М. Е. Васильев[†], *М. Н. Саенко*[‡]
[†] Институт славяноведения РАН (Россия, Москва); mvhumanity@gmail.com,
[‡] Институт славяноведения РАН (Россия, Москва); veraetatis@yandex.ru

К вопросу о точности глоттохронологии: датирование языковой дивергенции по данным романских языков

Статья продолжает начатое раннее исследование, посвященное определению точности лигвистических датировок, получаемых с помощью глоттохронологии. Предметом рассмотрения является датирование языковой дивергенции (процесса разделения двух или нескольких идиомов), которое производится на материале 110-словных списков современных романских языков. Предметом рассмотрения является датирование языковой дивергенции — т. е. определение даты разделения двух или нескольких современных языков. В статье сопоставляются как традиционные, так и вновь предложенные модели глоттохронологии. При этом особое внимание уделяется величине погрешности и надёжности глоттохронологических вычислений на различных временных глубинах. Результаты проведенного исследования позволяют не только количественно оценить границы точности глоттохронологии при датировании романских языков, но также сделать ряд практических выводов, касающихся возможностей применения глоттохронологии на любом другом языковом материале.

Ключевые слова: глоттохронология, лексикостатистика, список Сводеша, романские языки

Данная статья является продолжением нашего исследования, цель которого — оценка точности и надежности лингвистических датировок, получаемых с использованием глоттохронологических расчетов. Первая часть работы (Васильев, Саенко 2016) была посвящена рассмотрению лексических изменений, происходящих в списке базисной лексики одного языка с течением времени, и определению временной дистанции между языком-предком и его потомками на основе нескольких различных глоттохронологических методов. Во второй части мы обратимся к процессу языковой дивергенции — т. е. независимому изменению лексики родственных идиомов после их разделения.

Датирование языковой дивергенции представляет наибольший практический интерес для сравнительно-исторического языкознания (в частности, при изучении дальнего языкового родства), так как дает возможность получить датированные генеалогические деревья и тем самым — сопоставить результаты праязыковой реконструкции с историческими или доисторическими событиями, не имеющими письменных свидетельств. При этом глоттохронология, несмотря на многочисленные критические замечания, до сих остается наиболее популярным, а в случае с малоизученными семьями языков — единственно доступным методом для получения лингвистических датировок. Нередко глоттохронологические датировки, опубликованные в узкопрофильных статьях и имеющие, как правило, лишь предварительный, оценочный характер, используются также в научных и научно-популярных работах по смежным дисциплинам (история, археология и др.), что способствует их распространению и популяризации за пределами сравнительно-исторического языкознания. Зачастую такие данные преподносятся читателю уже как установленный научный факт, подтвержденный строгим лингвистическим методом[1].

[1] См., например, (Алексеев 2013: 63; Багаев 2015: 127). Более того, даже специалисты, известные критическим отношением к глоттохронологии, нередко сами пользуются её результатами в своих работах. См. например, работы *Л.* С. Клейна (Клейн 2010: 33–36, 122, 305–306, 466).

В этих обстоятельствах исследование точности глоттохронологических моделей приобретает особое значение как для профессионального лингвистического сообщества, так и для широкой аудитории, поскольку, с одной стороны, оно позволит специалистам получить представление о надежности и границах применимости метода при датировании языковой дивергенции, а неспециалистам — поможет избежать лишних разочарований, связанных главным образом с непониманием «действительных и мнимых»[2] возможностей глоттохронологии.

В настоящей статье представлена попытка такого исследования, выполненная на актуальном романском материале, представленном списками базисной лексики 56 языков и диалектов. Для датирования языковой дивергенции нами будут использованы три различных глоттохронологических метода: классическая глоттохронология М. Сводеша, усовершенствованная методика, разработанная С. А. Старостиным, и недавно предложенная модель, основанная на потоковом подходе к процессу лексических замен. При этом основные цели и задачи исследования останутся теми же, что и в первой части работы:

1. Сопоставить результаты применения известных глоттохронологических моделей (М. Сводеша, С. А. Старостина, потоковой модели) на романском материале.
2. Определить точность полученных датировок и при необходимости провести калибровку рассматриваемых моделей по имеющимся исходным данным (опорным точкам).
3. Оценить фактическую и теоретическую погрешность рассматриваемых моделей при датировании дивергенции между романскими идиомами и группами идиомов.

По итогам рассмотрения мы сделаем выводы о пределах точности и практической ценности глоттохронологических расчетов при датировании романских языков, а также о возможностях применения глоттохронологии на другом языковом материале и бóльших временных глубинах.

1. Исходные данные

Обе части нашего исследования проводятся на материале обновленной лексикостатистической базы романских языков, включающей в себя этимологизированные 110-словные списки 52-х современных и 4-х исторических литературных идиомов (архаическая и классическая латынь, староитальянский и старофранцузский)[3]. Используя приведенные в базе проценты совпадений[4] между парами или группами языков, а также сведения о дате их разделения, известные из экстралингвистических источников, сформируем набор исходных данных (или «опорных точек»), необходимых для измерения скорости расхождения языков, а также определения других параметров дивергенции — см. табл. 1. В полученной таблице для каждой сравниваемой пары идиомов (или групп идиомов[5])

[2] Из заглавия тезисов к докладу В. М. Иллич-Свитыча: «Мнимые и действительные возможности лексикостатистики» (Иллич-Свитыч 1966).

[3] Подробнее о составе и принципах формирования базы, а также методике сбора списков см. в первой части работы — Васильев, Саенко 2016: 262–263.

[4] Полная таблица попарных совпадений между романскими идиомами приводится в дополнительных материалах к вышеуказанной статье (там же), которые доступны на сайте журнала www.jolr.ru.

[5] При сравнении нескольких пар языков (строки 5, 7, 8), в таблице приводится минимальное, максимальное и среднее арифметическое значение совпадений между соответствующими списками. Среднее арифметическое значение рассчитывается как сумма всех полученных процентов совпадений, деленная на количество слагаемых. Подробнее см. в сноске 2 к статье Васильев, Саенко 2016: 263.

Таблица 1. Исходные данные для определения скорости дивергенции романских языков
(по данным 110-словных списков для 54-х романских идиомов)

№	Сравниваемые списки	Мин. % совп.	Средний % совп.	Макс. % совп.	Дата разделения, год	Время дивергенции, лет
1	*Любые идиомы*	-	100	-	0	0
2	Португальский — галисийский	-	97	-	1400	600
3	Старофранцузский (Кретьен де Труа, 1140 г. н.э.) — староитальянский (*Данте*, 1270 г.)	-	91	-	480	790*
4	Румынский — арумынский	-	89	-	900	1100
5	Франко-провансальский — пикардский, валлонский	86	87,5	89	850	1150
6	Португальский – кастильский испанский	-	86	-	710	1290
7**	Португальский — фриульский, ладинские, руманшские, итало-романские, франко-провансальский, окситанский, галло-романские	71	78,2	83	480	1520
	Каталанские — фриульский, ладинские, руманшские, итало-романские, галло-романские	73	79,3	85		
	Пикардский и валлонский — фриульский, ладинские, руманшские, итало-романские, каталанские, иберийские	69	78,0	83		
	Общее значение:	69	78,5	85		
8	Балкано-романские — остальные романские	61	69,4	78	270	1730

* Значение рассчитано относительно даты фиксации для языка *Данте* — 1270 г.

