

СРАВНИТЕЛЬНО-ИСТОРИЧЕСКОЕ ЯЗЫКОЗНАНИЕ И ЛЕКСИКОСТАТИСТИКА

В настоящее время можно уже с полным основанием утверждать, что в последние десятилетия (60–80-е годы) произошел коренной перелом в практике сравнительно-исторических исследований. Длительный период компаративистской работы с самыми различными языковыми семьями подготовил почву для исследования генетических отношений между отдаленно родственными языками и языковыми группами, причем решающий шаг в этом направлении был сделан В. М. Иллич-Свитычем в его основополагающем «Опыте сравнения ностратических языков» — труде, где на основании строгой методики и гениальной интуиции обосновывается родство ряда языковых семей Старого Света.

Выход сравнительно-исторической лингвистики на новые рубежи (сравнение реконструированных праязыковых состояний с целью воссоздать облик более древних «макросемей») представляется полностью закономерным. Действительно, если мы принимаем в качестве аксиоматической модели компаративистики теорию языковой дивергенции, то мы вынуждены признать факт существования на территории ойкумены в период приблизительно с 6-го тысячелетия до н. э. по 1-е тысячелетие до н. э. целого ряда «праязыков», успешно поддающихся реконструкции. Только языковых семей глубины около 5–6 тысяч лет (т. е. сравнимых с индоевропейской семьей по древности дивергенции) можно насчитать, по-видимому, несколько сот. По мере совершенствования реконструкции праязыковых состояний этих семей неизбежно встают вопросы: связаны ли между собой генетически какие-либо из этих праязыков и если да, то можно ли эту связь обосновать?

На первый из этих вопросов, несомненно, следует ответить положительно. Действительно, отсутствие генетических связей между «праязыками» (т. е. доисторическими языками) можно было бы мотивировать, только предполагая спонтанное возникновение этих языков одновременно во многих местах земного шара примерно в один и тот же сравнительно поздний период независимо друг от друга — что противоречит как всем нашим представлениям об истории человека и языка, так и элементарному здравому смыслу. Хотя бы некоторые из существующих языковых семей должны оказаться родственными на более древнем уровне.

Второй вопрос значительно сложнее. Какова методика обнаружения и доказательства древнейших родственных связей между языковыми семьями? Ответ здесь, по-видимому, может быть только один: классическая методика сравнительно-исторического языкознания, то есть обнаружение между «праязыками» системы регулярных фонетических соответствий (соблюдающихся в основной массе лексических и морфологических сопоставлений) и реконструкция на их основании исходных более древних состояний. Именно эту методику использовал В. М. Иллич-Свитыч при реконструкции общеностратического языкового состояния.

Однако оппоненты дальнего сравнения указывают здесь, что если два языка развиваются независимо в течение периода порядка 10 тысяч лет (а именно такие — а, возможно, и большие языковые глубины следует предполагать для распада «праязыков» макросемей), они неизбежно должны утратить практически весь исконный лексический фонд и заменить его новым¹. Что же в таком случае мы можем сравнивать между собой?

Действительно, в подобных случаях ситуация, по-видимому, весьма сложна: вероятно, именно так обстоит дело с языками-изолятами типа баскского, шумерского или айнского, вероятность определения генетической принадлежности которых нужно признать крайне малой (именно в результате чрезвычайно раннего их отделения от каких-то

¹ На самом деле это не вполне так. Теория глоттохронологии доказывает, что исходный 100-словный список базисной лексики в языке заменяется целиком за период приблизительно в 15–20 тысяч лет. Если учесть при этом, что всякое — пусть даже небольшое — изменение значения слова по глоттохронологической теории считается его выпадением, то становится ясно, что корневой состав языка — вещь довольно консервативная, и основные корни могут (пусть с несколько модифицированной семантикой) сохраняться значительно дольше. На этом, в частности, основывается наша методика «корневой» глоттохронологии.

более крупных языковых единств и дальнейшего независимого развития). Но в подавляющем большинстве случаев дело обстоит иначе: мы имеем языковые семьи, а следовательно, возможность реконструировать для каждой из них праязыковое состояние — с тем, чтобы впоследствии сравнивать между собой уже не современные языки, но промежуточные «праязыки». Если мы согласны с тем, что сходство между реконструированными прагерманским и праславянским языковыми состояниями больше, чем между современными немецким и русским языками, то мы должны согласиться с тем, что — при условии древнего родства — сходство между праиндоевропейским и пракартвельским окажется более значительным, чем сходство между современными русским и грузинским языками. Таким образом, выход из положения — и одновременно единственно возможный способ работы с отдаленно родственными языками — мы видим в так называемой «ступенчатой» реконструкции промежуточных праязыковых состояний, сравнении их между собой и реконструкции на их основании исходного единого праязыкового состояния.

Все это — общие соображения. В реальной работе компаративиста, занимающегося сравнением отдаленно родственных языков, подстерегает еще одна, и самая существенная, опасность — возможность сравнения генетически неродственных элементов. Эти случаи распадаются на несколько категорий:

а) случайные совпадения как таковые (ср. англ. *woman* и др.-яп. *wotina* ‘женщина’);

б) идеофоны (ср. рус. *кукушка* и англ. *cuckoo*);

в) заимствования (ср. праяп. **kui* и австронез. **kaju* ‘дерево’).

Довольно многие работы, в которых делалась попытка доказать родство различных языковых семей, оказались неудачными как раз из-за неучтенной возможности случайных совпадений и старых заимствований². Особенно велика вероятность ошибки в случае т. н. «бинарного» сравнения (пользовавшегося особой популярностью в американской лингвистике), т. е. сравнения только двух отдаленно родственных языков или праязыков. Действительно, если мы находим в

² Критики дальнего сравнения иногда придают большое значение также второй категории случаев — идеофонам — как фактору, препятствующему сравнению далеко родственных языков (т. н. теория «элементарного родства»); на наш взгляд, однако, роль идеофонов явно преувеличена (неясно, по каким причинам, например, дейктическим основам типа **mV* ‘местоимение 1-го л.’ или **tV* ‘указательное местоимение’ следует приписывать идеофоническое происхождение), и мы здесь не будем на ней специально останавливаться.

и.-е. корень **per-* ‘передний край’ и в драв. **pir-* ‘задний край’, это может быть и случайным совпадением (вероятность которого может быть оценена статистически), если же к этой паре мы добавим алт. **p'era* ‘нижний край’, урал. **pera* ‘задняя часть’ и картв. *pir-* ‘передний край’ (см. [ОСНЯ: I, 27]), то вероятность случайного совпадения между всеми этими основаниями снижается практически до нуля. Если при этом можно найти еще большое количество аналогичных изоглосс и продемонстрировать регулярный характер их фонетических соотношений, а также обнаружить морфологические параллели, то генетическое родство можно считать доказанным³.

Даже при признании применимости классической компаративистической модели к изучению далекого родства языков, остается еще круг вопросов, не получивших надлежащего ответа, но имеющих едва ли не самое принципиальное значение. Речь идет о тесно взаимосвязанных проблемах классификации языков и датировки языковой дивергенции.

Действительно, если ностратическая семья существует, то:

1) каково ее членение? Оправдано ли, например, деление на восточноностратические и западноностратические языки? Являются ли тюркские, монгольские и тунгусо-маньчжурские языки отдельными ветвями ностратической семьи или их необходимо объединять в одну ветвь — алтайскую?

2) каковы ее границы? В настоящее время имеются сомнения в правомерности отнесения к ностратическим афразийским языкам — не исключено, что это отдельная макросемья приблизительно такой же временной глубины, как и собственно ностратическая. С другой стороны, в ряде работ, имеющих к ностратике лишь косвенное отношение, делается вывод о том, что ностратический праязык, восстановленный В. М. Иллич-Свитычем, может представлять собой вообще праязык человечества — как доказать или опровергнуть эту точку зрения?

3) какова ее приблизительная датировка, т. е. оценка времени раздельного существования отдельных ветвей ностратических языков

³ Все приведенные рассуждения, естественно, имеют смысл только в рамках классической концепции «родословного древа». Если мы не признаем реальность «праязыков» (т. е. не считаем, что реконструируемые системы совпадают или по крайней мере существенно приближаются к реально существовавшим), то вопрос о сравнении праязыков между собой, естественно, отпадает. Споры вокруг проблемы дальнего родства языков, таким образом, сводятся к признанию или непризнанию аксиоматики сравнительно-исторического языкознания и как таковые выходят за рамки проблем, обсуждающихся в данной работе.

(в литературе давались различные оценки времени ностратического единства — от 9-го до 20-го тысячелетия до н. э., а иногда предлагались и более ранние даты).

Те же самые вопросы, естественно, относятся не только к ностратической, но и ко всем прочим постулируемым в настоящее время макросемьям. Более того, при ближайшем рассмотрении оказывается, что традиционное сравнительно-историческое языкознание часто дает противоречивые ответы на эти вопросы даже и в случае с более молодыми языковыми семьями, само существование которых не вызывает никаких сомнений. Так, если границы индоевропейской семьи можно считать надежно установленными (весьма маловероятно, чтобы еще какие-либо из существующих и засвидетельствованных языков можно было в будущем отнести к индоевропейским, равно как и чтобы какие-либо языки, считающиеся индоевропейскими, оказались не принадлежащими к этой семье), то вопросы о ее членении и о ее временной глубине (а также о ее предполагаемой «прародине») до сих пор вызывают оживленные споры.

Во многих случаях различие в расстоянии между родственными языками вполне очевидно: можно назвать множество признаков, общих для славянских языков в отличие, к примеру, от германских и позволяющих констатировать большую близость к русскому болгарского языка, чем, скажем, шведского. Очевидно, что если мы научимся давать единомасштабную количественную оценку этого расстояния, то мы сможем:

а) создавать внутреннюю классификацию языков данной семьи (или по крайней мере оценивать существующие классификации), т. е. отвечать на вопрос о членении семьи;

б) находить место языка или языковой семьи в рамках более крупных языковых единств (семей, макросемей), т. е. отвечать на вопрос о границах семьи;

в) накладывая на полученный масштаб временную шкалу, определять вероятное время языковой дивергенции, т. е. отвечать на вопрос о хронологии.

Таким образом, принципиальное значение приобретает обнаружение такого рода количественных характеристик языков, которые могли бы быть использованы в качестве объективных критериев для определения расстояния между языками⁴. Оценки фонетической,

⁴ Ясно, что такие критерии могут служить и критериями верификации родства языков. Так, между неродственными языками расстояние должно оцениваться как бесконечное.

морфологической и синтаксической близости языков, по-видимому, не могут быть использованы в качестве таких критериев: известны случаи чрезвычайно быстрых и радикальных изменений в области фонетики, морфологии и синтаксиса, а также примеры исключительной консервативности этих сфер языка. Скорость языковых изменений в упомянутых сферах, таким образом, явно неравномерна и вряд ли поддается измерению; статистической оценке скорости здесь препятствует также малый набор исходных элементов (фонем, грамматических морфем), из которых трудно получать достаточные для статистической обработки выборки⁵.

Единственной сферой языка, в которой скорость изменения, по-видимому, является равномерной (см. об этом подробнее ниже) и которая одновременно представляет собой вполне удобный объект для количественных и статистических оценок, является лексика. Поэтому мы хотим остановиться здесь на различных аспектах использования лексико-статистической (или глоттохронологической) методики.

Несмотря на большое количество критических работ, обозначивших многие недостатки изобретенной М. Сводешем еще в 50-е годы методики глоттохронологии (одно время казалось даже, что глоттохронология разбита наголову), представляется, что полностью отказываться от нее не только преждевременно, но и неверно. Всякий компаративист, имевший дело с глоттохронологическими подсчетами, знает, что близкие диалекты обычно обнаруживают в 100-словном списке Сводеша 90 и более процентов совпадений; близкородственные языки (типа славянских, романских, германских, тюркских — т. е. распавшихся около 1,5–2 тысяч лет назад) обнаруживают в этом списке 70–80% совпадений; наконец, языки семей типа индоевропейской (распавшиеся около 5–6 тысяч лет назад) обнаруживают между собой 25–30% совпадений. При переходе к членам более древних семей, например, уральской или алтайской, мы обнаруживаем между современными языками уже не более 10–20% соответствий; наконец, процент совпадений между современными членами разных ветвей

⁵ Из сказанного не следует, конечно, что на основании фонетических и морфологических данных нельзя строить генетические классификации языков (известно, что большинство имеющихся основано именно на этих данных). Утверждается лишь, что этих данных недостаточно для количественной оценки расстояния между языками. Так, мы можем отделить группу миньских диалектов Китая от всех прочих на том основании, что в них сохраняется противопоставление **b* — *bh*, **d* — *dh* и т. п., утраченное в других диалектах; но определить по этой черте время отделения миньских диалектов мы не можем.

