

МЕТОД ГЛОТТОХРОНОЛОГИИ В КУЛЬТУРОЛОГИЧЕСКОМ АСПЕКТЕ

В. А. Калмыков

Нижегородский государственный лингвистический университет имени Н. А. Добролюбова

Л. М. Осокина

*Национальный исследовательский университет «Высшая школа экономики»
(Нижегородский филиал)*

Поступила в редакцию 29 июля 2014 г.

Аннотация: в статье оценивается метод глоттохронологии в его нынешнем виде, анализируются полученные результаты с целью обнаружения всех преимуществ и недостатков метода в свете последних тенденций языковедческих исследований, предпринимается попытка его усовершенствования.

Ключевые слова: глоттохронология, лексикостатистическое датирование, базовая лексика, праязык, родственные языки, совпадающие слова, несовпадающие слова, время расхождения языков, новые языки.

Abstract: the article evaluates the method of glottochronology as it is at the moment; analyses the results in order to find out the advantages and disadvantages of the method in the wake of the latest linguistic trends; attempts to improve it.

Key words: glottochronology, lexicostatistic dating, basic vocabulary, protolanguage, kindred languages, convergences, divergences, language divergence time, new languages.

Настоящая статья является попыткой дать оценку глоттохронологии как метода, при помощи которого можно изучать концептуальные картины мира. При этом подчеркивается связь метода глоттохронологии и культуры, что не нашло отражения в работах таких выдающихся лингвистов, как М. Сводеш, В. В. Иванов, Г. А. Климов, М. В. Арапов, М. М. Херц, В. А. Звегинцев, С. А. Старостин, Г. Хойер, С. Е. Яхонтов, В. В. Кромер и др.

Новизна данного подхода заключается не в попытке изменить широко известный метод и его математический инструментарий, но в стремлении добавить к нему культурологический компонент, с целью оптимизации его валидности при установлении языковых особенностей.

Применение математического инструментария для определения в абсолютных цифрах эпох начала дифференциации языков, восходящих к одному языку-основе, нашло выражение в методе глоттохронологии, или лексикостатистическом методе датировки доисторических дивергенций внутри праязыковых единств.

Глоттохронология (*glotta* «язык», *chronos* «время» и *logos* «слово, учение») – область сравнительно-исторического языкознания, занимающаяся выявлением скорости языковых изменений и определением на этом основании времени разделения родственных языков и степени близости между ними [1, с. 109].

М. Сводеш, основатель глоттохронологии, предпринял достаточно успешную попытку определения объективного метода для установления момента расхождения языков, сущность которого определяется лексикостатистической доминантой [1, с. 452–463].

При этом лексикостатистическая глоттохронология определяет время разделения родственных языков, исходя из предположения об одинаковой скорости изменения той основной части словаря, которая нужна для обслуживания наиболее часто встречающихся и существенных ситуаций общения.

Согласно принципам реализации метода, представленным В. А. Звегинцевым, «степень сохранности элементов основного лексического ядра постоянна на протяжении всего времени», а «процент утраты слов основного ядра одинаков во всех языках». При этом необходимо учитывать, что фактический процент позволяет вычислить время начала их дивергенции [2, с. 12].

Рассматривая список слов базовой лексики, представляется возможным утверждать, что за одно тысячелетие сохраняется в среднем не менее 80 % словаря. Поэтому, сопоставив процент сохранившихся родственных слов в основных списках двух языков одной семьи, можно определить наименьшее время их разделения t по формуле, которая была предложена Р. Лисом [3, с. 56]:

$$t = \log C/2 \log r,$$

где C – доля совпадающих слов в основном списке, r – коэффициент, характеризующий степень сохран-

ности основного списка за интервал времени (принимаемый за $r = 81$ или 86%).

