный корень к праязыку равна  $1/(n/x)$ . Ясно поэтому, что в ситуации бинарного сравнения (при делении праязыка ровно на два языка-потомка) число реконструируемых (то есть сохранившихся в обоих языках) корней может быть очень невелико, в зависимости от глубины соответствующей семьи, однако весьма велико число потенциально реконструируемых корней: каждый корень в языке, имеющем лишь одного родственника, с вероятностью 0.5 может продолжать некоторый праязыковой корень. Эта вероятность естественно уменьшается при увеличении числа сравниваемых языков, однако при этом неизбежно возрастает число реально реконструируемых корней (естественно, если сравниваемые языки родственны).

Поэтому нет ничего удивительного в том, что, скажем, для прауральского восстанавливается меньше корней, чем для праностратического: уральская семья делится ровно на две ветви (финно-угорскую и самодийскую), и общеуральскими по определению считаются только корни, сохранившиеся в обеих ветвях — хотя, как мы показали выше, каждый, к примеру, финно-угорский корень имеет довольно высокую вероятность (0.5) быть возведенным к общеуральскому состоянию.

Мы видим, таким образом, что какие-либо теоретические препятствия на пути дальнейшего сравнения или реконструкции отсутствуют. Методика ступенчатой реконструкции, избранная основоположниками ностратического языкознания, представляется вполне эффективным решением вопроса. Реальная трудность — это резкое увеличение объема информации, которая должна быть обработана при дальнейшем сравнении языковых семей. Для успешного решения этой проблемы — помимо уникальных ученых-одиночек, способных держать в голове громадные количества корней всевозможных языковых семей, нужны два фактора:

1) Готовность к кооперации со стороны специалистов по отдельным языковым семьям. Финальным продуктом деятельности таких специалистов, несомненно, являются сравнительные грамматики и, главное, этимологические словари соответствующих семей. Желательно, чтобы эти этимологические словари могли быть использованы и специалистами по другим семьям, то есть чтобы в них был представлен по возможности полный и филологически качественно обработанный материал соответствующих семей. Должна быть, кроме того, готовность модифицировать результаты реконструкции некоторого частного праязыка в соответствии с открывающейся более глубокой генетической перспективой. Очевидно, что внешние данные весьма

---

Аналогичная ситуация налицо, к примеру, и в северокавказских языках. Число общих северокавказских корней невелико (около 700) — только в силу бинарного членения северокавказской общности (на западнокавказскую и восточнокавказскую подгруппы). При этом ясно, что реальный корневой состав общесеверокавказского языка был существенно больше, поскольку каждый отдельный восточнокавказский или западнокавказский корень имеет весьма высокий шанс быть возведенным к общесеверокавказскому состоянию.

Интересно, что в этимологически хорошо разработанных семьях с большим ветвлением (типа индоевропейской, финно-угорской, алтайской, восточнокавказской, австронезийской, синотибетской и т. п.) число реконструируемых корней приближается к некоторой константе, колеблющейся около 2–2.5 тысяч. Это число, по-видимому, не зависит от временной глубины соответствующей семьи, поскольку совершенно аналогичные цифры мы получаем для молодых семей типа славянской, германской, тюркской и т.п. Наличие подобной константы можно объяснить только тем, что реконструируемые праязыки представляли собой реальные языковые образования со средним для любого языка числом морфем. Гипотеза о том, что любой язык имеет конечное число корней, колеблющееся между 2–3 тысячами, вполне вероятна, хотя и нуждается в дополнительной типологической проверке. Так, словарь морфем современного русского языка [Кузнецова, Ефремова 1986] насчитывает 4400 корней, хотя реальное число, видимо, существенно меньше (в качестве отдельных единиц в словарь включены варианты корней типа *оберт-* при *верт-* и т. п.).

часто могут корректировать реконструкцию, полученную исключительно внутренним путем<sup>9</sup>.

2) Использование современных способов обработки информации, прежде всего, компьютерных баз данных, существенно облегчающих поиск этимологий и обработку больших массивов корней. Представляется, что компьютерные методы будут в будущем играть все большую роль в исследованиях по дальнему родству языков. Компьютерная компаративистика — отдельная тема, требующая специального рассмотрения. Здесь я ограничусь отсылкой на интернетовский сайт <http://starling.rinet.ru>, где представлено on-line большое количество баз данных, разработанных группой московских исследователей в рамках проекта «Вавилонская башня».

---

<sup>9</sup> Примеров можно привести множество. Так, реконструкция анлаутных звонких (\*d-, \*g-) в пратюркском принимается не всеми исследователями, поскольку реально они сохраняются только в одной — огузской — подгруппе; в прочих языках, начиная с древнетюркского, оппозиция глухости–звонкости в анлауте отсутствует. Внешние же (монгольские и тунгусо-маньчжурские) данные однозначно поддерживают реконструкцию звонких и таким образом позволяют уточнить пратюркскую реконструкцию. Праслав. \**noktb* ‘ночь’ реконструируется именно так (не \**notjb*) только в силу того, что известен индоевропейский прототип этого корня \**nok<sup>wt</sup>t*-. Даже реконструкция праславянских сочетаний типа \**TorT*, \**TolT* была бы вряд ли возможна чисто на славянской почве, без привлечения данных внешнего сравнения.