** Следует отдельно пояснить выбор идиомов в строке 7. Поскольку основное ядро романских языков представляет собой практически непрерывный диалектный континуум, процент совпадений между географически соседними идиомами может иметь некий «подскок» в силу наличия ареальных инноваций. Для смягчения нежелательного влияния завышенных значений на результаты глоттохронологических расчетов использовался следующий прием: при рассмотрении какого-либо звена диалектный цепи соседствующие с ним звенья исключались и проводилось сравнение только с географически не примыкающими идиомами. Например, португальский сопоставлялся со всеми остальными «ядерными» (т. е. всеми, кроме балкано-романских и сардинских) языками, кроме соседствующих с ним идиомов Испании.

указаны проценты совпадений между их 110-словными списками, предполагаемая дата их разделения, реконструируемая по экстралингвистическим данным, а также время их развития после разделения — т. е. собственно дивергенции.

Например, доля совпадений между списками франко-провансальского и пикардского составляет 89%, франко-провансальского и валлонского — 86%. Сложив оба значения и разделив сумму пополам, получаем среднее значение — 87,5%, представленное в таблице.

Для датирования разделения этих и других идиомов были выбраны исторические события, в значительной мере повлиявшие на историю романского мира, а тем самым — и на судьбу носителей романских языков. К примеру, разделение франко-провансальского с пикардским и валлонским, по всей вероятности, следует связывать с обособлением Прованса в ходе распада империи Каролингов в 850–860-х гг. Аналогичным образом можно соотнести отделение балкано-романской группы от основного массива романских языков — с выводом римских легионов из Дакии в 271 г.; распад основной романской общности — с крушением Западной Римской империи в 476 г.; расхождение португальского и кастильского испанского — с арабским завоеванием Пиренейского полуострова в 711–718 гг. Разделение румынского и арумынского в 900 г. датируется на основе свидетельств об упоминании арумын в качестве отдельной этнической группы в византийских хрониках IX в.[6]

Разумеется, любые попытки установить хронологическую корреляцию между лингвистическими изменениями и историческим контекстом могут вызвать обоснованные возражения. Прежде всего очевидно, что начало языковой дивергенции не всегда связано с переломными историческими событиями (такими как завоевания, миграции, природные бедствия и т. д.), а может происходить вследствие внутренних культурных, социальных, экономических и других причин. Таким образом разделение языков может как предшествовать физическому разделению их носителей, так и произойти спустя некоторое (иногда — продолжительное) время после него — например, при условии поддержания культурных контактов между ними[7].

Во-вторых, сам термин «дата разделения» можно применять лишь условно, поскольку в действительности начало дивергенции не является одномоментным событием, а представляет собой процесс, происходящий постепенно с течением времени. Иначе говоря, было бы корректнее говорить не о «дате», а о некотором «периоде» дивергенции, по завершении которого мы можем зафиксировать те или иные различия, свидетельствующие о независимом развитии идиомов. При этом, как показывают конкретные случаи дивергенции, момент первой замены не всегда являются надежным свидетельством начала разделения[8]. Если же связывать фактическое разделение с накоплением определённого числа различий между идиомами, то встаёт вопрос, какое количество (или качество) различий считать критическим.

В то же время следует отметить, что все известные на сегодняшний день глоттохронологические модели были получены с использованием (пусть и в имплицитном виде)

[6] См. подробнее в Нарумов 2001: 638.

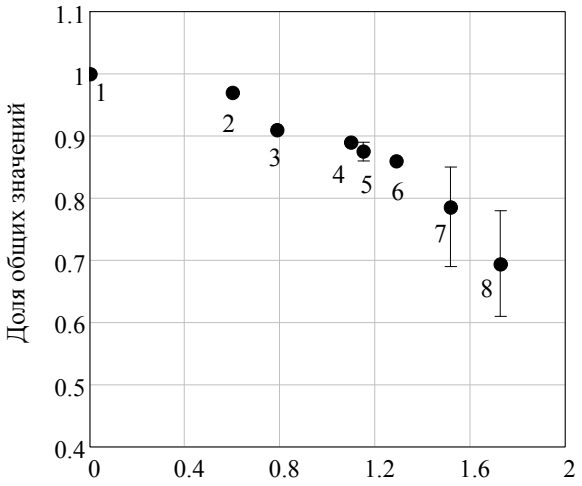
[7] Этот факт, в частности, может проявляться в значительном разбросе процентов совпадений, полученных для одной и той же точки исходных данных.

[8] Например, в сардинских идиомах континуанты *magnus* по-прежнему являются основным словом для *big*, в то время, как в остальном романском мире *magnus* было вытеснено *grandis*. Мы знаем, что эта замена является довольно старой, поскольку для языка Апулея базовым следует считать именно *grandis*, а не *magnus*. Однако связывать отделение сардинских от основного ядра романских языков с заменой *magnus* > *grandis*, было бы неправильным: в сардинских мы находим целый ряд более поздних романских инноваций (*ignis* > *focus*; *iecur* > *ficatum*; *vir* > *homo*; *os* > *bucca*; *cutis* > *pellis*; *brevis* > *curtus* и др.).

допущения о некотором *моменте времени*, соответствующем началу разделения языков в ходе дивергенции. При этом поиск и калибровка параметров моделей производилась, как правило, на основе исторических или доисторических (например, полученных с помощью археологии) сведений о жизни носителей рассматриваемых языков. Очевидным образом отказ от данного допущения и методики калибровки привел бы к невозможности получения числовых параметров моделей, а следовательно — к невозможности применения глоттохронологии в целом. Поэтому, осознавая всю проблематичность и несовершенство данного подхода, мы, тем не менее, должны признать его использование целесообразным и оправданным в рамках нашего исследования.

Полученные исходные данные можно представить в виде диаграммы, на которой каждая опорная точка соответствует строке таблицы 1 с тем же номером.

Рисунок 1. Изменение доли совпадений в базисной лексике романских языков в зависимости от времени дивергенции. Для точек 5, 7 и 8 показан диапазон разброса долей совпадений и среднее арифметическое значение.



На рисунке видно, что наблюдаемый процесс дивергенции (так же как и процесс изменения лексики одного языка, рассмотренный нами ранее[9]), имеет статистическую природу. В частности, для точки 7 доля совпадений между различными идиомами с одной и той же предполагаемой датой разделения (480 г.) варьируется от 69 до 85 %[10], что указывает на вероятностный характер лексических замен. Отметим также, что все опорные точки лежат в относительно небольшом временном диапазоне (до 2000 лет), которым и будут в основном ограничены хронологические рамки нашего рассмотрения. При этом мы наблюдаем компактное расположение опорных точек вдоль некоторой линии регрессии[11] на всем рассматриваемом интервале времени, что дает основания говорить о наличии *значимой* статистической зависимости между долей совпадений в лексике разделившихся языков и временем их дивергенции. Определение свойств этой зависимости мы будем производить на основе трех различных глоттохронологических методов,

[9] Ср. с аналогичной диаграммой на рис. 2 (Васильев, Саенко 2016: 264–265).

[10] Соответствующий диапазон для точки 8 (270 г.) еще шире — от 61 до 78%.

[11] Кривая, наиболее точно отражающая распределение экспериментальных данных.

два из которых (методика М. Сводеша и С. А. Старостина) хорошо известны и уже рассматривались в первой части нашей работы, а третий (потоковая модель) предложен относительно недавно[12]. Для этого мы сопоставим значения каждой модели с исходными данными и при необходимости проведем калибровку их параметров, а затем сделаем выводы о соответствии или несоответствии полученных моделей общему характеру процесса дивергенции и его особенностям.