макросемей типа ностратической еще меньше (на уровне 5–9%). Приведенные цифры, конечно, приблизительны (примеры более точных данных см. ниже). Однако следует подчеркнуть, что они обладают свойством систематической повторяемости: ни один славянский язык не имеет ни с одним другим славянским языком менее 75% совпадений; ни один индоевропейский язык не имеет ни с одним другим индоевропейским языком (не принадлежащим к той же самой группе) более 35% (и менее 20%) совпадений. По опыту лексикостатистических обследований материала мы знаем, что абсурдной была бы цифра в 50% совпадений, с одной стороны, между русским и польским языками, а с другой — между русским и немецким. Аналогичная равномерность распределения долей совпадений между членами равноудаленных друг от друга (по предварительным оценкам) языковых групп хорошо известна и в других языковых семьях (в австронезийской, уральской, синотибетской и т. д. и т. п.). Все эти соображения вселяют в нас уверенность в том, что скорость изменения в лексике (во всяком случае в некоторых ее сферах) действительно является равномерной и универсальной величиной. Поэтому имеет смысл еще раз критически рассмотреть основные постулаты глоттохронологии и принципы ее использования.

Основные принципы глоттохронологии М. Сводеша сводятся к пяти постулатам, хорошо сформулированным в работе [Арапов, Херц 1974: 21, 25]:

«1. В словаре каждого языка можно выделить специальный фрагмент, который мы будем называть дальше основной, или стабильной, частью.

2. Можно указать список значений, которые в любом языке обязательно выражаются словами из основной части... Будем говорить, что эти слова образуют основной список /ОС/. Через N_0 обозначим число слов в ОС.

3. Доля p слов из ОС, которые сохранятся (не будут заменены другими словами) на протяжении интервала времени t ... постоянна (то есть зависит только от величины выбранного промежутка, но не от того, как он выбран или какие слова какого языка рассматриваются).

4. Все слова, составляющие ОС, имеют одинаковые шансы сохраниться (соответственно, не сохраниться, 'распасться') на протяжении этого интервала времени.

5. Вероятность для слова из ОС праязыка сохраниться в ОС одного языка-потомка не зависит от его вероятности сохраниться в аналогичном списке другого языка-потомка».

Совокупность приведенных постулатов используется для вывода следующих основных математических зависимостей глоттохронологии:

$$(1) N(t) = N_0 e^{-\lambda t}$$

где время, прошедшее от начала момента развития до некоторого последующего момента, обозначается как t (и измеряется в тысячелетиях); N_0 есть исходный ОС; λ есть «скорость выпадения» слов из N_0 ; и $N(t)$ есть доля слов исходного ОС, сохранившихся к моменту t . Так, если исходный список насчитывает сто слов и обозначается как единица /1/, $\lambda = 0.14$ (т. е. принимается, что за тысячу лет из ста слов выпадает 14), а время, прошедшее с момента начала распада составляет, к примеру, две тысячи лет, то

$$N(2) = e^{-0.14 \cdot 2} = 0.76 \text{ (т.е. 76 слов).}$$

Зная долю слов, сохранившихся в данном языке из списка ОС, мы можем вычислить путем логарифмирования длину прошедшего промежутка времени:

$$(2) t = \frac{\ln N(t)}{-\lambda \cdot N_0}$$

Поскольку — согласно пятому постулату — мы считаем развитие двух языков из одного праязыка независимым, то, зная долю совпадающих слов в описках ОС двух или более родственных языков, мы можем вычислить время, отделяющее их от момента распада. Действительно, доля совпадений между n языками (при условии единого общего для них праязыкового ОС) составит

$$(3) N_n(t) = N_0 e^{-n\lambda t}$$

а время, разделявшее их, будет вычисляться по формуле

$$(4) t = \frac{\ln N_n(t)}{-n\lambda N_0}$$

Так, при указанном выше λ ($= 0.14$), доля совпадений между двумя языками, равная 0.8 (т. е. 80%) будет означать, что от распада их отделяет

$$t = \frac{\ln 0.8}{-2 \cdot 0.14} = 0.8 \text{ (т. е. около 800 лет)}^6.$$

Следует особо подчеркнуть, что получаемые датировки имеют сугубо вероятностный характер и допускается вероятность более или менее значительных ошибок в определении времени; доверительные интервалы в принципе вычислимы, но их рассмотрение сейчас не входит в нашу задачу.

«Скорость» λ представляет собой эмпирически определяемую константу (выводимую на основании замеров, сделанных по языкам, исторически зафиксированных на протяжении последней тысячи лет и более). Величина $\lambda = 0.14$, использованная в вышеприведенных примерах, не случайна: именно эту константу постулировал для своего стословного списка ОС основоположник глоттохронологии М. Сводеш (см. [Сводеш 1960а] и др.).

Описанная выше теория глоттохронологии (математический аппарат которой практически целиком заимствован из физической теории радиоактивного распада) сама по себе довольно красива и проста. Однако следует, к сожалению, с полной определенностью констатировать, что для лингвистических датировок глоттохронология М. Сводеша непригодна⁷. Практически во всех случаях датировки исторически

⁶ В работах по глоттохронологии часто можно встретить несколько упрощенную запись перечисленных математических зависимостей, где вместо «скорости» используется коэффициент сохраняемости лексики $r = 1 - \lambda$. Чаще всего можно встретить записи формулы:

(1) как (1') $c = r^t$ (где c соответствует $N(t)$, а N_0 принимается за единицу);

(2) как (2') $t = \frac{\log c}{\log r}$;

(3) как (3') $c = r^{nt}$ (чаще всего встречается формула $c = r^{2t}$ для $n = 2$);

(4) как (4') $t = \frac{\log c}{n \log r}$ (опять-таки чаще всего используется формула $t = \frac{\log c}{2 \log r}$

для $n = 2$).

⁷ По всей видимости, это обстоятельство обусловило тот факт, что многие исследователи, пользуясь методикой глоттохронологии, в настоящее время уже не пытаются использовать ее для абсолютных датировок, но лишь для относительной хронологизации (т. е. построения генеалогических деревьев) языков.

документированных лингвистических событий наблюдается один и тот же результат: сильное «умоложение» дат. Попробуем разобраться в причинах этого явления.

Прежде всего следует рассмотреть вопрос о постоянстве скорости распада лексики в ОС (3-й постулат глоттохронологии). Этот момент послужил предметом самых ожесточенных споров: ясно, что если такого постоянства не наблюдается, вся процедура глоттохронологии становится бессмысленной.

В литературе по глоттохронологии встречается точка зрения, что «утверждения об изменениях в лексике не распространяются на языки, имеющие культивируемую литературную норму» (см., например, [Яхонтов 1984: 45]). Это мнение сторонников глоттохронологии является реакцией на критику, высказанную в классической статье К. Бергсланда и Х. Фогта ([Bergsland, Vogt 1962]). Эти авторы, рассматривая материал скандинавских языков, показали, что скорость распада лексики ОС в исландском языке за последнюю тысячу лет равнялась всего ≈ 0.04 ($r \approx 0.96$), а в литературном норвежском (риксмоле) ≈ 0.2 ($r \approx 0.8$). Соответственно мы получаем (при применении константы М. Сводеша 0.14) нелепые результаты: для исландского языка около 100–150 лет, а для риксмолы около 1400 лет развития (при том что оба языка развились из одного источника и развивались независимо в течение около 1000 лет). Аналогичные результаты были получены О'Нейлом [O'Neil 1964] при сравнении исландского и фарерского ОС: исторически известна датировка их дивергенции (в X в.), но 94% совпадений между этими языками по формуле Сводеша указывают на дату их дивергенции всего лишь около 200 лет назад.

Эти, казалось бы, очевидные факты и заставляют исследователей делать скидку на «литературную норму», замедляющую развитие лексики (но в случае с риксмомом имеем, как мы видели выше, наоборот, ускоренное развитие). Попробуем, однако, критически разобраться в скандинавской ситуации.

Проще всего объясняется «ускоренное» развитие риксмолы. Этот язык является, по существу, гибридом норвежского и датского языков, и его 100-словный список включает в себя 11 датских, 3 шведских и 2 немецких заимствования. При подсчете всех этих заимствований как замен по сравнению с древнеисландским мы и получаем исключительно высокую скорость норвежского развития. В то же время, если исключить эти заимствования из рассмотрения, мы получим для исландского скорость, равную 0.06 (заменилось 6 слов: *eta* 'есть' > *borda*,

lauf ‘лист’ > *blað*, *verr* ‘мужчина’ > *maður*, *svima* ‘плавать’ > *synda*, *varmr* ‘теплый’ > *hlyr*, *mani* ‘луна’ > *tungl*), и для риксмолы — 0.05 (заменялось четыре слова: *eldr* ‘огонь’ > *varme*, *lauf* ‘лист’ > *blad*, *hold* ‘мясо’ [‘flesh’] > *kjøtt* и *sa* ‘тот’ > *den* из 84 при неучете заимствований). Очень близкую скорость получаем для прочих скандинавских языков: для фарерского — 0.05, для шведского, датского и норвежских диалектов Гвестал и Санднес — 0.04⁸. Итак, мы видим, что к резким искажениям приводит учет заимствований в качестве лексических замен. По-видимому, из этого следует важный вывод: при глоттохронологических подсчетах заимствования (во всяком случае, в ситуации массовых заимствований в одном языке из соседнего) следует устранять из списка ОС, учитывая только скорость изменения исконной лексики⁹.

При этом условии скорость развития лексики ОС скандинавских языков оказывается в целом стабильной и незначительно уклоняется от величины 0.05. Ясно, однако, что эта скорость значительно медленнее, чем постулированная М. Сводешем константа 0.14. Остается как будто бы только — вслед за Бергсландом и Фоггом — констатировать более медленную скорость развития скандинавских языков по сравнению с остальными.

Обратимся, однако, к прочим языкам. В нашем распоряжении находятся материалы по развитию ОС во многих документированных языках мира, которые показывают следующую картину:

⁸ Величины эти могут несколько варьировать (от 0.04 до 0.06) в зависимости от того, какие обозначения ‘земли’ (*mold* или *jgrð*) и ‘мяса’ (*hold* или *kjøtt*) мы выберем в качестве основных для древнеисландского. На общую оценку скорости этот выбор, конечно же, не влияет.

⁹ Таким образом, замены в пределах исконной лексики мы считаем равномерным процессом, а замены исконных слов на заимствования — более или менее случайными «сбоями», на которые нужно делать поправку. Из этого сразу же следует вывод о том, что глоттохронологические подсчеты в условиях неразработанной сравнительной фонетики и этимологии (т. е. при невозможности отличить исконную лексику от заимствованной) существенным образом обесцениваются. Следует, однако, специально подчеркнуть нашу позицию: никакое лексикостатистическое обследование вообще невозможно без предварительной тщательной компаративистской обработки материала. Поэтому, к примеру, лексикостатистические подсчеты Хаттори Сиро, сравнивавшего японский 100-словный список со списками самых разных языков мира [Hattori 1959], или подсчеты самого М. Сводеша, оценивавшего степень лексикостатистической близости самых разных языков Евразии и Америки («дене-финская» теория, см. [Сводеш 1965]), очевидно, лишены какого бы то ни было смысла.