В отечественном языкознании С. А. Старостин предложил такой способ глоттохронологии, который предполагает подсчет не слов, а корневых морфем (основ), сохраняющихся в текстах определенной длины. Вычисления по уточненной методике дают лучшее приближение к датам, основанным на исторических данных. Вместе с тем установлен факт «старения» лексики языка, влекущий необходимость при оперировании древними языками до их сопоставления с современными введения определенного количественного коэффициента (поправки на «старение») [4, с. 789–791].

Р. Лис считал, что при однолинейном развитии, т.е. в истории одного языка, время t , за которое в языке сохранился определенный процент первоначального состава основного словаря C , вычисляется по равенству:

$$t = \log C / \log r,$$

где r – индекс сохранения слов основного словаря за одно тысячелетие; при взаимной дифференциации из общего источника двух языков $t_1 = t_2$ и, следовательно:

$$\begin{aligned} t_1 + t_2 &= \log C / \log r, \\ 2t &= \log C / \log r, \\ t &= \log C / 2\log r \text{ [3, с. 56–57].} \end{aligned}$$

Первое, что обращает на себя внимание, – это отсутствие видимых причин использования логарифмов для достижения поставленных целей. Чтобы ответить на данный вопрос, необходимо самим вывести корректную формулу и попытаться применить ее, поскольку только в этом случае мы будем избавлены от необходимости делать догадки об эффективности этого метода.

В наших рассуждениях мы отталкиваемся от того, что существует r , индекс сохранения слов основного словаря, равный $80\text{--}85\%$.

Предположим, что мы имеем какой-либо язык А, тогда через тысячу лет из 100% слов его базовой лексики останется 80% ; еще через одну тысячу лет оставшиеся от предыдущего этапа 80% сохранят опять же 80% и т.д. Для иллюстрации этой динамики приведем табл. 1.

Т а б л и ц а 1

0	1	2	3	4
1	0,8	0,64	0,512	0,4096

Поясним: $C_1 = C_0 \times 0,8$, где $C_0 = 1$, так как праязык мы приняли за 100% . Таким образом:

$$C_1 = 1 \times 0,8,$$

$$\begin{aligned} C_2 &= 1 \times 0,8 \times 0,8, \\ C_3 &= 1 \times 0,8 \times 0,8 \times 0,8, \\ C_4 &= 1 \times 0,8 \times 0,8 \times 0,8 \times 0,8 \text{ и т.д.} \end{aligned}$$

Таким образом, любое C , выявленное нами в списках слов базовой лексики, есть не что иное, как $0,8$ в степени x . В виде формулы наш вывод выглядит так:

$$C = 0,8^x, \text{ т.е. } C = r^x,$$

при этом x будет равен t . Отсюда получаем, что $t = \log_{0,8} C$ или $t = \log_r C$.

Эту формулу можно значительно упростить, если выразить ее в десятичных логарифмах:

$$t = \log_r C = \log C / \log r.$$

Таким образом, нами построена та самая формула, которая фигурирует в источниках:

$$T = \log C / \log r.$$

Благодаря этому разбору математической части метода возможно, на наш взгляд, уточнить и выявить ошибку в расчетах некоторых ученых-лингвистов, которые недостаточно корректно с математической точкой зрения исследовали данный вопрос. Указанную выше формулу они интерпретировали как пропорцию, что, конечно же, не совсем верно, и поспешили заявить о ее непригодности. Пример ошибочного вывода можно найти у таких авторов, как Я. В. Лоя [5, с. 177], Ф. М. Березин, Б. Н. Головин [2, с. 306]. Они рекомендовали перевернутую формулу, подкрепляя свои рассуждения законом М. Сводеша, в соответствии с которым время, прошедшее с момента разделения языка на два родственных диалекта (а затем и языка), обратно пропорционально проценту совпадений слов опытного списка в этих языках. В. В. Кромер ввел свою собственную математическую модель при исследовании лингвистических явлений на разных этапах их развития, а также понятие «диахронический скачок», определяющее динамику языкового процесса в зависимости от его статистических данных типологического характера. Модель В. В. Кромера характеризует решающий фактор развития определенного языкового явления, но не сам процесс во всей временной протяженности. Исследования В. В. Кромера важны в плане определения «диахронического скачка» как переломного момента в развитии [4].