2. Анализ глоттохронологических моделей

2.1. Глоттохронология М. Сводеша

В соответствии с методикой М. Сводеша развитие языков-потомков после распада их общего предка (праязыка), происходит независимо друг от друга, что отражено в одном из главных постулатов классической глоттохронологии:

> Вероятность того, что слово из О[сновного] С[писка] праязыка сохранится в О[сновном] С[писке] одного языка-потомка, не зависит от вероятности его сохранения в аналогичном списке другого языка-потомка (Арапов, Херц 1974: 25).

Данное утверждение позволило Сводешу перейти от общего уравнения глоттохронологии, имеющего вид $N_{Sw}(t) = e^{-\lambda \cdot t}$, к модели дивергенции двух или нескольких языков-потомков путем возведения исходного выражения в соответствующую степень. В частности, для двух идиомов мы получаем формулу:

$$N2_{Sw}(t) = N_{Sw}(t)^2 = e^{-2\lambda \cdot t}.$$

Таким образом, «скорость» дивергенции двух языков относительно друг друга (2λ) оказывается вдвое больше «скорости» изменения одного языка относительно своего предка (λ), что соответствует принятому утверждению о независимом развитии идиомов после их разделения.

Подставляя в качестве «коэффициента потерь» λ[13] значение 0,16, предложенное Сводешем[14], получим итоговую модель:

$$N2_{Sw}(t) = e^{-2 \cdot 0{,}16 \cdot t}$$

Используя данную формулу, можно подсчитать время дивергенции двух родственных языков (t), по известному проценту совпадений между их основными списками (N). Например, согласно этой модели разделение румынского и арумынского с долей совпадения 89% (N=0,89), должно было произойти около 370 лет назад:

---

[12] Впервые данная методика была описана в статье Васильев, Милитарёв, 2008: 509–536.

[13] Коэффициент потерь (λ) в формуле Сводеша определяет темп замен в базисной лексике языка: чем больше λ, тем больший процент значений изменится в списке за определенный промежуток времени. Не следует путать «коэффициент потерь» с «коэффициентом сохраняемости» (r), который также часто используется в работах по глоттохронологии и означает долю слов, сохранившихся (т.е. оставшихся неизменными) в списке за 1000 лет.

[14] Данное значение λ соответствует «коэффициенту сохраняемости» r=0,85, первоначально вычисленному Сводешем для 200-словных списков (Сводеш 1960: 34). Позднее величина λ неоднократно уточнялась и корректировалась (в том числе — по стословным спискам). Тем не менее, наибольшую известность приобрело именно исходное значение, которое еще долго использовалось в дискуссиях как сторонниками, так и критиками глоттохронологии. Подробнее см. в Васильев, Саенко 2016: 260–261.

$$t = -\frac{\ln(N)}{2 \cdot \lambda} = -\frac{\ln 0{,}89}{2 \cdot 0{,}16} = \frac{0{,}117}{0{,}32} = 0{,}366 \text{ тыс. лет},$$

— т. е. примерно в XVII в., что существенно позже предполагаемой даты — IX в. (см. табл. 1, строка 4).

Результаты аналогичных расчетов, проведенных для диапазона возможных значений N, представлены на рис. 2.

Рисунок 2. Сравнение модели М. Сводеша с исходными данными: $N2_{Sw}(t) = e^{-2 \cdot 0{,}16 \cdot t}$.

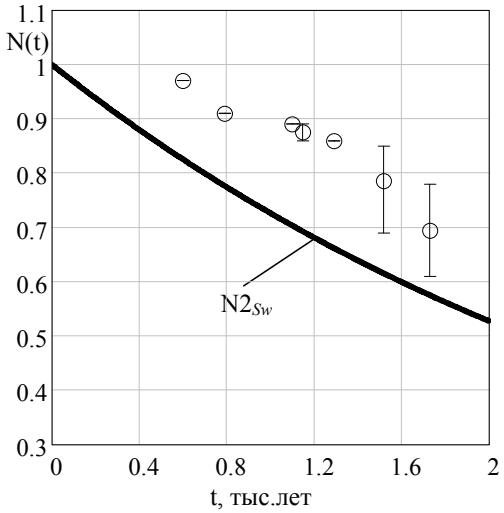


Рисунок показывает, что на всем рассматриваемом интервале использование модели приводит к существенному (в 2 и более раз) «омоложению» расчетных датировок по отношению к предполагаемым датам разделения. Величина отклонения выглядит особенно внушительно при сравнении с результатами, полученными при моделировании процесса замен в лексике одного языка[15], где применение формулы Сводеша с тем же коэффициентом λ=0,16 и на том же романском материале обеспечило очень хорошее совпадение расчетных и фактических значений[16].

Для повышения точности модели попробуем провести калибровку коэффициента λ по имеющимся исходным данным. Для этого воспользуемся уже знакомым нам методом наименьших квадратов[17]. Смысл метода заключается в поиске такого значения λ, при котором суммарное отклонение (ε) между фактическими и расчетными долями совпадений, вычисленное для всех опорных точек, окажется минимальным. В общем виде формулу для поиска оптимального значения λ можно представить следующим образом:

$$\varepsilon = \sum_{i} (N_{p,i} - N_{\phi,i})^2 \rightarrow \min,$$

где N_p — расчетное значение доли совпадений, вычисленное по формуле $N2_{Sw}(t) = e^{-2 \cdot \lambda \cdot t}$, i — номер опорной точки, а N_ϕ и t — фактические значения доли совпадений и времени дивергенции[18].

[15] Ср. с аналогичным графиком на рис.3 (Васильев, Саенко 2016: 266).

[16] Напомним, что соответствие оказалось настолько точным, что после калибровки модели по исходным данным величина коэффициента λ не изменилась и совпала с исходным значением — 0,16.

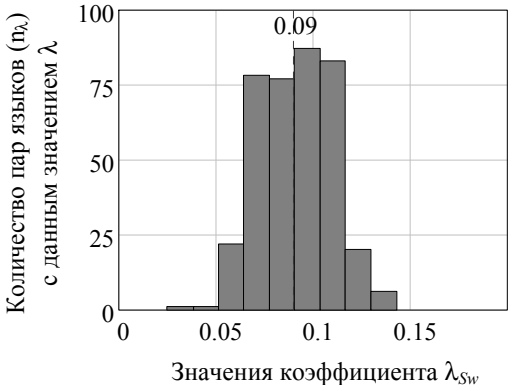
[17] См. подробное описание метода с примерами его использования в первой части исследования (там же: 265–267).

[18] Например, для случая с румынским и арумынским языками доля совпадений между их списками составляет 89% (N_ϕ=0,89), а время дивергенции 1100 лет (t=1,1). Подставляя значение времени в формулу

Подставляя в данную формулу данные из табл. 1 и выполнив необходимые вычисления[19], получаем коэффициент λ=0,09, удовлетворяющий условию наименьшего суммарного отклонения, величина которого составила ε=0,76 (см. рис. 3). Таким образом, калиброванная модель Сводеша будет иметь вид:

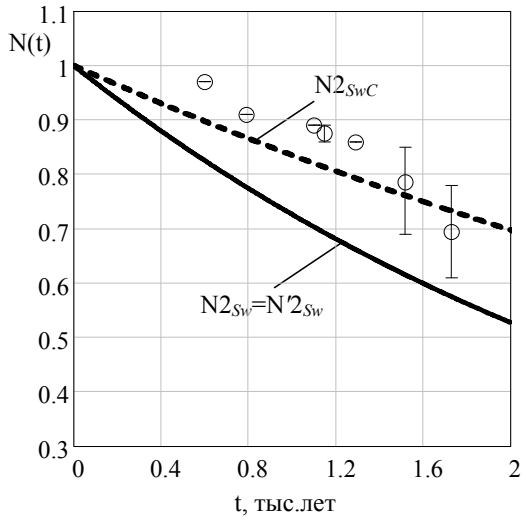
$$N2_{SwC}(t) = e^{-2\cdot 0,09\cdot t}.$$

Рисунок 3. Распределение значений коэффициента λ_{Sw}, рассчитанных по опорным точкам (табл. 1) с помощью модели Сводеша. Найденное оптимальное значение λ_{Sw} соответствует математическому ожиданию 0,09 при среднем квадратическом отклонении σ_λ=0,02.