Язык	t	Л ₁ (при учете заимствований как замен)	Л ₂ (при неучете заимствований)
Японский	1,2	0.11	0.06 ¹⁰
Китайский	2,6	0.1 ¹¹	
Английский	1,3	0.14	0.1 ¹²
Немецкий	1,2	0.08	0.05 ¹³
Французский	1,5	0.09	0.07 ¹⁴
Испанский	1,5	0.07	0.06 ¹⁵
Румынский	1,5	0.09	0.06 ¹⁶

¹⁰ В японском языке по сравнению с др.-яп. (VIII в. н. э.) заменены слова: *кар-и* 'есть' > *taberu*, *карйгю* 'яйцо' > *tamago*, *касира* 'голова' > *atama*, *исагю* (*манагю*) 'песок' > *suna*, *ви-ри* 'сидеть' > *suwaru*, *ка-но* 'тот' > *ano*, *наре* 'ты' > *anata* (*кинй*), на заимствования заменены *kokoro* 'сердце' > *shinzō*, *кимо* 'печень' > *kanzō*, *сиси* 'мясо' > *niku*, *кара* 'кожа' > *hifu*, *пи* 'солнце' > *taiyō*.

¹¹ В китайском языке по сравнению с др.-кит. (VII в. до н. э.) при полном отсутствии заимствований заменены слова: *крāj* 'все, весь' > *yiqie*, *дуо*, *пук* 'живот' > *duzi*, *чрū?* 'коготь, ноготь' > *jia*, *кх^мин* 'собака' > *gou*, *ʔат* 'пить' > *he*, *лак* 'есть' > *chi*, *рōн* 'яйцо' > *dan*, *чок* 'нога' > *jiao*, *пiс* 'давать' > *gei*, *рап* 'хороший' > *hao*, *кеп* 'шея' > *bozi*, *кхек* 'красный' > *hong*, *ват* 'сказать' > *shuo*, *кēнн* 'видеть' > *kan*, *рар* 'стоять' > *zhan*, *нит* 'солнце' > *taiyang*, *сэ, де?* 'этот' > *zhe*, *слхай, пай?* 'тот' > *na*, *тхэ?* 'зуб' > *ya*, *мōк* 'дерево' > *shu*, *нис* 'два' > *liang*, *грāп* 'идти, ходить' > *zou*, *гхāj* 'что' > *shemta*.

К сожалению, довольно трудно составить 100-словные списки для промежуточных периодов истории китайского языка (хорошую фиксацию разговорных слов мы имеем только в древнейших текстах, с одной стороны, и в современном языке — с другой).

¹² В английском языке по сравнению с др.-англ. (IX в. н. э.) заменены слова: *wamb* 'живот' > *belly*, *micel* 'большой' > *big*, *фуыл* 'птица' > *bird*, *волцен* 'облако' > *cloud*, *hund* 'собака' > *dog* (возможно, нижненемецкое заимствование), *γута* (*вер*) 'мужчина' > *man*, *flæsc* 'мясо' > *meat*, *heals* 'шея' > *neck*, *wey* 'дорога' > *road*, *rec* 'дым' > *smoke*, *se* 'тот' > *that*, *þu* 'ты' > *you*; на заимствования заменены *rinde* 'кора' > *bark*, *steorfan* 'умирать' > *die*, *beory* 'гора' > *mountain*, *sine-wealt* 'круглый' > *round*, *hyd* 'кожа' > *skin*.

¹³ В немецком языке по сравнению с др.-в.-нем. (VIII в. н. э.) заменены слова: *bein* 'кость' > *Knochen*, *wamba* 'живот' > *Bauch*, *mihhil* 'большой' > *groß*, *luzzil* 'маленький' > *klein*, *zagal* 'хвост' > *Schwanz*, *wib* 'женщина' > *Frau*; на заимствования заменены: *feizzit* 'жир' > *Fett*, *houbit* 'голова' > *Kopf* и *sinwēl* 'круглый' > *rund*.

¹⁴ Во французском языке по сравнению с вульгарной латынью (IV–V вв. н. э.) заменены слова: *penna* 'перо' > *plume*, *caput* 'голова' > *tête*, *audire* 'слышать' > *entendre*, *occidere* 'убивать' > *tuer*, *multum* 'много' > *beaucoup*, *carnem* 'мясо' > *viande*, *arenam* 'песок' > *sable*, *sementem* 'семя' > *graine*, *ambulare* 'идти, ходить' > *aller* (возможно, заимствование), *natare* 'плавать' > *nager*, *parvus* 'маленький' > *petit*; на заимствование заменено *albus* 'белый' > *blanc*.

¹⁵ В испанском по сравнению с вульгарной латынью (IV–V вв. н. э.) заменены слова: *canem* 'собака' > *perro* (возможно, заимствование), *occidere* 'убивать' > *matar*, *geniculum* 'колено' > *rodilla*, *stare collocatum* 'лежать' > *estar acostado*, *longus* 'длинный' > *largo*, *parvus* 'маленький' > *pequeño*, *ambulare* 'ходить' > *andar*, *galbinus* 'желтый' > *amarillo*; на заимствования заменены *penpat* 'перо' > *pluma* и *albus* 'белый' > *blanco*.

Этот список можно было бы продолжить (в частности, И. Фодором [Fodor 1961] показана аналогичная медленная скорость развития лексики славянских языков); однако ясно, что скорость развития списков ОС на протяжении последней тысячи — полутора тысяч лет в среднем значительно меньше, чем константа 0.14, постулированная М. Сводешем; последняя наблюдается только в случае с английским языком при учете заимствованной лексики¹⁷. Сравнительно высокую (0.1) скорость развития китайской лексики мы надеемся объяснить ниже.

Проще всего было бы, учитывая все вышеизложенное, заменить константу $\lambda = 0.14$ М. Сводеша на $\lambda = 0.06$ (среднюю величину для всех λ_2 , приведенных выше). В случае с разобранными языками при константе $\lambda = 0.06$ мы бы получили несколько замозаженных даты для скандинавских языков и более или менее точную датировку для других романских и германских (кроме английского), однако уже для китайского мы получим совершенно нереальную датировку (середина 3-го тысячелетия до н. э.); приблизительно на тысячу лет удревется и датировка древнеанглийского. Более того, если мы попробуем датировать при помощи этой константы даты языковых дивергенций (по формуле 4), мы потерпим полное фиаско. Так, для распада белорусского и украинского языков (имеющих 97% совпадений в 100-словном списке) мы получим:

$$t = \frac{\ln 0.97}{-2 \cdot 0.06} = 0.25$$

т. е. 250 лет — дата, по-видимому, ничему не соответствующая¹⁸. Аналогичное время самостоятельного развития будет отделять друг от

¹⁶ В румынском языке по сравнению с вульгарной латынью заменены слова: *ventrem* 'живот' > *pîntece*, *grandem* 'большой' > *mare*, *avem* 'птица' > *pasăre*, *ustulare (cremare)* 'жечь' > *arde*, *frigidus* 'холодный' > *rece*, *siccus* 'сухой' > *uscat*, *terram* 'земля' > *pămînt*, *cordem* 'сердце' > *inima*, *buccam* 'рот' > *gura*; на заимствования заменены *collum* 'шея' > *gît*, *samminum* 'дорога' > *drum*, *arenam* 'песок' > *nisip*, *parvus* 'маленький' > *mic*.

¹⁷ В остальных языках даже при учете заимствованной лексики скорость существенно меньше, чем 0.14. Несколько «ускоренное» развитие английского следует, по-видимому, объяснить активной языковой интерференцией при формировании новоанглийского языка (участие в этом процессе как скандинавского, так и романского компонента).

¹⁸ Для распада белорусского и украинского по формуле Сводеша мы получаем датировку серединой XIX века: для разделения исландского и феррерского, как это

друга, к примеру, современные персидский и таджикский языки, идиш и немецкий язык и др. Датировка здесь оказывается явно «умоложенной». Напротив, если мы перейдем на индоевропейский уровень и попытаемся продатировать, к примеру, распад русского и персидского языков (имеющих 28% совпадений в стословном списке), мы получим:

$$t = \frac{\ln 0.28}{-2 \cdot 0.06} = 10,6$$

т. е. начало 9-го тысячелетия до н. э. — при более или менее общепринятой точке зрения на распад общеиндоевропейского как произошедший около 4-го тысячелетия до н. э. (другие пары индоевропейских языков будут давать аналогичные, чрезвычайно удревленные датировки).

Теперь становится понятным выбор М. Сводешем константы $\lambda = 0.14$. При нормальной экспоненциальной зависимости между временем и долей сохраняющейся лексики эта величина обеспечивает отсутствие излишнего удревления дат на больших временных глубинах, хотя в то же время дает эффект некоторого стандартного умоложения дат на всем отрезке измеряемого времени. Поскольку, как мы видели, константа 0.14 реально не наблюдается нигде (кроме разве английского языка — и то с определенными оговорками насчет заимствований), ей никак не может быть придан статус эмпирически определенной величины. В то же время эмпирически наблюдаемые величины λ позволяют удовлетворительно датировать события тысячелетней или полуторатысячетней давности, но оказываются (как было показано выше) совершенно непригодными для датировок в пределах более поздних и более ранних эпох. Требуется понять сущность этого противоречия и найти более или менее радикальный выход из положения.

Рассмотрим еще раз данные романских языков. Выше мы определяли величину скорости на основании сравнения ОС романских языков с ОС вульгарной латыни (представляющей собой праязык по отношению к современным романским и распавшейся предположительно в IV–VI вв. н. э.). Перейдем теперь к сравнению с латынью

уже указывалось выше, — XVIII век; для разделения праславянского, прагерманского и прароманского (т. е. народной латыни) мы получаем соответственно датировки XI–XII вв. н. э.; для распада общеиндоевропейского — середина или конец 3-го тысячелетия до н. э.

классической, ОС которой эпохи II–IV вв. до н. э. хорошо известен¹⁹, и посмотрим, какие изменения произошли во французском, испанском и румынском ОС по сравнению с классической латынью.

Во французском заменены слова: *omnis* > *tout* ‘весь’, *magnus* > *grand* ‘большой’, *edere* > *manger* ‘есть’, *pinguedo (adeps)* > *graisse* ‘жир’, *penna* > *plume* ‘перо’, *ignis* > *feu* ‘огонь’, *caput* > *tête* ‘голова’, *audire* > *entendre* ‘слышать’, *occidere* > *tuer* ‘убивать’, *cubare* > *être couché* ‘лежать’, *jecur* > *foie* ‘печень’, *vir* > *homme* ‘мужчина’, *multus* > *beaucoup* ‘много’, *caro* > *v viande* ‘мясо’, *os* > *bouche* ‘рот’, *arena* > *sable* ‘песок’, *semen* > *graine* ‘семя’, *cutis* > *peau* ‘кожа’, *parvus* > *petit* ‘маленький’, *nare (natate)* > *nager* ‘плавать’, *ire* > *aller* ‘идти, ходить’, *flavus (fulvus)* > *jaune* ‘желтый’; на заимствования заменены слова *via* > *chemin* ‘дорога’, *lapis* > *Pierre* ‘камень’ и *albus* > *blanc* ‘белый’. Таким образом, имеем 77% сохранившихся слов: полагая $t \approx 2,3$ получаем величину $\lambda t = 0.11$ (при скорости от вульгарной латыни к современному французскому $\lambda = 0.7$, см. выше).

В испанском заменены слова: *omnis* > *todo* ‘весь’, *magnus* > *grand* ‘большой’, *urere* > *quemar* ‘жечь’, *canis* > *perro* ‘собака’, *pinguedo (adeps)* > *grasa* ‘жир’, *ignis* > *fuego* ‘огонь’, *occidere* > *matar* ‘убивать’, *genu* > *rodilla* ‘колено’, *cubare* > *estar acostado* ‘лежать’, *jecur* > *higado* ‘печень’, *longus* > *largo* ‘длинный’, *vir* > *hombre (varón)* ‘мужчина’, *os* > *boca* ‘рот’, *via* > *camino* ‘дорога’, *cutis* > *piel* ‘кожа’, *parvus* > *pequeño* ‘маленький’, *ire* > *andar* ‘идти, ходить’, *flavus (fulvus)* > *amarillo* ‘желтый’; на заимствования заменены *penna* > *pluma* ‘перо’, *lapis* > *piedra* ‘камень’ и *albus* > *blanco* ‘белый’. Сохранилось, таким образом, 80% слов, что (при $t = 2,3$) дает $\lambda t = 0.1$ (ср. со скоростью от вульгарной латыни к испанскому $\lambda = 0.06$, см. выше).