С. Е. Яхонтов использует формулу М. Сводеша, но вводит уточнение в виде понятия о «стандартном отклонении», значение которого зависит от общего числа слов и процентов их совпадения в двух сравниваемых языках.

Уточняя рассматриваемую формулу М. Сводеша, следует отметить некорректность цифры 2 в ее знаменателе $t_1 + t_2 = \log C / \log r$:

$$2t = \log C / \log r;$$

$$t = \log C / 2\log r.$$

С нашей точки зрения, ее избыточность подтверждается нижеследующими аргументами.

Во-первых, если мы подставим гипотетический результат C , равный 80 %, то мы по теории должны получить результат t , равный 1 (т.е. одной тысяче лет). На практике же эта формула дает результат лишь 0,5 (т.е. пятьсот лет назад).

Во-вторых, если $t_1 = t_2$, то это означает, что $t_1 = \log C / \log r$ и $t_2 = \log C / \log r$. Из этого следует, что при сложении t_1 и t_2 мы получаем $2t = 2 (\log C / \log r)$. Далее цифры «два» можно сократить как в числителе, так и в знаменателе. Окончательная формула будет иметь вид $t = \log C / \log r$. Проверив ее подстановкой, мы получим результат, предсказанный теорией, т.е. цифру 1, что соответственно составляет одну тысячу лет.

Для прояснения вопроса о возможных результатах применения формулы рассмотрим рис. 1:

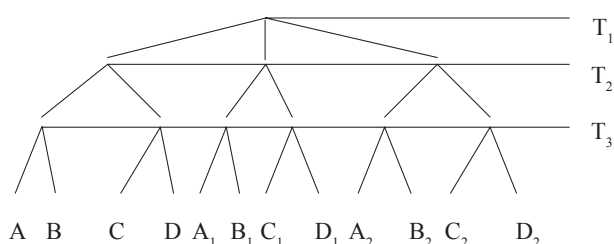


Рис. 1

К анализу подобного дерева нас подталкивает возможность получения различных результатов при сравнении родственных языков друг с другом. В представленном выше дереве мы можем получить три различных результата T_1, T_2, T_3 .

Результат T_3 мы получаем при сравнении таких близких языков, как А и В, С и D, A_1 и B_1, C_1 и D_1, A_2 и B_2, C_2 и D_2 .

Результат T_2 мы получаем при сравнении таких языков, как А и С, А и D, В и С, В и D и т.д.

Результат T_1 мы получаем при сравнении таких языков, как А и A_1, A и B_1, A и C_1, A и D_1, B и A_1, B и B_1 и т.д.

Для нас результаты T_1, T_2, T_3 имеют вполне очевидное значение, поскольку они подразумевают, что исследуемые языки имеют три «ступени» совпадений в списках базовой лексики, которые обусловлены историческим процессом их разделения.

В реальности дерево может оказаться другим, но это не меняет сути рассуждения.

Также для оценки эффективности метода глоттохронологии необходимо привлечь праязык, поскольку он поможет нам установить принцип срав-

нения списков базовой лексики. Метод выявления и подсчета несовпадающих пар в списках представляет особую важность в проводимой нами работе, поэтому нам важно выявить все возможные недостатки и найти пути их устранения.

Для начала предположим, что мы имеем два списка слов базовой лексики, один – принадлежащий праязыку (А), а другой – взятый из словаря современного языка (В). В этом случае при их сравнении мы получим гипотетическую ситуацию, отраженную в табл. 2, где \bullet – совпадающие слова, а \circ – несовпадающие.

Исходя из табл. 2, мы можем определить время разъединения с абсолютной точностью, так как 20 % расхождений будут приблизительно соответствовать одной тысяче лет.

Самое же интересное заключается в том, что в нашем случае мы гарантированно знаем, что среди несовпадающих пар несоответствия вызваны инородными элементами в родственном языке, которые появились в нем с течением времени.