Для сопоставления исходной и новой модели, а также оценки результатов калибрования обратимся к диаграмме на рис. 4.

Рисунок 4. Сравнение исходной и калиброванной моделей Сводеша с исходными данными:
$N2_{Sw}(t) = e^{-2\cdot 0,16\cdot t}$ (ε=8,46) — исходная модель Сводеша[20];
$N2_{SwC}(t) = e^{-2\cdot 0,09\cdot t}$ (ε=0,76) — калиброванная модель Сводеша.



Сводеша, мы получаем: $N_p=N2_{Sw}(t)=e^{-2\cdot\lambda\cdot 1,1}$. Очевидно, для наилучшего соответствия между фактическим и расчетным значением необходимо найти такое значение λ, при котором величина отклонения ε будет минимальной:

$$\varepsilon = (e^{-2\cdot\lambda\cdot 1,1} - 0,89)^2 \rightarrow \min.$$

[19] Большинство расчетов и построение графиков проводилось с помощью системы MathCad.

[20] Как уже говорилось выше (сноска 16), при калибровке коэффициента λ по исходным данным для изменения лексики одного языка его значение совпало с исходным (0,16). Таким образом, вид моделей $N2_{Sw}$ и $N'2_{Sw}$ (с исходным и калиброванным коэффициентами) также совпадает (см. в первой части работы — там же: 267).

Как следует из представленной диаграммы, переход к калиброванной модели с коэффициентом λ=0,09 позволил заметно уменьшить расхождение между расчетными и фактическими значениями[21]. В то же время для большинства опорных точек полученные датировки по-прежнему оказались моложе ожидаемых. При этом форма полученной кривой указывает, что дальнейшая калибровка модели не позволит добиться существенного улучшения результатов в силу принципиального несоответствия между простой экспоненциальной зависимостью и общим характером процесса дивергенции.

Попытка преодолеть этот сущностный недостаток классической глоттохронологии Сводеша была осуществлена в рамках усовершенствованной методики С. А. Старостина, к рассмотрению которой мы переходим.

2.2. Глоттохронологический метод С. А. Старостина

В работе (Starostin 2000: 233–259) С. А. Старостин устанавливает, что основной причиной неудач методики Сводеша является невыполнение двух основных постулатов глоттохронологии: о постоянной скорости лексических изменений и одинаковой стабильности значений в основном списке. Анализируя конкретные примеры развития базисной лексики, он предлагает ввести в исходную глоттохронологическую формулу $N_{Sw}(t) = e^{-\lambda \cdot t}$ две поправки: *замедляющую*, связанную с проявлением в списке наиболее устойчивой части лексики (λ=λ·N(t)), и *ускоряющую* — отражающую «устаревание» сохранившейся лексики, а следовательно — ускорение замен среди устаревших значений (λ=λ·t). Таким образом, процесс лексической дивергенции в одном языке должен описываться более сложным соотношением:

$$N_{St}(t) = e^{-\lambda \cdot N_{St} \cdot t^2}$$

При этом Старостин вслед за Сводешем принимает постулат о независимости развитии языков-потомков[22], что позволяет использовать такую же методику получения модели дивергенции — путем возведения исходной формулы во вторую степень:

$$N2_{St}(t) = N_{St}(t)^2 = e^{-2 \cdot \lambda \cdot N_{St} \cdot t^2} = e^{-2 \cdot \lambda \cdot \sqrt{N2_{St}} \cdot t^2}.$$

Апробируя полученную формулу на различном языковом материале (в том числе романском), Старостин определяет также константу λ, которая по разным подсчетам варьируется около величины 0,05. Таким образом, конечная модель для датирования относительного развития языков будет иметь вид:

$$N2_{St}(t) = e^{-2 \cdot 0{,}05 \cdot \sqrt{N2_{St}} \cdot t^2}.$$

Сопоставление графика полученной модели с опорными точками (рис. 5) подтверждает, что введение поправок позволило добиться значительно лучшего соответствия между фактическими и расчётными датировками, причем на всем рассматриваемом временном диапазоне. При этом параметры модели, найденные нами в ходе калибровки, существенно не отличаются от предложенных. Так, с помощью метода наименьших квадратов получаем коэффициент λ=0,07 (см. рис. 6), близкий к исходному значению 0,05[23].

---

[21] Об этом же свидетельствует изменение величины суммарного отклонения, которое уменьшилось с ε=8,46 для исходной модели до ε=0,76 для калиброванной формулы.

[22] См. п. 2.1 выше.

[23] Наблюдаемое при этом заметное уменьшение суммарного отклонения (которое снизилось с 1,97 до 0,72 после калибровки модели), объясняется в первую очередь неравным количеством сравниваемых идиомов в различных опорных точках. Так, опорные точки 7 и 8 содержат результаты сравнения для нескольких

Рисунок 5. Сравнение моделей Старостина с различными коэффициентами λ:
$N2_{St}(t) = e^{-2\cdot0,05\cdot\sqrt{N2_{St}}\cdot t^2}$ (ε=1,97) — исходная модель Старостина;
$N2_{StC}(t) = e^{-2\cdot0,07\sqrt{N2_{StC}}\cdot t^2}$ (ε=0,72) — калиброванная модель Старостина;
$N'2_{St}(t) = e^{-2\cdot0,11\cdot\sqrt{N'2_{St}}\cdot t^2}$ (ε=7,07) — модель Старостина с коэффициентом λ, калиброванным по данным для развития лексики одного языка.

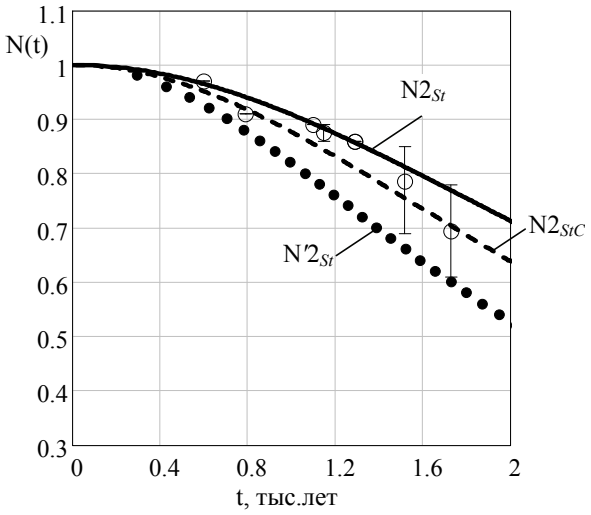
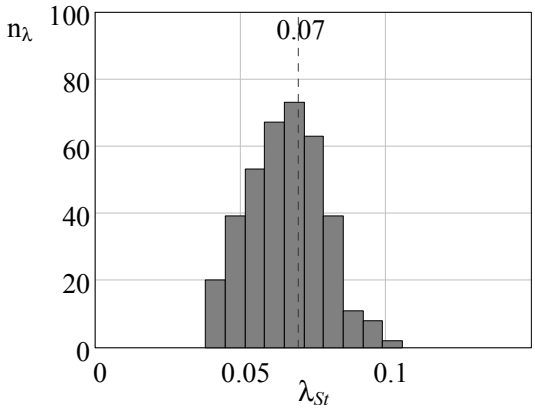


Рисунок 6. Распределение значений коэффициента λ, рассчитанных для опорных точек (табл. 1) по модели Старостина. Математическое ожидание коэффициента λ_{St} составляет 0,07; значение среднего квадратического отклонения σ_λ=0,013.