В румынском заменены слова: *omnis* > *tout* ‘весь’, *venter* > *pîntece* ‘живот’, *magnus* > *mare* ‘большой’, *avis* > *pasăre* ‘птица’, *urere* > *arde* ‘жечь’, *frigidus* > *rece* ‘холодный’, *siccus* > *uscat* ‘сухой’, *terra* > *pământ* ‘земля’, *edere* > *mînca* ‘есть’, *pinguedo (adeps)* > *grasime* ‘жир’, *ignis* > *foc* ‘огонь’, *cor* > *inima* ‘сердце’, *cubare* > *sta culcat* ‘лежать’, *jecur* > *ficat* ‘печень’, *os* > *gură* ‘рот’, *vir* > *om (barbat)* ‘мужчина’, *cutis* > *piele* ‘кожа’, *ire* > *umbla* ‘ходить, идти’, *flavus (fulvus)* > *galben* ‘желтый’; на заимствования заменены *collum* > *gât* ‘шея’, *via* > *drum* ‘дорога’, *arena* > *nisip* ‘песок’, *parvus* > *mic* ‘маленький’ и *lapis* > *piatra* ‘камень’. Сохранилось столько же слов,

¹⁹ Народная латынь датируется равными исследователями по-разному [Muller 1929; Гурычева 1959]; несомненным, однако, представляется единый характер, хотя, вероятно, с некоторыми диалектными различиями, народной (вульгарной) латыни до V в. н. э., ее все более активное дробление на диалекты в период с V по VIII в. н. э. и начало эпохи романских языков с VIII в. н. э.

сколько и в испанском (80%), что дает нам ту же скорость за 2,3 тысячи лет $\lambda t = 0.1$ (опять при скорости от вульгарной латыни к румынскому $\lambda = 0.06$, см. выше).

Мы сталкиваемся, таким образом, с явным феноменом увеличения скорости (т. е. ускорения) распада ОС при увеличении времени распада Δt . Этот факт объясняет чрезвычайно малую скорость распада лексики при разделении белорусского и украинского, а также персидского и таджикского языков²⁰ (при 97% попарных совпадений между этими языками и временем, приблизительно в 600 лет, отделяющим каждый из этих языков от даты их распада, скорость λ нужно определять примерно как 0.03). С другой стороны, объясняется также сравнительно высокая (0.1) скорость распада китайской лексики: период, прошедший от древнекитайского до современного китайского, сравним по величине с временем, разделяющим классическую латынь и современные романские языки.

Как уже указывалось выше, математический аппарат глоттохронологии был целиком заимствован из теории радиоактивного распада. Но существенным отличием слов от нейтронов является, по-видимому, то обстоятельство, что первые — в отличие от вторых — «стареют». Действительно, вне зависимости от того, сколько тот или иной нейтрон «прожил» к моменту времени $t = 0$, его вероятность уцелеть к моменту времени t будет все равно представлять собой величину $e^{-\lambda t}$. Напротив, вероятность сохранения слова (в том числе слова из ОС) оказывается связанной со временем, которое это слово уже «прожило». Только таким образом можно объяснить, например, факт универсальной замены во всех современных языках общеиндоевропейского **gũ(e)ru-* 'тяжелый': это слово представлено в др.-инд. *guru-*, др.-гр. βαρῦς, лат. *gravis*, гот. *kaurus*, но утрачено в большинстве современных языков (в значении 'тяжелый' эта основа сохраняется, вероятно, только в новогреческом и в немногих новоиндийских языках: в большинстве индийских, германских, романских, славянских и пр. и.-е. языков она либо изменила значение, либо исчезла вообще). Примеры такого рода (широкое распространение некоторого слова или корня в древних языках при почти полной утрате в современных) хорошо известны компаративистам и, по всей видимости, связаны именно со «сроком жизни» слов.

²⁰ Для белорусского и украинского по историческим данным разделение можно датировать XIII–XIV вв., если соотносить его с отделением Белоруссии после захвата ее Литвой; для персидского и таджикского наиболее вероятная датировка разделения конец XIV — начало XV в. (отделение Таджикистана и части восточного Ирана в результате монгольского нашествия).

Предположение о «старении» слов автоматически предполагает отказ от 3-го постулата глоттохронологии (т. е. от предположения о постоянстве скорости развития ОС) и переход к гипотезе о зависимости величины λ от времени t . Действительно, если мы предположим, что чем больше время t , тем больше становится вероятность выпадения каждого отдельного олова из списка ОС, то вместо формулы (1) мы должны будем принять для глоттохронологии основную зависимость:

$$(5) N(t) = N_0 e^{-\lambda t^2}$$

т. е. формулу «равномерно ускоренного движения»²¹.

Соответственно, время вычисляется здесь по формуле:

$$(6) t = \sqrt{\frac{\ln N(t)}{-\lambda * N_0}}$$

а время, разделяющее n языков, — по формуле:

$$(7) t = \sqrt{\frac{\ln N_n(t)}{-n\lambda N_0}}$$

Подставляя в эти формулы приведенные выше данные различных языков, получаем следующие коэффициенты ускорения λ :

Язык	t	λ
Японский	1,2	0.05
Английский	1,3	0.08
Китайский	2,6	0.04
Немецкий	1,2	0.04
Французский	1,5	0.05
Исландский	1	0.06
Риксмол	1	0.05

— и т. д. Итак, за исключением английского (обнаруживающего несколько большее ускорение), для всех прочих языков величина λ в целом стабильна и незначительно варьирует от 0.04 до 0.06. При датировании языкового распада в пределах последних тысячи — полутора тысяч лет использование описанного материала также дает в целом

²¹ Возможность такой зависимости была предусмотрена (хотя лишь чисто теоретически) уже в работе [Арапов, Херц 1974: 50].

хорошие результаты. Так, для белорусского и украинского (а также для персидского и таджикского) получаем:

$$t = \sqrt{\frac{\ln 0.97}{-2 \cdot 0.05}} = 0.55 \text{ (т. е. конец XIV в. н. э.)}$$

При датировке распада исландского и фарерского языков (94% совпадений) получаем примерно XII в. (вместо XVIII в., по М. Сводешу) и т. д.

Однако можно показать, что и использование зависимости (5) при переходе к большим временным глубинам начинает давать весьма значительное умоложение датировок. Так, для русского и персидского при использовании формулы (7) и величины $\lambda = 0.05$ время дивергенции будет составлять:

$$t = \sqrt{\frac{\ln 0.28}{-2 \cdot 0.05}} = 3,6$$

т. е. примерно середина 2-го тысячелетия до н. э. Даже, по Сводешу, при 28% совпадений мы должны получить конец 3-го тысячелетия до н. э., а реальнее всего была бы датировка 4-м тысячелетием до н. э. (см. выше).

Более того, можно показать, что при принятии зависимости (5) период полного распада списка ОС должен составить менее 10 тысяч лет, а у двух разошедшихся языков какие бы то ни было совпадения в 100-словном списке должны исчезнуть приблизительно за 7 тысяч лет дивергенции. Нет никакого сомнения в том, что принятие зависимости

$$N(t) = N^0 e^{-\lambda t^2}$$

несколько улучшает датировку (и лучше согласуется с реальной моделью развития лексики) для близких к нам эпох, но по мере углубления в древность дает еще худшие результаты, чем Сводешевская глоттохронология. Как же разрешить это противоречие?

Для того, чтобы продвинуться дальше, нам необходимо разобрать 4-й постулат глоттохронологии: «все слова, составляющие ОС, имеют одинаковые шансы сохраниться на протяжении некоторого интервала времени» (см. выше, с. 413). Именно этот момент всегда вызывал наибольшее сомнение как у сторонников, так и у противников глоттохронологии. Действительно, всякий лингвист, работавший с глоттохронологией, хорошо знает, что разные слова списка ОС на самом деле

обладают разной устойчивостью: шансов на выпадение из 100-словного списка слов 'маленький' или 'кожа' в целом значительно больше, чем у слов 'я', 'ты' или 'ухо'. Поэтому неоднократно высказывалось предположение о том, что коэффициент сохраняемости всего списка в целом должен определяться как математическое ожидание индивидуальных коэффициентов сохраняемости (т. е. вероятностей сохраниться на протяжении отрезка времени t) каждого слова; см., например, [Merwe 1966]. Попытку эмпирического определения этих индивидуальных коэффициентов (на австронезийском и индоевропейском материале) предприняли И. Дайен и Э. Джеймс [Dyen, James 1967]; однако предложенный ими математический аппарат для датировок представляется слишком громоздким и малоэффективным. Более того, если обоснованно предположение о том, что разные слова имеют разный коэффициент сохраняемости, то не менее вероятен и факт того, что коэффициент сохраняемости индивидуального слова может варьировать в зависимости от культурного и лингвистического окружения (так, слова 'облако' и 'хвост' весьма устойчивы в тюркских языках, но весьма нестабильны в германских; слово 'живот' очень устойчиво в романских, но неустойчиво в славянских и т. д. и т. п.). При том, что некоторые из слов ОС действительно обнаруживают универсально высокую устойчивость ('я', 'ты', 'солнце', 'глаз', 'ухо' и др.), все же учет индивидуальных коэффициентов слов, по-видимому, не может решить поставленной проблемы (либо же решение будет неуниверсальным и слишком громоздким).

Представим себе некоторый идеальный список ОС с некоторой средней скоростью распада λ , в котором все слова ранжированы по вероятности их выпадения за промежуток времени $\Delta\lambda$, причем первое слово имеет вероятность, близкую (или равную) 0, а i -е слово имеет вероятность, близкую (или равную) 1. В таком случае слова должны в принципе выпадать поочередно: сначала менее устойчивые, потом — более устойчивые. При этом по мере разрушения списка скорость распада его части, сохраняющейся в момент времени t , должна замедляться по сравнению с моментом времени t_{n-1} . Скорость распада такого списка будет, следовательно, переменной величиной, зависящей в каждый момент времени от доли сохранившихся (и, соответственно, выпавших) слов:

$$(8) \lambda t_n = \lambda_0 * N(t_n)$$

Подставляя выражение (8) в формулу (5), мы получаем следующую зависимость:

$$(9) N(t) = N_0 e^{-\lambda N(t)t^2}$$

Время t будет вычисляться здесь по формуле:

$$(10) t = \sqrt{\frac{\ln N(t)}{-\lambda * N(t) * N_0}}$$

Обратим внимание на то обстоятельство, что в случае с датировкой распада n языков зависимость (9) приобретает вид:

$$(11) N_n(t) = N_0 e^{-\lambda N(t)t^2}$$

где $N(t)$ — величина, реально не наблюдаемая, но с большой или меньшей точностью вычисляемая по формуле

$$(12) N(t) = \sqrt[n]{N_n(t)}$$

Из (11) и (12) получаем способ вычисления времени дивергенции по данным о доле совпадающих слов в ОС n языков:

$$(13) t = \sqrt{\frac{\ln \frac{N_n(t)}{N_0}}{-n\lambda \sqrt[n]{N_n(t)}}}$$

²² Для практического использования (в наиболее частом случае, когда $n = 2$, $N_0 = 1$ (100 слов) при $\lambda = 0.05$ и при более традиционном обозначении $N(t)$ как c формулу можно упростить:

$$(14) t = \sqrt{\frac{\ln c}{-0.1\sqrt{c}}}$$

Эту же зависимость (беря за основу не «коэффициент выпадения» λ , а «коэффициент сохранения» γ) можно представить в виде:

$$(15) t = \sqrt{\frac{\log c}{\sqrt{c} \log 0.9}}$$

Формула (9) отражает «противоречивый» характер процесса распада лексики в ОС: показатель степени при t отражает ускорение распада по мере «старения» слов, а коэффициент $N(t)$ в показателе степени отражает, напротив, замедление скорости распада по мере выпадения из исходного ОС менее устойчивых слов и сохранения только более устойчивых²³. Заметим, что при малых t (и, соответственно, больших $N(t)$) датировки по формулам (9) и (5) весьма близки: напротив, при увеличении t (и уменьшении $N(t)$) датировки начинают существенно удревняться. Принимая λ в среднем за 0.05 (см. об этом выше, с. 417), приведем таблицу выборочных датировок (обозначения: $N(t)$ — доля сохранившихся слов из ОС в одном языке, $N_2(t)$ — в двух языках, t — время развития и, соответственно, время дивергенции в тысячах лет):

$N(t)$	$N_2(t)$	t	(t , по М. Сводену)
0.99	0.99	0,3	0,03
0.97	0.94	0,8	0,2
0.95	0.9	1	0,35
0.9	0.81	1,5	0,7
0.85	0.72	2	1,1
0.8	0.64	2,4	1,5
0.75	0.56	2,8	1,9
0.7	0.49	3,2	2,4
0.65	0.42	3,7	2,9
0.6	0.36	4,1	3,4
0.55	0.3	4,7	4
0.5	0.25	5,3	4,6
0.45	0.2	6	5,3
0.4	0.16	6,8	6,1
0.35	0.12	7,8	7
0.3	0.9	9	8
0.25	0.06	10,7	9,3
0.2	0.04	12,7	10,7
0.15	0.02	16,6	13
0.1	0.01	21,5	15,3

²³ Подробная критика математического аппарата глоттохронологии была принята уже Д. Кретьеном [Chrétien 1952], но на все возражения последнего вполне обоснованный ответ был предложен уже Дж. Дайеном. Мы поэтому на критике Д. Кретьена останавливаться не будем (достаточно сказать, что возражения Кретьена к самой сущности проблемы «постоянной» или «переменной» скорости изменений отношения не имеют).