В данном случае авторитет праязыка неоспорим, поскольку он стал прародителем языка В и не подвергался никаким изменениям в течение времени, так как в нашем понимании праязык – это условное понятие, которое характеризует состояние того или иного языка на определенный момент в прошлом.

Результативность рассмотренной ситуации мы сможем полностью оценить далее, когда перейдем к ситуации с двумя современными языками.

Ситуация с двумя современными языками

В табл. 3 в чисто гипотетической ситуации мы находим 60 % совпадений, что будет соответствовать точке во времени приблизительно 2,3 тысяч лет назад. Однако нас в данной ситуации больше интересуют несовпадения, так как среди них могут быть и слова из праязыка. Если их не учесть, то результаты могут оказаться искаженными.

Т а б л и ц а 2

A	B
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●

Т а б л и ц а 3

A	B
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●
●	●

Известно, что в каждой паре несовпадающих слов может содержаться по одному слову из праязыка. Если их будет больше, то пара автоматически будет учтена как совпадающая.

Состав несовпадающих пар представляет особую важность для метода глоттохронологии, вследствие того, что изменения в языках могли идти не параллельно: в одном языке могло измениться слово, которое обозначало жилище, а в другом – слово, обозначающее родственные отношения. Всё это могло произойти за один и тот же промежуток времени. Такая ситуация объясняется уникальными, свойственными для каждого народа в отдельности, историческими условиями.

Таким образом, не имея возможности заглянуть внутрь слов, мы вынуждены обращаться за помощью к логике. Для иллюстрации снова прибегнем к помощи таблиц.

Т а б л и ц а 4

○	●
○	●
○	●
○	●

Т а б л и ц а 5

●	○
○	●
○	●
○	●

Т а б л и ц а 6

●	○
●	○
○	●
○	●

Т а б л и ц а 7

●	○
●	○
●	○
○	●

Т а б л и ц а 8

●	○
●	○
●	○
●	○

Варианты табл. 4 и табл. 8 можно теоретически исключить, ввиду того, что они предполагают существование языка, который за историю не изменился.

Варианты табл. 5 и табл. 7, по сути, представляются одинаковыми и предполагают, что один из языков имеет 90 %, а другой – 70 % сходств. Однако по теории они должны быть равны. Чтобы не противоречить теории, мы можем взять среднее число, т.е. 80 % совпадений.

Вариант табл. 6 предполагает, что оба языка сохранили по 80 % базовой лексики.

При сравнении полученных результатов с первоначальными обнаруживается значительное отклонение, которое равно 20 %. А это значит, что время

выделения этих двух условных языков будет равно не 2,3, а 1 тысяче лет назад. Как видно, ошибка в 1,3 лет является очень существенной.

В ходе рассуждений очень важно не терять из виду праязык. Мысленно сравнивая оба языка с праязыком, который неизменен, мы получаем одно время t не на двоих, а сразу для каждого из них в отдельности. Именно этот аргумент и является главным в наших доводах против цифры 2 в знаменателе формулы Р. Лиса. Еще раз хочется подчеркнуть те предпосылки, которые нас подтолкнули к такому радикальному выводу:

- время t приблизительно равно для всех родственных языков;
- скорость и количество изменений в этих родственных языках за одно и то же t должны тоже быть приблизительно равны;
- мы фактически сравниваем праязык с неким средним современным языком, который как бы содержит в себе все конкретные современные родственные языки;
- чем больше родственных языков участвует в анализе, тем выше точность получаемых результатов.

Ситуация с несколькими языками

Обратимся к табл. 9, где отражена ситуация с несколькими языками: А – праязык, В, С, D, Е – современные языки. Средний процент совпадений с праязыком – 75 %. Среднее t выделения для современных родственных языков $B, C, D, E = \log 0,75 / \log 0,8 = 1,28$ тысячи лет назад.