Примечательно, что при использовании найденного в первой части работы[24] коэффициента λ=0,11, полученного для модели Старостина по романским данным для дивергенции одного языка, результаты оказываются даже хуже, чем в случае с исходной константой (0,05) — см. рис. 5. Это несоответствие указывает на некорректность[25] применяемой как Сводешем, так и Старостиным методики перехода от модели независимого

десятков пар языков, в то время как предыдущее — всего для одной-двух пар. Таким образом, при вычислении суммарного отклонения, соответствие кривой последним двум точкам имеет гораздо больший «вес», чем всем остальным вместе взятым.

[24] См. Васильев, Саенко 2016: 268–269.

[25] Очевидно, в противном случае найденные коэффициенты λ должны были совпасть или иметь сходные значения.

развития одного идиома к модели относительной дивергенции и тем самым свидетельствует о невыполнении постулата Сводеша о независимом развитии языков-потомков. В свою очередь отказ от третьего постулата приводит к необходимости учитывать согласованные изменения в базисной лексике родственных языков после их разделения. Примером такого подхода к описанию процесса дивергенции является потоковая модель, которую мы рассмотрим далее.

2.3. Потоковая глоттохронологическая модель[26]

В отличие от представленных выше методик М. Сводеша и С. А. Старостина потоковая модель базируется на предположении, что развитие двух языков-потомков обладает определенной согласованностью, в результате чего даже спустя значительное время после их разделения в списках обоих идиомов могут заменяться одни и те же значения. При этом в каждом из списков можно выделить две составляющие, одна из которых соответствует значениям, которые развиваются сходным образом в обоих языках, а вторая — независимо развивающейся части списка. Причем в начальный момент разделения все значения будут развиваться согласованно (т. е. относиться к связанной составляющей), а в ходе дальнейшего развития — постепенно утрачивать эту согласованность и переходить в независимую составляющую, увеличение которой и будет соответствовать собственно дивергенции двух языков[27]. Если предположить, что убывание связанной составляющей происходит экспоненциально, а сам процесс замен внутри каждой из составляющих представляет собой сумму потоков[28] случайных событий, то формула, описывающая общий процесс дивергенции между двумя языками, примет вид[29]:

$$N2_P(t) = c_0 + c_1\left(\frac{\mu}{\mu-\eta}\cdot e^{-\eta t} + \frac{\eta}{\eta-\mu}\cdot e^{-\mu t}\right),$$

где константы c_0 и c_1 соответствуют количеству значений в наиболее устойчивой и изменяющейся частях списков, коэффициент η определяет скорость потерь в изменяющейся части списка, а величина μ является показателем увеличения его независимой составляющей.

---

[26] На всякий случай подчеркнем, что обсуждаемая ниже модель дивергенции *не идентична* потоковой модели, описывающей процесс замен в лексике одного языка и рассмотренной в первой части работы (Васильев, Саенко 2016: 269–270). Использование термина «потоковая» применительно к обеим моделям отражает тот факт, что в их основе лежит одно и то же представление о процессе замен в базисной лексике как о совокупности потоков редких случайных событий, каждое последующее из которых не зависит от предыдущего. Несмотря на теоретический характер, данное представление имеет ряд практических следствий, непосредственно влияющих на результаты и саму методику проводимого исследования. В частности, численная оценка точности и надежности глоттохронологических моделей, становится возможной благодаря известным статистическим свойствам стационарных потоков, используемых при моделировании.

[27] Более подробное описание и теоретическое обоснование модели содержится в статье Васильев, Милитарев 2008: 523–529.

[28] Каждый из таких потоков соответствует процессу замен одного из значений списка.

[29] Полный вывод формулы дается в Приложении к статье (Васильев, Милитарев 2008: 535–536). Особый интерес представляет содержательный анализ этой формулы и, в частности, вопрос соотношения значений η и μ, а также возможность перехода к упрощенному виду формулы при их равенстве ($\eta=\mu$). Обсуждение этих особенностей требует отдельного подробного рассмотрения, которое, к сожалению, выходит за рамки настоящей статьи.

Путем калибровки модели по исходным данным были получены следующие значения параметров[30]:

$$c_0=0{,}000;\ c_1=1{,}000;\ \eta=0{,}612^{31};\ \mu=0{,}611.$$

При подстановке данных значений в исходное выражение получаем модель:

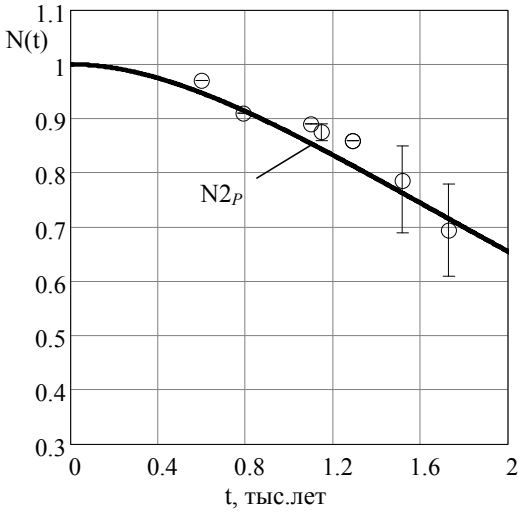
$$N2_P(t)=0{,}000+1{,}000\cdot\left(1297\cdot e^{-0{,}612t}-1296\cdot e^{-0{,}611t}\right),$$

которая в силу близости значений η и $\mu\approx0{,}61$ может быть представлена в упрощённой форме:

$$N2_P(t)=e^{-0{,}611t}(1+0{,}61t).$$

График полученной модели (рис. 7) наглядно демонстрирует хорошее совпадение расчетных значений с исходными данными на всем временном интервале[32].

Рисунок 7. Соответствие потоковой модели исходным данным: $N2_P(t)=e^{-0{,}611t}(1+0{,}61t)$; ε=0,54.

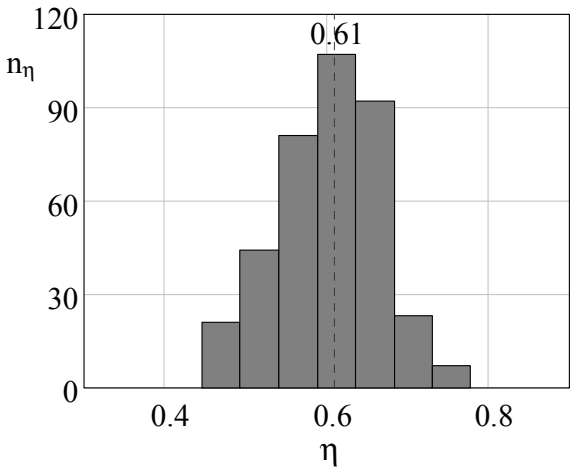


[30] Напомним, что калибровка всех моделей производилась типовым методом наименьших квадратов отклонений, см. также сноску 18.

[31] Распределение значений коэффициента η, полученных в ходе калибровки, представлено на рис. 8. Отметим, что найденное значение η (0,61) не совпадает с полученным ранее (0,45) на другом материале, включающем данные нескольких языковых семей (см. Васильев, Милитарев 2008: 529). Несоответствие полученных значений лишний раз подтверждает, что параметры глоттохронологический моделей в значительной мере определяются набором и качеством исходных данных, используемых при их калибровке. Поэтому при поиске параметров модели особенно важно привлечение максимального широкого языкового материала, как с точки зрения географического охвата, так и временной глубины.