Представляется, что датировки по формуле (13) существенно надежнее, чем датировки классической глоттохронологии. Так, для распада белорусского и украинского языков (97% совпадений) мы получаем дату XIV в. н. э.; для различных пар германских, романских, славянских и тюркских языков мы получим даты, соответствующие 1-му тысячелетию н. э.; для распада балтославянских языков мы получаем датировку рубежом первого и второго тысячелетий до н. э. Данные глоттохронологии, конечно, не следует переоценивать (возможны разного рода статистические колебания и возмущения²⁴), однако она представляется все же существенным подспорьем при классификации и оценке степени родства языков.

Все оказанное выше относится к «стандартному» 100-словному списку ОС М. Сводеша. Факт различия индивидуальных коэффициентов сохраняемости слов говорит о том, что в принципе возможно составление списков, отличающихся друг от друга не только общим коэффициентом λ , но и формулами зависимости между скоростью распада и временем. Не исключено, что, варьируя набор слов в списке, можно подбирать списки, удовлетворяющие различным формулам зависимости. Это, однако, весьма кропотливая и вряд ли выполнимая без компьютера работа²⁵.

²⁴ Несомненным возмущающим фактором являются языковые контакты, способные вызывать вторичное сближение уже разошедшихся родственных языков. В таких случаях часто нарушается 5-й постулат глоттохронологии: в активно взаимодействующих родственных языках имеется тенденция к сохранению и/или выпадению из списка ОС одних и тех же слов (при этом далеко не всегда можно говорить о заимствовании). Иногда такие контакты могут вызывать «подскок» доли совпадений даже на 5–6%. Подобную картину мы наблюдаем, например, для белорусского и западнославянских языков, немецкого и скандинавских и пр. Для устранения этого эффекта при оценке степени близости языков существуют некоторые рекомендации, но они носят частный характер и заслуживают специального рассмотрения. Попытка обобщения здесь была предпринята Хаттори Сиро [Hattori 1954], предложившего в зависимости $N_2(t) = e^{-2\lambda t}$ вводить в степени e не коэффициент 2, но переменный коэффициент в зависимости от степени близости языков; однако конкретные способы определения этого коэффициента все же остаются неясными.

²⁵ В качестве эксперимента мы попытались, комбинируя слова из 100-словного и 200-словного списков М. Сводеша, составить 55-словный список, удовлетворяющий классической «радиоактивной» зависимости $N(t) = N_0 e^{-\lambda t}$ при $\lambda = 0.1$ (т. е. при коэффициенте сохраняемости 0.9 за тысячу лет). В этот список вошли слова: кора, живот, большой, черный, кровь, кость, жечь, умирать, собака, сухой, ухо, есть, яйцо, глаз, огонь, нога, полный, волос, голова, я, колено, лист, печень, длинный, много, мясо, луна, близкий, ночь, нос, круглый, короткий, змея, звезда, камень, плавать, хвост, тот, тонкий, ты, язык, дерево, два, вода, мы, что, белый, женщина, червяк, год, желтый, кто, шея,

Классическая лексикостатистика — даже с улучшенным способом датировки — обладает все же рядом существенных недостатков. Известны затруднения, связанные с выбором основного слова при наличии в некотором значении двух или более синонимов (какое слово выбирать в итальянском для 'головой' — *capo* или *testa*?; для 'песка' — *sabbia* или *rena*? и т. п.). Кроме того, при лексикостатистическом анализе в некотором смысле отсутствует статистика как таковая: сравнивая ОС двух языков, мы получаем только один результат, а для повышения надежности выводов хотелось бы иметь серию результатов (для которых можно было бы вычислить математическое ожидание и возможные пределы отклонений).

Мы полагаем, что оба названных недостатка можно преодолеть при помощи методики, которую мы назвали «этимологической статистикой» (или «корневой глоттохронологией»). Сформулируем ее основные постулаты.

1. В каждом языке имеется определенное число исконных, т. е. не заимствованных за период независимого существования языка, корневых морфем (по предварительным оценкам число таких морфем для любого живого разговорного языка не превышает 2–3 тысячи).

2. Такие корневые морфемы характеризуются различной частотностью, т. е. вероятностью встретиться в произвольном языковом тексте.

3. Частотность некоторой корневой морфемы в данном языке в данный момент времени t есть стабильная величина, не зависящая (или мало зависящая) от характера выбираемого текста.

4. Все корневые морфемы обладают свойством «старения», т. е. постепенного снижения частотности до нуля (после чего корень считается исчезнувшим из языка). При этом, однако, скорость выпадения

новый, рот. Этот список мы специально пытались составить так, чтобы он удовлетворял «классической» процедуре глоттохронологии (в частности, заимствования в этом списке учитываются как замены). Этот список (проверенный уже на довольно большом материале) позволяет составлять в целом надежные классификационные схемы и давать датировки, в целом (хотя — из-за меньшего набора слов — нередко с довольно большими отклонениями) совпадающие с датировками по стандартному списку Сводеша; однако при малых временных глубинах датировки получаются все же довольно сильно умоложенными.

Список из 35 «наиболее устойчивых» слов с $\lambda = 0.07$ или 0.08 был составлен С. Е. Яхонтовым; он оказывается весьма полезным при верификации родства языков (процент совпадений в пределах 35 слов для родственных языков должен быть выше, чем процент совпадений в пределах остальной части списка), но плохо пригоден для датировок и классификационных целей.

различных корней не одинакова (как и слова, корни делятся на более устойчивые и менее устойчивые).

5. Корневой состав языка распадается с равномерной скоростью (т. е. из определенного набора корней, характеризующихся заданной частотностью, за промежуток времени Δt выпадает фиксированное количество корней).

Наиболее существенным для излагаемой ниже методики является постулат 3, и он же априорно вызывает наибольшие сомнения. Действительно, характер лексики и распределение частоты слов в текстах разных жанров довольно сильно варьирует (если не говорить о максимально частотных местоимениях и служебных словах). Однако от одного этимологического корня образуется обычно масса производных слов с разнообразными значениями: поскольку частотность корня существенным образом зависит от его продуктивности, постольку наиболее частотные корни, как правило, встречаются в основах, относящихся к самым разнообразным семантическим сферам. Поэтому чем выше продуктивность (и частотность) корня, тем больше шансов встретить его в тексте любого содержания и жанра. Во всяком случае, эмпирические данные явно демонстрируют «внежанровый», нейтральный характер корневого состава любого текста.

5-й постулат «корневой глоттохронологии» аналогичен 3-му постулату обычной лексикостатистики (см. с. 413 выше) и нуждается в эмпирической проверке (см. ниже).

Возьмем некоторую выборку — например, 100 — незаимствованных корневых морфем какого-либо языка А (заметим, что уже эта процедура предполагает наличие предварительного этимологического анализа) и будем подбирать для этих корней этимологические соответствия в родственных языках. Ясно, что в языках, наиболее близких генеалогически к языку А, обнаружится максимальное число соответствий, а по мере удаления языков это число будет снижаться. Естественно ожидать, что русский язык, например, будет иметь больше всего совпадающих корней со славянскими, затем — с балтийскими, затем с прочими индоевропейскими и т. п. Таким образом, уже эта процедура позволяет строить классификацию и относительную хронологизацию дивергенции внутри языковых семей. Анализ нескольких выборок такого рода в принципе должен показывать аналогичные результаты.

Однако — в силу постулатов 2 и 4 — абсолютные числа этимологических сопадений между корнями языка А и родственных ему языков в пределах произвольных выборок корней будут сильно варьировать. Для того, чтобы они приобрели стабильный характер, необходимы выборки

корней, характеризующихся одним и тем же распределением частотности.

Можно было бы создавать частотные словари корней и делать выборки из них (т. е. пытаться проделать с корнями работу, аналогичную проделанной М. В. Араповым и М. М. Херц со словами); однако такая работа довольно трудно осуществима.

Здесь и вступает в силу постулат 3, согласно которому для каждой исконной корневой морфемы языка задана определенная вероятность встретиться в произвольном тексте; соответственно, всякий текст должен характеризоваться определенным — и одним и тем же — распределением вероятностей (или частотностей) входящих в него корневых морфем.

Если это так, то ожидается, что в выборках исконных корневых морфем из различных текстов языка А обнаружится одинаковое или сходное число этимологических совпадений с каждым из родственных языков.

Рассмотрим некоторый русский текст, например, текст данной статьи. Будем выбирать из него незаимствованные корневые морфемы (корни самостоятельных слов: все префиксальные и суффиксальные морфемы, а также имена собственные условимся не учитывать), при этом повторяющиеся учитывать не будем. Для каждой морфемы такого рода будем находить этимологические соответствия в польском, литовском, немецком и французском языках (если такие этимологические соответствия в каждом из этих языков имеются, но, в свою очередь, являются не исконными, а заимствованными, мы их также учитывать не будем).

рус.	пол.	лит.	нем.	фр.	и.-е.
1. в	+	+	+	+	*en-
	(w)	(ĩ)	(in)	(en)	
2. сто(й)	+	+	+	+	*sta-
	(stać)	(stó-ti)	(stehen)	(être)	
3. мог-	+	+	+	—	*mągh-
	(moc)	(mėgti)	(mag)		
4. у-(же)	+	+	+	+	*io-u,
	(juz)	(jaũ)	(ja)	(dé-jà)	*ia-m
5. с, сам	+	+	+	+	*sem-
	(z)	(san-)	(zu-sammen)	(simple etc.)	
6. полн-	+	+	+	+	*pələ(n)-
	(pelnų)	(pilnas)	(voll)	(plein)	

рус.	пол.	лит.	нем.	фр.	и.-е.
7. снов-	+ (snowac)	-	-	+ (neuf)	*sneu-
8. ч-(то), г-(де), к-(ак), с-ко(-лько), (не)-ко(тор-)	+ (co)	+ (ka-s)	+ (wa-s)	+ (que)	*k ^w e- *k ^w i-
9. след-	+ (ślad)	+ (slýsti)	+ (Schlitten)	-	*sleidh-
10. десят-	+ (dziesięc)	+ (dešimt)	+ (zehn)	+ (dix)	*deġmt
11. лет-	+ (lato)	-	-	-	*lēt-
12. шест-	+ (sześć)	+ (šeši)	+ (sechs)	+ (six)	*s(ū)ēks-
13. восьм-	+ (ośm-)	+ (aštuōni)	+ (acht)	+ (huit)	*oktō(u)-
14. год-	+ (gody)	+ (guōdas)	+ (Gatte)	-	*ghodh-
15. шед-	+ (szed-ł)	-	-	-	*sed ²⁶
16. корен-	+ (korzeń)	+ (kēras)	-	-	*k(ū)er-
17. лом-	+ (łomic)	-	+ (lahm)	-	*lem-
18. дл(и)-, дал-	+ (długi)	+ (ilgas)	+ (lang)	+ (long)	*delə-, de- lə-(n)gh-
19. готов-	+ (gotowy)	-	-	-	*got- (?)
20. работ-	+ (robić)	-	+ (Erbe)	-	*Horbh-
21. тек- / теч-	+ (ciec)	+ (tekėti)	+ (dienen)	-	*tek ^w -
22. лик-	+ (lice, liczyc)	-	-	+ (langue)	*leik ²⁷
23. язык	+ (język)	+ (liežūvis)	+ (Zunge)	+ (langue)	*lenġhu- / *denġhu

²⁶ И.-е. *sed- 'ходить, идти' (слав. *ход-/ *šċd-, др.-инд. sad-, гр. ὀδός) мы предпочитаем отдалять от омонимичного *sed- 'сидеть'.