Т а б л и ц а 9

А	В	С	D	Е
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	●	○	○	○
○	●	●	○	○
○	●	●	●	○
○	●	●	●	●

С представленной в табл. 9 ситуацией в реальности мы не столкнемся, так как у нас не будет столбца А (праязыка) для сравнения с другими столбцами (современными родственными языками). В реальности мы не сможем идентифицировать совпадение слова в девятой строчке у языка Е. Для нас оно будет просто черной точкой. Безусловно, мы никак не сможем учесть в наших расчетах, что неизбежно повле-

чет некоторое искажение результата вычислений. Теперь нам необходимо быть внимательными, так как зона несовпадений, начинающаяся с шестой строчки табл. 9, таит в себе много интересного.

С седьмой строчкой все достаточно ясно – три совпадающих слова повлияют на средний результат.

В восьмой строчке все тоже достаточно ясно, если только мы не имеем ситуацию, где два совпадающих слова (восходящих к праязыку) противопоставляются двум другим (также совпадающим словам, но не восходящим к праязыку). В такой ситуации нет принципиальной разницы, поскольку 50%-ный результат совпадений обязательно отразится на общем среднем результате. Да и сама ситуация, при которой в словари базовой лексики сразу нескольких, хотя и родственных, языков входят слова какого-то одного неродственного языка, очень редка. Следовательно, такая возможность маловероятна.

Таким образом, применяя опробованный нами способ вычислений, получаем средний процент совпадений, который равен 70,25 %. Далее по формуле мы получаем:

$$t = \log 0,70 / \log 0,8 = 1,598.$$

Как видим, разница в результатах составляет около 300 лет.

Все вероятные недочеты и ошибки с нашей стороны могут значительно исказить полученные результаты. Кроме того, нам показалось, что сама идея определять время разделения родственных языков с точностью до года не является конструктивной. Представляется более разумным говорить о некотором периоде времени, в течение которого произошли достаточные изменения в языках, которые позволяют сделать окончательный вывод о том, что появились новые языки. Хотя, как и в предыдущем случае, прочертить точную границу, где кончается диалект и где начинается новый язык, практически невозможно. Поэтому идея о промежутке времени получает дополнительную актуальность. Проверим на практике истинность наших предположений.

Во-первых, мы имеем первую границу этого временного периода в виде группы полных совпадений во всех сравниваемых языках (в нашем случае это шесть строчек полных совпадений). Переведем их в цифры, мы получим $(\log 0,6 / \log 0,8 = 2,289)$ t , равное приблизительно 2,4 тысяч лет. Данный результат подтверждается подсчетами, выполненными С. Е. Яхонтовым по предложенной им формуле стандартных отклонений [6, с. 134].

Во-вторых, эта цифра для нас является твердой, так как в ее основе лежат неоспоримые результаты полных совпадений. Согласно исследованиям С. Е. Яхонтова, совпадения составляют более 50 %. Однако в нашей работе мы учитываем и неполные

совпадения, что дает нам основание утверждать, что в качестве второй границы периода дивергенции языков представляется возможным взять t , полученные из среднего результата 70 %, включающего и неполные совпадения. Это t , как мы уже выяснили, равно приблизительно 1,6 тысячи лет.

В-третьих, данная цифра не является для нас постоянной, так как результат может колебаться в зависимости от точности наших вычислений.

Для наглядности полученных результатов используем числовую ось на рис. 2, где 0 – это начало нашей эры, 1 – тысячный год нашей эры, 2 – двухтысячный год:

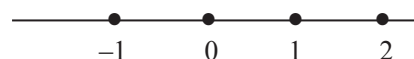


Рис. 2

Теперь отобразим полученные ранее результаты на этой оси. 2,4 тысячи лет назад – это первая граница временного периода, а 1,6 тысячи лет – вторая. Для их отображения на этой оси необходимо понимать, что эти границы надо откладывать с момента настоящего времени (в нашем случае это 2010 г.).

Таким образом, период разделения языков в нашем экспериментальном случае будет укладываться в рамках периода с 400-х гг. до нашей эры и до 400-х гг. нашей эры. Другими словами, процесс их разделения произошел за период в 800 лет (рис. 3).