[32] Видимая неравномерность распределения опорных точек относительно кривой связана с неравномерностью распределения исходных данных. Как уже отмечалось выше, наибольшим весом обладают две крайние опорные точки №№7 и 8 (см. табл. 1 и ссылку 23), поэтому в результате калибровки методом наименьших отклонений именно они главным образом определяют значения параметров модели. При этом легко убедиться, что указанные точки расположились симметрично по обе стороны расчётной кривой $N2_P$.

Рисунок 8. Распределение значений коэффициента η, рассчитанных по опорным точкам (табл. 1) с помощью потоковой модели. Полученное математическое ожидание величины η составляет 0,61; значение среднего квадратического отклонения σ_η=0,065.



Завершив подробное рассмотрение каждой из глоттохронологических моделей, перейдем теперь к сравнению и анализу полученных результатов.

2.4. Сравнение полученных моделей и их оценка

Сопоставление полученных моделей будем производить с помощью графиков, представленных на рис. 9 (а, б, в), а также их числовых параметров, приведенных в табл. 2.

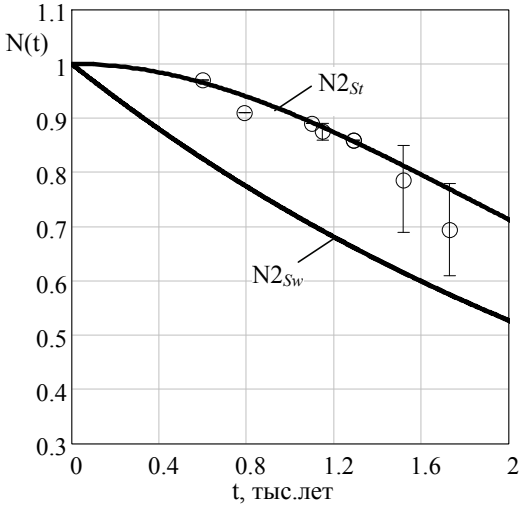
В первую очередь отметим, что калибровка формулы М. Сводеша ($N2_{Sw}$) как по данным общей, так и относительной дивергенции не приводит к существенному увеличению точности расчетных датировок, что, как уже говорилось выше, вызвано несоответствием экспоненциальной зависимости характеру процесса замен при расхождении двух языков-потомков. Так, при использовании модели с калиброванным коэффициентом λ_{SwC}=0,09 мы получаем правдоподобные даты разделения для интервала 1500–1700 лет и сильно заниженные (или наоборот — завышенные) значения за его пределами.

Гораздо лучшие результаты демонстрирует модель С. А. Старостина ($N2_{St}$), которая, после отказа от двух постулатов Сводеша и внесения соответствующих поправок в классическое уравнение глоттохронологии, позволила добиться хорошего совпадения с опорными точками на всем рассматриваемом временном отрезке. Причём оптимальное значение коэффициента λ_{StC} (0,07), найденное в ходе подбора параметров модели по фактическим данным, оказалось очень близко к исходному (0,05) — см. рис. 9а,в и табл. 2. При этом данное значение λ_{StC} заметно отличается от полученного при калибровке модели Старостина по тем же романским данным, но для одного языка (λ'_{StC}=0,11) (рис. 9б). Обнаруженное несоответствие свидетельствует о том, что процесс дивергенции между родственными языками не может быть смоделирован на основе двух независимых процессов развития каждого из них[33] и, следовательно, указывает на несостоятельность постулата Сводеша о независимом развитии языков-потомков после их разделения. Отказ от принципа независимости создает предпосылку для перехода к модели, которая могла бы учитывать согласованность процесса лексических замен в разделившихся идиомах.

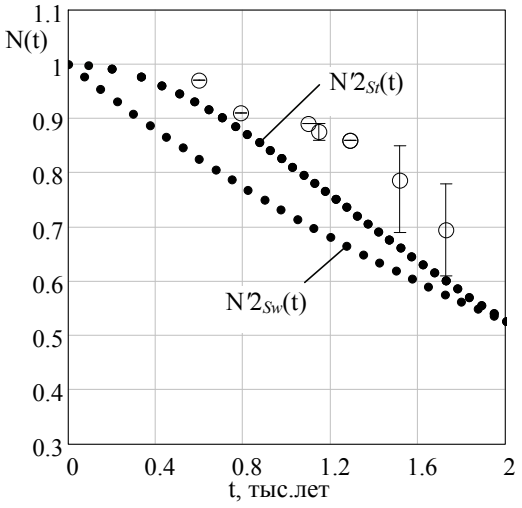
[33] Что подразумевается методикой Сводеша и Старостина при получении модели дивергенции из общей формулы глоттохронологии.

Данный подход был реализован при построении потоковой модели ($N2_P$), эффективность использования которой при датировании процесса дивергенции подтверждается как графически (рис. 9в), так и численно — наименьшей (по сравнению с остальными моделями[34]) величиной суммарного отклонения ε_P=0,54 (см. табл. 2).

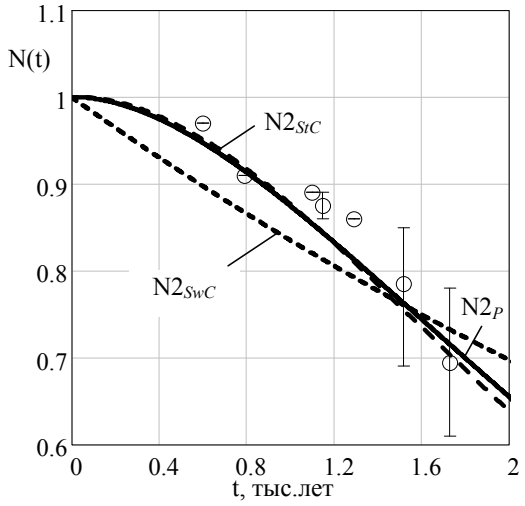
Рисунок 9. Сравнение исходных и калиброванных моделей дивергенции М. Сводеша, С. А. Старостина и потоковой.



а) Исходные модели М. Сводеша ($N2_{Sw}$) и С. А. Старостина ($N2_{St}$)



б) Модели Сводеша ($N'2_{Sw}$) и Старостина ($N'2_{St}$) с коэффициентом λ, калиброванным по данным процесса замен в одном языке



в) Калиброванные по фактическим данным модели Сводеша ($N2_{SwC}$), Старостина ($N2_{StC}$) и потоковая модель ($N2_P$).

[34] Тем не менее, калиброванная модель Старостина, несмотря на выявленные методические недостатки, численно даёт значения, почти идентичные потоковой модели на большей части временного интервала (ср. кривые $N2_{SwC}$ и $N2_P$ на рис.9в), что позволяет использовать её для датирования дивергенции (по крайней мере — в рамках указанного интервала времени).