²⁷ Несколько спорен вопрос об объединении *leik- 'лицо' и *leik- 'подходить, пододать'; если это разные (но омонимичные) корни, фр. loisir следует отделять от рус. лик- в 'различный'.

рус.	пол.	лит.	нем.	фр.	и.-е.
24. семьи-	-	+	+	+	* $\text{k}^{\text{h}}\text{ei-m-}$ / * $\text{k}^{\text{h}}\text{ei-}\mu\text{-}$
		(šeima)	(Heim)	(ci-té)	
25. шв- (почва < *родъ-šьва)	+	+	+	+	* $\text{s}\check{\text{j}}\text{eu}\text{a-}$
	(szyć)	(siúti)	(Saum)	(coudre < *cōsē-re < *con-sue-re)	
26. дл-, де-д-	+	+	+	+	*dhē-
	(dla, działo)	(dėl)	(tun)	(faire)	
27. нос-	+	+	+	-	*(e)neĕ-
	(nieść)	(nėšti)	(genug)		
28. род-	+	+	-	-	* $\mu\text{erdh-}$
	(rod)	(rasmė)			
29. и-, е-(го), и-х	+	+	+	+	*e-/i- (местоим. основа)
	(i)	(ji-s)	(e-r, e-s)	(ce < *ecce-iste)	
30. реш-	-	+	? +	-	* $\mu\text{reik-}$
		(raišyti)	(Rist)		
31. шаг- (< *sęg-)	+	-	-	-	*se(n)g ^{wh} - ²⁸
	(sięgas)				
32. (э)-то-, то, то-	+	+	+	+	*to-
	(ten)	(tą-s)	(de-r)	(te-l)	
33. прав-, перв-, при	+	+	+	+	*per, *pr-ō
	(prawy)	(prō)	(ver-, für)	(pour)	
34. бы-	+	+	+	+	*bhe $\mu\text{-}$
	(być)	(būti)	(bauen, bin)	(fu-t)	
35. лаг- / лег- / лог-	+	-	+	+	*legh
	(leżeć)		(liegen)	(lit)	
36. пыт-	+	-	-	+	*pūt-
	(pytać)			(com-pt-er)	
37. труд-	+	+	+	+	*treud-
	(trud)	(triūdas)	(ver-drießen)	(intrus)	
38. на	+	+	+	-	*an-a, *an-ō
	(na)	(nuō)	(an)		
39. ряд-	+	+	-	-	*rend(h)-
	(rząd)	(rinda)			

²⁸ Так в [WP II: 482–483]. В [Фасмер IV: 393–394; III: 825] слав. *sęg- возводится к и.-е. *se(n)g-, откуда — из интересующих нас языков — нем. *Senkel* ‘шнурок (богинка)’. Лит. *sėgti* ‘застегивать’ следует, по-видимому, отделять как от одного, так и от другого корня и сравнивать со слав. *steg-* (с тем же соотношением в анлауте, что и слав. *sterg- : лит. *sergeti* ‘стеречь’).

рус.	пол.	лит.	нем.	фр.	и.-е.
40. стар-	+ (stary)	+ (stóras)	-	-	*stār- ²⁹
41. свет-	+ (świat)	+ (šviēsti)	+ (weiß)	? + (verre)	*k _u eit-
42. нов-	+ (nowy)	+ (naūjas)	+ (neu)	+ (neuf)	*ne _u -
43. бол-	-	-	-	-	*bel- / *bol-
44. древ-н-	+ (drzewni, drzewo)	+ (drėve, dervà)	+ (treu, Teer)	-	*der _u - / *dre _u -
45. кон-[/чн-]	+ koniec	-	-	?+ (ré-cent)	*ken-
46. мер-	+ (miara)	+ (mė-tai)	+ (messen)	+ (mesure)	*mė-
47. мы	+ (my)	+ (mėš)	-	-	*me- / *mo-
48. -им-/-я-	+ (mieć, jęć)	+ (iĩnti)	-	+ (vend-ange, r-an-con)	*em-
49. зна-	+ (znać)	+ (žinóti)	+ (kennen)	+ (connaître)	*ĝnō-
50. близ-	+ (blizki)	-	-	-	*bhleiĝ-
51. тысяч-	+ (tysiąć)	+ (túkstantis)	+ (Tausend)	-	*tūs(k̂)mt ³⁰
52. до, да	+ (do)	+ (-dà)	+ (zu)	+ (de)	*dō- / *dē-
53. на-/ш/	+ (nasz)	-	+ (uns)	+ (nous)	*nō-(s)/ *n̥-s
54. по, позд-	+ (po)	+ (pa-, pō)	+ (ab)	+ (po-ndre)	*apo / *pō
55. цел-	+ (cały)	-	+ (heil)	-	*kail-
56. спе-	+ (śpiech, śpiać)	+ (spėti)	+ (sputen)	+ (ésperer)	*spē-

²⁹ Не исключена производность этой основы от *stā- 'стоять' (№ 2), хотя это не очень надежно.

³⁰ Если *tūs(k̂)mt действительно в конечном счете восходит к корню *teuə- 'быть толстым, сильным', то из фр. можно привлечь *tout* и другие формы.

рус.	пол.	лит.	нем.	фр.	и.-е.
57. да-	+ (dać)	+ (duóti)	-	+ (donner)	*dō-
58. глуб-	+ (głęboki)	-	+ (klauben)	-	*gleubh-
59. кол-	+ (koło)	+ (kāklas)	-	+ (quenouille)	*k ^w el-
60. пят-	+ (pięć)	+ (penkì)	+ (fünf)	+ (cinq)	*penk ^w e-
61. ес-/ть/	+ (jest)	+ (es-mì)	+ (is-t)	+ (es-t)	*es-
62. чит-	+ (czytać)	+ (skaiťyti)	-	-	*k ^w eit-
63. вид-	+ (widzieć)	+ (-vísti)	+ (wissen)	+ (voir)	*uēid-
64. ст-	+ (sto)	+ (šiñtas)	+ (Hundert)	+ (cent)	*k̥mt-
65. верх-	+ (wierzch)	+ (viršùs)	-	+ (verrue)	*uērs-
66. бег-	+ (bieżec)	+ (bėg-ti)	-	-	*bhēg ^w -
67. прос-	+ (prosić)	+ (prašyti)	+ (fragen)	+ (prier)	*prek-
68. вяз-	+ (wiązać)	-	-	-	* _(u) engh ⁻³¹
69. ли	+ (li)	+ (-li)	-	-	*le(i)-
70. со-б- / се-б-	+ (sie-bie)	+ (sāvo)	+ (sich)	+ (se)	*s _(u) e-
71. из	+ (z)	+ (iš)	-	+ (é-)	*eĝhs
72. мне-	+ (po-mniec)	+ (miñti)	+ (mahnen)	+ (mentir)	*men-
73. вет-	-	+ (vaitenù)	-	-	*ueit-
74. од/и/н-	+ (jeden)	+ (vīenas)	+ (ein)	+ (un)	*oĭ-n-
75. мног-	+ (mnóstwo)	? + (minià)	+ (manch)	-	*menegh-
76. мест-	+ (miasto)	+ (maĩstas)	-	-	*mejt-

³¹ Связь с и.-е. *anĝh- 'узкий, суживать' сомнительна фонетически, см. [Фасмер: I, 374].

рус.	пол.	лит.	нем.	фр.	и.-е.
77. зем-	+ (ziemia)	+ (žẽmẽ)	-	-	*ghõem-
78. шар	-	-	-	-	?
79. же	+ (-že)	+ (-gu, -gi)	-	+ (-hui)	*ghe-
80. вис-	+ (wiesieć)	-	-	-	?
81. друг-	+ (drugi)	+ (draũgas)	-	-	*dhreugh-
82. от	+ (od)	+ (at-)	-	-	*at-i
83. яв-	+ (jawić)	+ (ovyje)	-	(?) + (in-oui)	*ãu-
84. рек-	+ (rzec)	+ (rẽkti)	-	-	*rek-
85. вес-	+ (wszy-stki)	+ (visas)	-	-	*ui[s]-
86. о/б/	+ (/o(b)/)	-	(?) + (bei)	+ (oub-lier)	*(o)bhi
87. чело-/век/	+ (człowiek)	? + (kiltis)	-	-	*k ^w el-
88. /чело-/век	+ (człowiek)	+ (vaĩkas)	-	-	*uok-
89. мысл-	+ (myśl)	+ (maũsti)	-	-	*meudh-
90. хот-	+ (chcieć)	-	-	-	?
91. долг-	+ (dług)	-	-	-	*dh _l gh-
92. каз-	+ (kazać)	-	-	-	*k ^w ek- / *k ^w eğ-
93. ровн-	+ (równy)	-	+ (Raum)	+ (rustre)	*reu- /er _u -
94. втор-	+ (wtóry)	-	? + (wider-)	-	*ui-ter- (?) ³²
95. руж-	-	-	-	-	*roug-
96. з-/де-сь/	+ (/dzi/-ś)	+ (šis)	+ (hier, hin)	+ (ce)	*ke- / *ki-

³² Имеется несколько этимологий слав. *vьtorь (*vьtorь) из *ui-tero, *η-tero- или *u-tero [Фасмер: I, 364–365]; неясно, какой из них следует отдать предпочтение.

рус.	пол.	лит.	нем.	фр.	и.-е.
97. блюд-	+	+	+	-	*bheudh-
	(budzić)	(baūsti)	(bieten)		
98. им/ен-/	+	-	+	+	*(e)nomŋ-
	(imię)		(Name)	(nom)	
99. два	+	+	+	+	*d̥uō(u)
	(dwa)	(du)	(zwei)	(deux)	
100. ви-	+	+	-	+	*uejə-
	(wi-)	(vų-ti)		(vis, vrille)	

Таким образом, в данном отрезке текста на 100 корневых заимствованных морфем³³ оказалось (если сомнительные этимологии принимать за 0.5%) 93 корня, имеющих соответствия в польском; 72 корня, имеющих соответствие в литовском; 55 корней, имеющих соответствия в немецком; и 51.5 корней, имеющих соответствия во французском.

Оказывается, что эти цифры довольно стабильны. Сравним данные подсчетов по другим русским текстам (в скобках указаны номера текстов, см. ниже. с. 447):

	польский	литовский	немецкий	французский
(4)	98	77	55	52
(1)	96	76	58	50
(5)	97	70	51.5	50

Итак, мы видим, что цифры здесь в каждом случае варьируют вокруг некоторого статистического среднего: для польского это $95 \pm 3\%$, для литовского — $74 \pm 3\%$, для французского — $54 \pm 3\%$.

³³ Из ста выбранных морфем три (*шар, вис-* и *хот-*) не имеют и.-е. этимологии: для первой из них отсутствует также надежная славянская этимология. Тем не менее заимствованный их характер не доказан, и мы оставили их в списке; в любом случае число их слишком мало для того, чтобы повлиять на результаты анализа. Напротив, из списка мы исключили большое количество основ, заимствованный характер которых надежно установлен, а именно (в порядке встречаемости): *время, утверждать, практика, сравнительный, исторический, период, компаративистский / компаративистика, генетический, между, группа, ностратический, строгий, методика, гениальный, интуиция, лингвистика, рубеж, реконструированный / реконструкция, цель, воссоздать, макро-, если, аксиоматический, модель, теория, дивергенция, вынужден, факт, существование, территория, ойкумена, эра, мотивировать, спонтанный, возникновение, элементарный, здравый, классический, система, регулярный, фонетический, масса, лексический, морфологический, использовать, обций, оппонент.*

Приведем еще некоторые данные по различным парам языков (язык, текст которого анализируется, будем называть «язык текста», а язык, со словарем которого производится сравнение, — «язык словаря»).