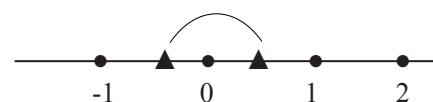


Рис. 3

При использовании формулы, предложенной М. Сводешом и Р. Лисом, получается, что дивергенция польского и русского языков произошла в XVI–XVII вв., что противоречит общеизвестным историческим фактам развития двух народов и языков. Значит, есть все основания предположить, что одной лексикостатистикой невозможно объяснить и сделать корректную оценку развития языков и их дивергенции. Так, С. Е. Яхонтов, рассматривающий глоттохронологию как лексикостатистику при определении возраста родственных языков по количеству сохранившихся в них слов общего происхождения, отмечает наличие других факторов, способных влиять на момент дивергенции языков. К таким факторам он относит математические ошибки, собственно лингвистический анализ и главное – существование в определенных языках степеней вежливости, способных привести к ускорению изменения лексики. Следовательно, требуется учитывать культурологический компонент, который является единственным и реша-

ющим в реализации метода глоттохронологии. С учетом рассмотренных в данной статье поправок, дивергенция русского и польского языков произошла в VIII–IX вв., что соответствует известным историческим данным. В связи с этим В. В. Кромер подчеркивает зависимость отдельных слов от информационного пространства, которое он определяет как «плотность упаковки лексики» в процессе дивергенции [3]. В. В. Кромер объясняет имеющий место «диахронический скачок» употребительности любого лингвистического явления на различных этапах его истории моментом «перелома» лингвистической ситуации, который способен «растягивать процесс во времени», что выходит за рамки чисто математических выкладок [4].

Таким образом, при анализе и оценке литературы по данному вопросу становится очевидным, что данный метод может быть перспективным только при учете культурологической составляющей, которая не принималась в расчет основателем глоттохронологии.

Нижегородский государственный лингвистический университет имени Н. А. Добролюбова

*Калмыков В. А., кандидат филологических наук,
доцент кафедры английского языка
E-mail: kalmykov1950@mail.ru
Тел.: 8-910-121-42-39*

*Национальный исследовательский университет
«Высшая школа экономики» (Нижегородский филиал)*

*Осокина Л. М., старший преподаватель кафедры
иностранных языков
E-mail: lilichka_1@rambler.ru
Тел.: 8-905-865-98-44*

ЛИТЕРАТУРА

1. Большой энциклопедический словарь языкознания / гл. ред. В. Н. Ярцева. 2-е изд. – М. : Большая Российская энциклопедия, 1998. – С. 109.
2. *Иванов В. В.* Глоттохронология / В. В. Иванов // Языкознание / гл. ред. В. Н. Ярцева. – М. : Большая Российская Энциклопедия, 1998.
3. *Кромер В. В.* Об одной поправке к каноническому закону / В. В. Кромер. – Режим доступа: http://old.kpfu.ru/gum_konf/ot2t.htm
4. *Сводеш М.* К вопросу о повышении точности в лексикостатистическом датировании / М. Сводеш // Новое в лингвистике. – Вып. 1. – М., 1960. – С. 53–87.
5. *Кромер В. В.* Арктангенс или логиста? / В. В. Кромер. – Режим доступа: http://www.philol.msu.ru/~humlang/articles/pilot_law.htm
6. *Сводеш М.* Лексикостатистическое датирование доисторических этнических контактов / М. Сводеш // Новое в лингвистике. – Вып. 1. – М., 1960. – С. 23–52.

Nizhny Novgorod State Linguistic University named after N. A. Dobrolyubov

*Kalmykov V. A., Candidate of Philology, Associate Professor of the English Language Department
E-mail: kalmykov1950@mail.ru
Tel.: 8-910-121-42-39*

National Research University «Higher School of Economics» (Nizhny Novgorod Branch)

*Osokina L. M., Senior Lecturer of Foreign Languages Department
E-mail: lilichka_1@rambler.ru
Tel.: 8-905-865-98-44*