Таблица 2. Сравнение параметров исходных и калиброванных моделей

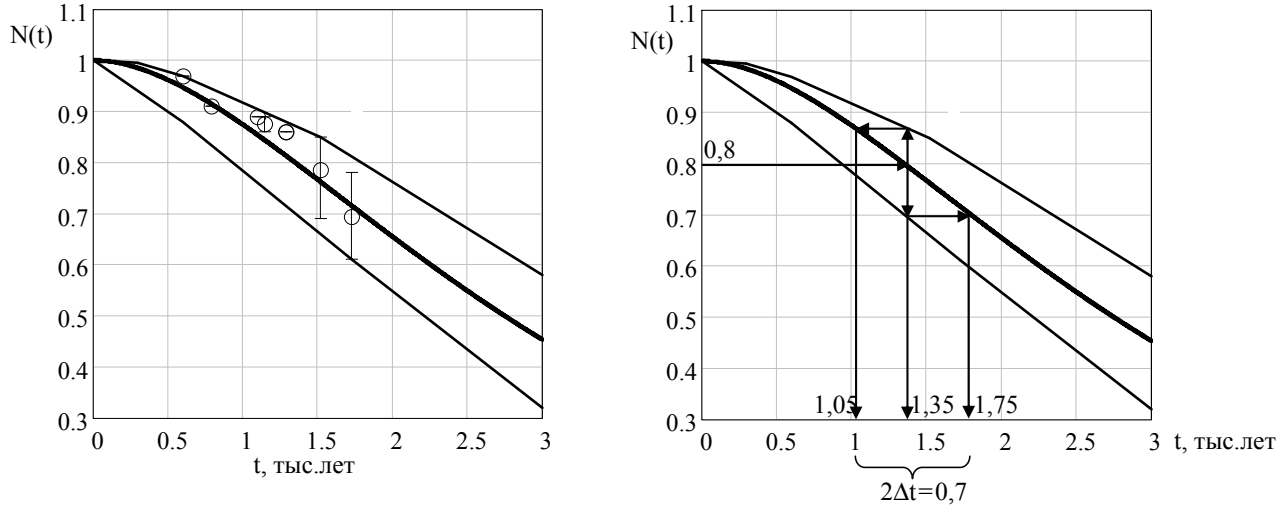
Название и общий вид модели		Исходные параметры модели	Параметры, калиброванные по данным процесса замен в одном языке	Параметры, калиброванные по фактическим данным процесса дивергенции
Модель М. Сводеша	$N2_{Sw}(t) = e^{-2\cdot\lambda\cdot t}$	λ_{Sw}= 0,16 (ε_{Sw}=8,46)	λ'_{Sw}= 0,16 (ε_{Sw}=8,46)	λ_{SwC}=0,09 (ε_{Sw}= 0,76)
Модель С. А. Старостина	$N2_{St}(t) = e^{-2\cdot\lambda\cdot\sqrt{N2_{St}}\cdot t^2}$	λ_{StC}= 0,05 (ε_{St}=1,97)	λ'_{StC}= 0,11 (ε_{St}=7,07)	λ_{StC}=0,07 (ε_{St}=0,72)
Потоковая модель	$N2_{P}(t) = e^{-\eta t}(1+\eta t)$	—		η=0,61 (ε_{P}=0,54)

Завершив сравнение существующих моделей и установив их основные особенности, мы можем перейти к вопросу о теоретической и практической погрешности глоттохронологических датировок, а также их статистической достоверности.

3. Погрешности и достоверность глоттохронологических датировок

Определение погрешностей, возникающих при глоттохронологических расчетах, начнём с оценки фактического разброса значений в исходных данных, которые очевидным образом и будут определять минимальную погрешность наших расчетов (Васильев, 2010: 538; Васильев, Коган: 2013: 156–159). Для этого воспользуемся данными из табл. 1, а также диаграммой (рис. 10а), на которой верхняя и нижняя кривые соединяют соответственно максимальные и минимальные значения процентов совпадений, известные для каждой опорной точки, а средняя линия отражает расчетные значения потоковой модели ($N2_P$). Например, в соответствии с табл. 1, процент совпадений между списками языков, разделившихся 1520 лет назад (точка 7), варьируется в пределах от 69 до 85% (ΔN=16%). Еще больший разброс долей совпадений (ΔN=17%) мы наблюдаем для даты разделения 1730 лет назад (точка 8) — от 61 до 78%.

Рисунок 10



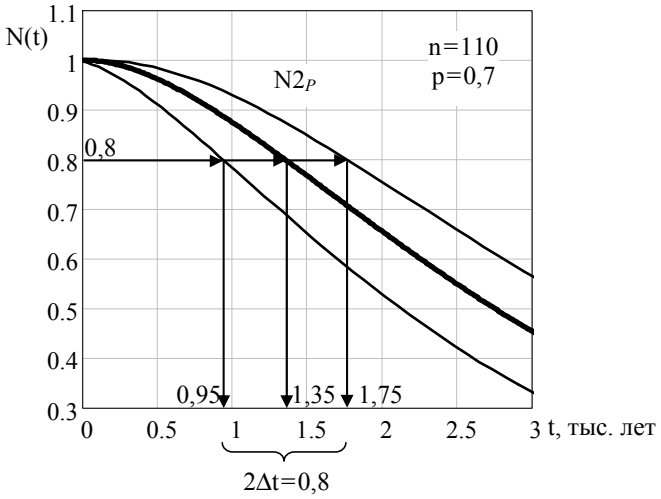
а) иллюстрация разброса фактических долей совпадений по отношению к расчётным значениям N(t), полученным по модели $N2_P$

б) иллюстрация разброса фактических датировок по отношению к расчетным значениям t, вычисленным по модели $N2_P$

Аналогично с помощью диаграммы на рис. 10б можно оценить разброс датировок, вычисленных с помощью потоковой модели для выбранного значения процента совпадений (N). Так, подставляя в формулу $N2_P(t)$ значение N=0,8 (80%), получаем расчетную датировку t=1350 лет назад. При этом, как видно на рисунке, фактическое разделение идиомов могло произойти в диапазоне времени от 1050 до 1750 года — т. е. с разницей в 700 лет. Это означает, что на практике дата разделения рассматриваемых идиомов не может быть определена точнее, чем в диапазоне 1350±350 лет. Таким образом, фактический разброс исходных данных (вне зависимости от используемых моделей) вносит неизбежную и существенную погрешность в результаты любых глоттохронологических расчетов[35].

Величину установленной *фактической* погрешности, связанной со статистическим характером процесса замен, полезно сопоставить с погрешностью *теоретической*, обусловленной особенностями самих используемых моделей. Мерой этой погрешности, применительно к потоковой модели, является величина *доверительного интервала*, описанная нами ранее в первой части исследования (Васильев, Саенко 2016: 274–275)[36]. В частности, доверительный интервал, вычисленный для некоторого известного процента совпадений, позволяет определить временной диапазон, в который с заданной вероятностью укладывается расчетная датировка.

Рисунок 11. Доверительный интервал модели $N2_P(t) = e^{-0,61t}(1+0,61t)$, рассчитанный для 110-словного списка с заданной вероятностью p=0,7.



Например, для доли совпадений 80% и соответствующей ему расчетной датировки 1350 лет назад, теоретическая величина доверительного интервала составляет 800 лет (рис. 11) — т. е. искомая датировка с вероятностью 0,7[37] будет располагаться в диапазоне 1350±400 лет назад.

Как видно на рис. 11 и 12, с увеличением временной дистанции доверительный интервал также увеличивается, однако в процентном отношении его значение убывает по

[35] Причиной такого разброса, как уже говорилось выше, является, с одной стороны случайная природа рассматриваемого процесса дивергенции (см. Васильев, Г. Старостин, 2014: 60), а с другой — невозможность абсолютно достоверного датирования опорных точек по известным историческим событиям.

[36] Методика расчета доверительных интервалов для потоковой модели дивергенции основана на вычислении плотности распределения вероятностей первых замен в списках каждого из языков-потомков (Вентцель, Овчаров 1969: 235–237).