Язык текста	Язык словаря	Доля совпадений	Текст
французский	русский	0,50	(7)
немецкий	русский	0,55	(6)
немецкий	литовский	0,57	(6)
латинский	русский	0,50	(8)
латинский	немецкий	0,55	(8)
латинский	литовский	0,53	(8)
др.-греческий	др.-индийский	0,69	(9)
др.-греческий	латинский	0,67	(9)
др.-греческий	русский	0,52	(9)
др.-греческий	немецкий	0,54	(9)
др.-греческий	литовский	0,53	(9)
латинский	др.-индийский	0,71	(8)
латинский	др.-греческий	0,72	(8)
русский	литовский	0,76	(3)
русский	литовский	0,74	(2)
др.-индийский	др.-греческий	0,77	(10)
др.-индийский	латинский	0,64	(10)
др.-индийский	русский	0,54	(10)
др.-индийский	литовский	0,53	(10)
др.-индийский	немецкий	0,57	(10)

Мы видим, что при довольно высокой стабильности цифр наблюдается интересный феномен: при подсчетах от «языка текста» к «языку словаря» и в обратную сторону (т. е. если «язык словаря» становится «языком текста») цифры в целом совпадают у синхронных языков; цифры повышаются при подсчете от современного «языка текста» к древнему «языку словаря», что представляется закономерным. Однако подсчеты от древних «языков текста» к современным «языкам словаря» обнаруживают практически те же самые цифры, что и при сравнении современных языков. Можно сформулировать, следовательно, важное правило: возраст текста не влияет на результат статистического анализа.

При ближайшем рассмотрении этот результат оказывается вполне понятным: действительно, мы — по исходным условиям — каждый раз измеряем, собственно, не расстояние от одного языка до другого,

но лишь расстояние от «праязыка» до «языка словаря», и получаемые цифры, следовательно, характеризуют только это расстояние.

Проведем теперь еще один опыт: возьмем в качестве текста 100-словный список М. Сводеша и будем проделывать с ним те же операции (т. е. выписывать подряд исконные корни, не учитывая заимствования и повторения). Проведем анализ на материале тех же языков — польского, литовского, немецкого, французского в сравнении с русским.

	русс.	пол.	лит.	нем.	фр.	и.-е.
1.	бел-	+	+	-	-	*bhel-
		(biały)	(baltas)			
2.	бол-	-	-	-	-	*bel- / *bol-
3.	вес-	+	+	-	-	*u[s]-
		(wszy-stki)	(visas)			
4.	вид-	+	+	+	+	*ueid-
		(widzieć)	(-vísti)	(wissen)	(voir)	
5.	вод-	+	+	+	+	*ued-
		(woda)	(vanduõ)	(Wasser)	(onde)	
6.	волос-	+	-	-	-	*uolk-
		(włos)				
7.	вошь	+	+	? +	-	*ues? *lius-
		(wesz)	(vievesa)	(Laus)		
8.	глаз	+	-	-	-	*gloĝ-(?)
		(głas)				
9.	голов-	+	+	-	-	*ghelṽ-
		(głowa)	(galvà)			
10.	гор-	+	+	-	-	*g ^w er-
		(góra)	(girià)			
11.	груд-	-	-	-	+	*g ^w rendh-
					(grand)	
12.	да-	+	+	-	+	*dō-
		(dać)	(duóti)		(donner)	
13.	два-	+	+	+	+	*dṽō(u)
		(dwa)	(dù)	(zwei)	(deux)	
14.	дерев-	+	+	+	-	*derṽ-
		(drzewo)	(dervà)	(treu, Teer)		
15.	дл-	+	+	+	+	*delə-,
		(długi)	(ilgas)	(lang)	(long)	*delə-(n)gh-
16.	дождь	+	+	-	+	*deju-
	(< *dus-dius)	(deszcz)	(diēvas)		(dieu)	

русс.	пол.	лит.	нем.	фр.	и.-е.
17. дорог-	+ (droga)	-	+ (tragen)	-	*dh(e)ragh-
18. дым	+ (dym)	+ (dūmai)	-	+ (fumée)	*dhū-mo-
19. ед-	+ (jeść)	+ (ėsti)	+ (essen)	-	*ed-
20. желт-	+ (zółty)	+ (geļtas)	+ (Gold)	+ (fiel)	*gh ^w el-
21. жен-	+ (żona)	-	-	-	*g ^w en-
22. ж/е/г-	-	+ (dėgti)	+ (Tag)	+ (fièvre)	*dheg ^w h-
23. жи-/р/, жи-/вот/	+ (żyć)	+ (gyti)	+ (er-quicken)	+ (vivre)	*g ^w ei(ʏ)-
24. звезд-	+ (gwiazda)	+ (žvaigzde)	-	-	*gh ^w eisd-
25. зелен-	+ (zielony)	+ (žalias)	+ (gelb)	-	*ghel-
26. зем-	+ (ziemia)	+ (žėme)	-	-	*gh ^h ēm-
27. зна-	+ (znać)	+ (žinóti)	+ (kennen)	+ (connaître)	*gnō-
28. зола-	+ (zoła)	-	+ (Kohle)	-	*g ^u el-
29. зүб-	+ (ząb)	+ (žėmbti)	+ (Kamm)	? + (gemme)	*g ^u embh-
30. имя	+ (imię)	-	+ (Name)	+ (nom)	*(e)nom ^u
31. камень	+ (kamień)	+ (akmuõ)	+ (Kammer)	-	*akm- / *kām-
32. кож-/коз-/	+ (koża)	-	-	-	*kag-
33. колен-	+ (kolano)	+ (kelỹs)	? + (Schulter)	-	*kel-
34. кор-	+ (kora)	+ (karnà)	-	+ (écorce)	*(s)ker-
35. корен-	+ (korzeń)	+ (kėras)	-	-	*k(ʏ)er-
36. кост-	+ (kość)	-	-	+ (côte)	*kost-

русс.	пол.	лит.	нем.	фр.	и.-е.
37. крас-	+ (krasa)	-	? + (Ruhm)	-	*kerə- ?
38. кровь	+ (krew)	+ (kraūjas)	+ (roh)	+ (cru)	*kreu-
39. круг-	+ (kraḡ)	-	+ (Ring)	-	*krengh-
40. к-/го/, ч-/го/	+ (kto, co)	+ (ka-s)	+ (wa-s)	+ (que)	*k ^w e- / *k ^w i-
41. кус-	+ (kąsac)	+ (kąsti)	-	-	*kənəd-
42. лег-	+ (ležeć)	-	+ (liegen)	+ (lit)	*legh-
43. лет-	+ (latac)	+ (lėkti)	-	-	*lek-
44. лист-	+ (list)	+ (láiskas)	-	-	*leis(t)-
45. лун-	+ (łuna)	+ (laūkas)	+ (Licht)	+ (lune)	*leuk-
46. мал-	+ (mały)	-	+ (schmal)	+ (mal)	*(s)mal-
47. мног-	+ (mnóstwo)	? + (minià)	+ (manch)	-	*menegh-
48. муж-	+ (mąż)	-	+ (Mann)	-	*mon
49. мы-	+ (my)	+ (mės)	-	-	*me- / *mo-
50. мясо	+ (mięso)	+ (mēsà)	-	+ (membre)	*mēms-
51. не	+ (nie)	+ (nè)	+ (nicht)	+ (non)	*ne
52. нов-	+ (nowy)	+ (naūjas)	+ (neu)	+ (neuf)	*neu-
53. ног-	+ (noga)	+ (nagà, nāgas)	+ (Nagel)	+ (ongle)	*(o)nogh ^w -
54. нос	+ (nos)	+ (nósis)	+ (Nase)	+ (nez)	*nas
55. ночь	+ (noc)	+ (naktis)	+ (Nacht)	+ (nuit)	*nok ^w t-
56. ог/о/н-	+ (ogień)	+ (ugnìs)	-	-	*[e]gni-

русс.	пол.	лит.	нем.	фр.	и.-е.
57. /од/-ин-	+ (jeden)	+ (v́ienas)	+ (ein)	+ (un)	*oĭ-n-
58. пер-	+ (pióro)	? + (spaĩnas)	+ (Feder)	+ (panne)	*pet- / *pter- ³⁴
59. пес-/ок/	+ (piasek)	-	-	-	*pēns(u)
60. печ/ен/	+ (pieczén)	+ (kēpenas)	-	? + (cuire)	*pek ^w -
61. пи-	+ (pić)	-	-	+ (boire)	*pō(i)-
62. плы-	+ (pywać)	+ (pláuti)	+ (Flut)	+ (pluie)	*pleu-
63. полн-	+ (pełny)	+ (pilnas)	+ (voll)	+ (plein)	*pələ(n)-
64. /при/-ид-	+ (iść)	+ (eĩti)	-	+ (i-ra)	*ei-
65. пт-/иц/-	+ (ptak)	+ (putýtis)	-	+ (petit)	*peut-
66. рог-	+ (róg)	+ (rāgas)	-	-	*rog(h)-
67. р/о/т	+ (ryć)	+ (ráuti)	-	+ (ruer)	*reu(ə)-
68. рук-	+ (ręka)	+ (rankà)	-	-	*ronk(ɯ)-
69. рыб-	+ (ryba)	-	+ (Raupe)	-	*rūbh-
70. се/-мя/	+ (siemię)	+ (sémens)	+ (Same)	+ (semence)	*sē-(men-)
71. серд-/ц/-	+ (serce)	+ (širdis)	+ (Herz)	+ (cœur)	*k̥rd-
72. сид-	+ (siedzieć)	+ (sēdėti)	+ (sitzen)	+ [(as)-sis]	*sed-
73. каз-	+ (kazać)	-	-	-	*k ^w ek- / *k ^w eg-
74. слы-/ш/-	+ (słyszeć)	+ (klausyti)	+ (laut, lauschen)	-	*k̥leu-

³⁴ Ввиду греч. πτερόν представляется (все же) целесообразным объединять и.-е. названия 'пера', первоначально, видимо, связанные с и.-е. *pet- 'летать'. Ввиду развития *pter- > *per- в ряде языков (в т. ч. в балтославянских) возможны были, однако, контаминации с и.-е. корнем *per- 'двигать/ся/' (рус. 'парить', нем. *fahren* etc.).

	русс.	пол.	лит.	нем.	фр.	и.-е.
75.	солн-/ц-/	+	+	+	+	*s _u el-
		(słońce)	(sáule)	(Sonne)	(soleil)	
76.	сп-	+	+	-	+	*s _u ep-
		(spać)	(sãpnas)		(sommeil)	
77.	сто/й/	+	+	+	+	*stā-
		(stać)	(stó-ti)	(stehen)	(être)	
78.	сух-	+	+	-	-	*saus-
		(suchy)	(saūšas)			
79.	теп-/л/	+	-	-	+	*tep-
		(ciepły)			(tiède)	
80.	т-/от/, /э/-т/от/	+	+	+	+	*to-
		(ten)	(tā-s)	(de-r)	(te-l)	
81.	ты	+	+	+	+	*tu
		(ty)	(tu)	(du)	(tu)	
82.	/y/-би-	+	-	+	-	*bhei(e)-
		(bić)		(Beil)		
83.	/y/-мер-	+	+	+	+	*mer-
		(mrzeć)	(miřti)	(Mord)	(mourir)	
84.	ух-	+	+	+	+	*ōus-
		(ucho)	(ausis)	(Ohr)	(oreille)	
85.	хвост-	+	-	-	-	?
		(chwoszcz)				
86.	ход-	+	-	-	-	*sed-
		(chodzić)				
87.	холод-	+	+	-	-	*řkol(dh)- ³⁵
		(chłód)	(řáltas)			
88.	хорош-	-	-	-	-	?
89.	чело-/век/	+	?+	-	-	*k ^w el-
		(człowiek)	(kiltis)			
90.	/чело-/век	+	+	-	-	*uoi _k -
		(człowiek)	(vaĩkas)			
91.	черн-	+	+	-	-	*kers-
		(czarny)	(kerřas)			
92.	шей-	+	+	+	+	*sjeuā-
		(szyja)	(siūti)	(Saum)	(coudre)	
93.	я	+	+	+	+	*eğō
		(ja)	(ař)	(ich)	(je)	

³⁵ Реконструкция и.-е. корня со значением 'холодный' представляет очень большие затруднения. Не исключено, что следует восстанавливать варьирование *keldh- (*řkelt) ~ *ř(h)el-; в таком случае со слав. *холод* будут объединяться также нем. *kalt* и лат. *gelidus, gelare* (фр. *geler*) [WP: I, 622, 629–630; ЭСРЯ: IV, 256].

	русс.	пол.	лит.	нем.	фр.	и.-е.
94.	язык	+	+	+	+	*lenĝhū- /
		(język)	(liežūvis)	(Zunge)	(langue)	*denĝhū
95.	яй/-ц-/	+	-	+	+	*ōujō-
		(jaje)		(Ei)	(œuf)	

На 95 корневых морфем русского стословного списка³⁶ мы имеем 91 совпадение с польским ($\approx 95.8\%$); 68.5 совпадений с литовским ($\approx 72.1\%$); 50.5 совпадений с немецким ($\approx 53.2\%$); 50 совпадений с французским ($\approx 52.6\%$).

Таким образом, мы получили практически те же самые результаты, что и при обследовании выборок русских корней, взятых из произвольных текстов. Это, по-видимому, говорит о том, что распределение индивидуальных частотных характеристик корней в списке Сводеша совпадает с их обычным распределением в тексте³⁷. Из этого можно сделать несколько важных выводов:

а) «устойчивость» корней не зависит от «устойчивости» слов, производных от них (это вытекает прежде всего из третьего постулата корневой глоттохронологии, но также хорошо демонстрируется на разобранном выше примере со списком Сводеша: слова, входящие в него, заведомо «устойчивее», чем большинство слов в произвольном тексте, а «устойчивость» корней оказывается такой же);

б) при отсутствии текстов (а, к сожалению, для довольно многих малоизученных языков отсутствуют текстовые записи, но имеются лишь небольшие словники) 100-словный список М. Сводеша может быть использован в качестве текста для корневой глоттохронологии;

в) математический аппарат, разработанный для классической глоттохронологии, может быть перенесен и на корневую (естественно, с выбором иного коэффициента λ).

³⁶ Из списка выброшены два заимствования (*облако* и *собака*): объединены однокоренные пары *жир, живот; кто, что; нога, ноготь; тот, этот* (дейктический префикс *э* мы не принимаем за корень). В одной статье стословного списка (*человек*), напротив, обнаруживается сложение из двух корней, каждый из которых мы оцениваем отдельно.

³⁷ Причина этого явления, вероятно, в том, что список Сводеша включает в себя наиболее распространенные и обычные понятия из самых разных семантических сфер, которые, таким образом, действительно образуют некий элементарный концептуальный «текст» (на тему «человек и его окружение»). Характерно, что из корней русского языка, имеющих в списке Сводеша, довольно многие (25!) представлены также и в разобранном выше (с. 431) тексте, а именно: *сто[й]-, полн-, ч-[то] / к-[то], шед-/ход-, корен-, язык-, ши- [шей-/шив-], дл- [долг-], т- [от], лег-, нов-, бол-, мы, зна-, да-, вид-, [од]-ин-, мног-, зем-, вес-, чел[о]-, -век-, каз-, им[ен]-, два.*

Последнее утверждение — самое сильное, и оно еще нуждается в основательной практической проверке. По предварительным данным, однако, время, вычисленное по доле корневых совпадений с применением формулы

$$(14) t = \sqrt{\frac{\ln(N(t)/N_0)}{-\lambda N(t)}} \text{ при } \lambda = 0.035^{38}$$

в целом совпадает с вычисленным по классической глоттохронологической процедуре с применением формулы (13). Ср. даты дивергенции русского соответственно с польским, литовским, немецким и французским языками (время в тысячах лет):

	пол.	лит.	нем.	фр.
t по классической глоттохронологии	1,3	3,1	4,7	4,7
t по корневой глоттохронологии	1,2	3,2	4,9	5,1 ³⁹

³⁸ В последнее время Д. Лещинер предложил иную формулу для датировки языковой дивергенции по методике корневой глоттохронологии:

$$t = \frac{1 - \ln(N(t))}{1 - \ln(1-N(t))} * T$$

где T (период «полураспада») принимается равным 5. Эта формула, по нашим наблюдениям, дает оптимальные датировочные результаты.

³⁹ В русском и польском не совпадают следующие слова 100-словного списка: *живот* — *brzuch*, *большой* — *wielki (duzy)*, *грудь* — *piers*, *жечь* — *palić*, *глаз* — *oko*, *жир* — *tłuszcz*, *хороший* — *dobry*, *печень* — *wątroba*, *много* — *wiele (duzo)*, *луна* — *księżyc*, *рот* — *usta*, *красный* — *czerwony*, *сказать* — *powiedzieć (rzec)*, *кожа* — *skóra*, *хвост* — *ogon*, *женщина* — *kobieta*, *холодный* — *zimny*. При учете двух заимствований в русском списке (*облако* и *собака*) это дает нам 85% совпадений.

В русском и литовском совпадают следующие слова 100-словного списка: *весь* — *visas*, *кусать* — *kąsti*, *кровь* — *krūjas*, *жечь* — *dėgti*, *ноготь* — *nāgas*, *холодный* — *šaltas*, *прийти* — *ateiti*, *умереть* — *mir̃ti*, *сухой* — *sausas*, *ухо* — *ausis*, *земля* — *žemė*, *огонь* — *ignis*, *лететь* — *lėkti*, *полный* — *pilnas*, *дать* — *dūoti*, *зеленый* — *žalias*, *рука* — *ranka*, *голова* — *galva*, *сердце* — *širdis*, *рог* — *nāgas*, *я* — *aš*, *знать* — *žinoti*, *колено* — *kėlis*, *печень* — *kėpenys*, *длинный* — *ilgas*, *мясо* — *mėsa*, *новый* — *naūjas*, *ночь* — *naktis*, *нос* — *nosis*, *не* — *ne*, *один* — *vienas*, *семья* — *sėkla*, *сидеть* — *sėdėti*, *дым* — *dūmai*, *стоять* — *stovėti*, *звезда* — *žvaigždė*, *камень* — *akmuo*, *солнце* — *saulė*, *плавать* — *plaukti*, *том* — *tàs*, *ты* — *tu*, *язык* — *liežuvis*, *два* — *dù*, *вода* — *vanduo*, *мы* — *mēs*, *кто* — *kàs*, *белый* — *baitas*, *что* — *kàs*, *желтый* — *geltonas*, т. е. (при неучете слов *облако* и *собака*) — ровно 50% совпадений.

Некоторые из этих датировок, возможно, все же еще несколько умоложены (это касается прежде всего русско-польской дивергенции, где повышение процентов совпадений, возможно, объясняется вторичными контактами: другие славянские данные позволяют удревнить датировку славянской дивергенции на 300 лет), однако в целом они представляются достаточно правдоподобными.

Методика «корневой глоттохронологии» имеет некоторые преимущества перед классической глоттохронологией: в частности, отсутствуют затруднения, связанные с отбором «основного» слова из нескольких синонимов; возможен статистический анализ большого количества результатов (поскольку количество текстов — в отличие от списков ОС — в принципе неограничено)⁴⁰. Можно предвидеть, однако, что этимостатистический анализ столкнется с большими трудностями при обработке мало обследованных (в сравнительно-историческом плане) языков. Следует, однако, подчеркнуть, что, по нашему мнению, любые лексикостатистические и этимостатистические штудии возможны только после этапа основательной сравнительно-исторической обработки языкового материала и ни в коей мере не могут подменить собой компаративистику.

В настоящее время этимостатистика уже успешно применяется на семитском и афразийском языковом материале (А. Ю. Милитарев), а также на австронезийском материале (Ю. Х. Сирк).

В русском и немецком совпадают следующие слова стословного списка: *ноготь* — *Nagel*, *ухо* — *Ohr*, *есть* — *essen*, *яйцо* — *Ei*, *перо* — *Feder*, *полный* — *voll*, *сердце* — *Herz*, *я* — *ich*, *знать* — *kennen*, *лежать* — *liegen*, *длинный* — *lang*, *мужчина* — *Mann*, *имя* — *Name*, *новый* — *neu*, *ночь* — *Nacht*, *нос* — *Nase*, *не* — *nicht*, *один* — *ein*, *семья* — *Same*, *сидеть* — *sitzen*, *стоять* — *stehen*, *солнце* — *Sonne*, *этот* — *dieser*, *ты* — *du*, *язык* — *Zunge*, *два* — *zwei*, *вода* — *Wasser*, *что* — *was*, *кто* — *was*. При учете двух заимствований в русском (*облако*, *собака*) и трех заимствований в немецком (*Fett*, *Kopf*, *rund*) списках это дает нам 30% совпадений.

В русском и французском совпадают следующие слова стословного списка: *кора* — *écorce*, *ноготь* — *ongle*, *умирать* — *mourir*, *пить* — *boire*, *ухо* — *oreille*, *яйцо* — *œuf*, *полный* — *plein*, *дать* — *donner*, *сердце* — *cœur*, *я* — *je*, *знать* — *connaître*, *длинный* — *long*, *луна* — *lune*, *имя* — *nom*, *новый* — *neuf*, *ночь* — *nuît*, *нос* — *nez*, *не* — *ne*, *один* — *un*, *видеть* — *voir*, *сидеть* — (*être*) *assis*, *дым* — *fumée*, *стоять* — *être* (*debout*), *солнце* — *soleil*, *ты* — *tu*, *язык* — *langue*, *два* — *deux*, *что* — *que*, *кто* — *qui*. При учете двух заимствований в русском (*облако*, *собака*) и трех заимствований во французском (*chemin*, *Pierre*, *blanc*) это дает нам опять же 30% совпадений.

⁴⁰ Еще одной областью применения корневой глоттохронологии может стать датировка текстов. Это, однако, отдельная область со своими проблемами, которые мы здесь обсуждать не будем.

Как представляется, комбинация лексикостатистики и «этимостатистики» позволит добиться уточнения датировок и классификационных схем как обычных семей, так и «макросемей». По предварительным данным, современные ностратические языки по методике корневой глоттохронологии обнаруживают 15–20% совпадающих корней (при сравнении произвольных текстов с достаточно надежными словарями). Это существенно большие величины, чем совпадения между современными ностратическими языками в 100-словном списке (в среднем 5–9%), и мы думаем, что в дальнейшем — после уточнения сравнительной фонетики и этимологии — корневая глоттохронология должна сыграть важную роль в обосновании теорий дальнего родства языков и построении генеалогических классификаций.

Тексты, использованные для этимологической статистики

1. Мих. Зоценко. Избранное в двух томах. Л. 1982. т. 1, с. 31–32.
2. Е. Молоховец. Подарок молодым хозяйкам, или Средство к уменьшению расходов в домашнем хозяйстве. СПб., 1904, с. 288.
3. Культурное наследие Востока (проблемы, поиски, суждения). Л., 1985, с. 3–4.
4. О. М. Фрейденберг. Миф и литература древности. М., 1978, с. 206–207.
5. Корней Чуковский. Стихи и сказки. М., 1984, с. 37–39.
6. Heinrich Boll. Wanderer, kommst du nach Spa... (Erzählungen). München, 1971, S. 7–8.
7. Albert Dauzat. Dictionnaire etymologique de la langue française. Paris, 1938, pp. V–VI.
8. G. Julius Caesar. Commentarii de bello Gallico (liber primus). М., 1946, pp. 37–48.
9. Sophoclis tragoediae. М., 1884, pp. 111–112.
10. Т. Aufrecht. Die Hymnen des Rigveda. Erster Teil. Berlin, 1955, S. 1–2.