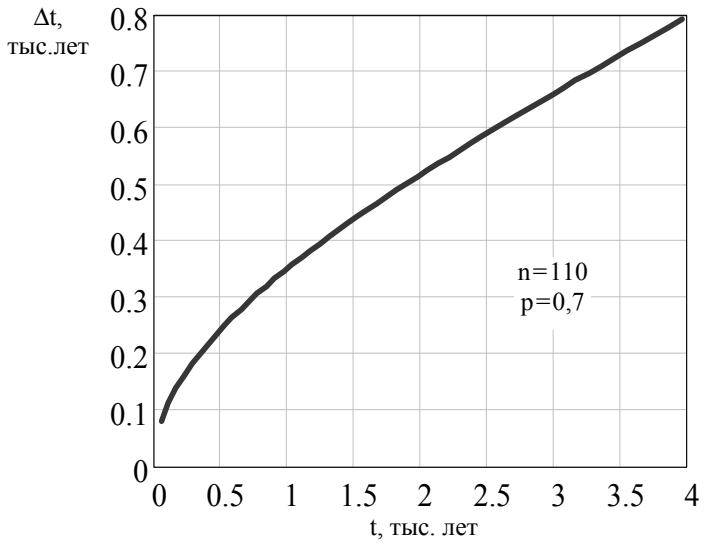
[37] Т. е. в 70 случаях из 100.

мере углубления датировок (табл. 3). Например, в соответствии с таблицей 3, при расчетной дате разделения 1000 лет назад доверительный интервал составляет ±350 лет (т. е. начало дивергенции с вероятностью 0,7 может датироваться от 650 до 1350 лет назад). Аналогично для времени разделения 3500 лет назад получаем доверительный интервал ±730 лет. Таким образом, абсолютная величина доверительного интервала выросла более чем в два раза (от ±350 до ±730 лет), в то время как его относительное значение снизилось с 35% до 21% (табл. 3). Это означает, что, несмотря на уменьшение абсолютной точности, практическая ценность глоттохронологических датировок будет заметно выше при бóльших временных интервалах.

Таблица 3. Значения доверительного интервала, рассчитанные для потоковой модели $N2_P$ с заданной вероятностью 0,7[38]

t, *лет*	200	400	600	800	1000	1200	1400	1600	1800	2000	2500	3000	3500	4000
Δt, *лет*	±150	±220	±270	±310	±350	±390	±420	±450	±480	±510	±590	±660	±730	±790
100Δt/t	75%	55%	45%	39%	35%	32%	30%	28%	27%	26%	24%	22%	21%	20%

Рисунок 12. Изменение величины 70-процентного доверительного интервала (Δt) в зависимости от времени (t) для 110-словного списка

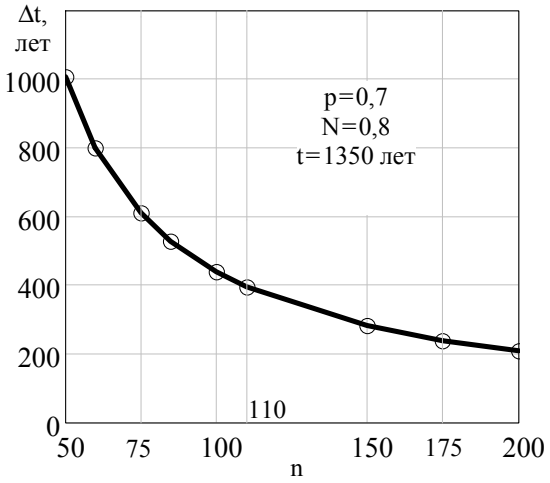


Пользуясь установленными свойствами потоковой модели, мы можем также определить зависимость величины доверительного интервала от количества значений в списках сравниваемых языков. Результаты проведенных расчетов представлены в виде графика на рис. 13.

Так, при использовании двухсотсловных списков для датирования дивергенции двух идиомов с долей совпадений 0,8 мы получаем дату разделения 1350 лет назад с доверительным интервалом ±200 лет (погрешность 15%). При уменьшении размера списка до 110 слов доверительный интервал увеличивается до ±400 лет (30%), а в случае с пятидесятисловным списком достигает значения ±1000 лет (74%!).

[38] Величина доверительного интервала очевидным образом зависит также от выбранной вероятности. Например, при использовании вероятности 0,95, расчетные значения доверительного интервала увеличатся примерно в два раза.

Рисунок 13. Изменение ширины доверительного интервала (Δt) в зависимости от числа лексических значений (n) в списках сравниваемых языков (для времени дивергенции t=1350 лет)



Полученная зависимость Δt(n) указывает на бесперспективность использования для глоттохронологического анализа коротких списков, что, однако, не умаляет полезности этих списков при установлении генеалогических связей между языками.

Сопоставляя между собой рис. 10 и 11, а также полученные нами расчетные значения, несложно убедиться в том, что величина доверительного интервала (при выбранной вероятности 0,7) лишь незначительно превышает фактический разброс исходных данных на рассматриваемом интервале времени. Следовательно, мы можем предположить, что решающее значение при оценке общей точности глоттохронологических датировок будет иметь именно эта объективная погрешность. Проверим справедливость нашего предположения на конкретных примерах дивергенции между языками романской группы, а также некоторыми другими языками.

В первой части табл. 4 приведены данные для нескольких пар идиомов с предположительной датой разделения 480 г. н.э. При этом средний процент совпадений между их списками варьируется от 75% (между португальским и галло-романскими) до 87% (между фриульским и лигурийскими). Расчетные датировки, полученные для этих значений с помощью потоковой модели составляют 410 и 960 г. соответственно. Таким образом, диапазон разброса фактических значений для всей группы из 14-ти романских языков составил 550 лет (или ±275 лет), что с запасом «укладывается» в теоретический доверительный интервал ±415 лет, вычисленный для среднего значения совпадений N=78,5 (см. табл. 1) и вероятности p=0,7. При рассмотрении отдельных пар языков (см. выделенные строки табл. 4) в 3-х случаях из 14-ти, (т. е. в 22% случаев) отклонение фактических дат распада от расчётных значений выходит за рамки 70-процентного доверительного интервала, что также согласуется с теоретической оценкой его статистической значимости. Так, большинство пар с участием фриульского дают сильно завышенные проценты совпадений, что приводит к «омоложению» расчетных датировок почти в два раза по сравнению с предполагаемой датой разделения[39]. Например, для фриульского и лигурийского с долей совпадения 86,7% получаем дату 960 г. с доверительным интервалом ±310 лет, в который очевидным образом не укладывается фактическое значение 480 г. В то же время для большинства остальных пар величина доверительного интервала оказывается избыточной, а отклонение расчетных датировок от фактической

[39] Заметим, впрочем, что подобный «подскок» значений может также объясняться более поздним отделением фриульского от сравниваемых с ним идиомов.

Таблица 4. Даты дивергенции языков, а также их доверительные интервалы, рассчитанные на основе потоковой модели $N2_P$.

Сравниваемые языки	Средний % совпадений	Фактическая датировка (лет)	Расчётная датировка (лет)	Доверительный интервал, p=0,7 (лет)
Руманшские — лигурийские	78,8	480	590	±410
Руманшские — сицилийские	80,7	480	680	±380
Руманшские — португальский/галисийский	78,3	480	570	±420
Руманшские — галло-романские	77,1	480	520	±430
Лигурийские — португальский/галисийский	80,3	480	660	±400
Лигурийские — галло-романские	77,0	480	510	±430
Сицилийские — португальский/галисийский	81,5	480	720	±370
Сицилийские — галло-романские	80,2	480	660	±400
Португальский/галисийский — галло-романские	74,7	480	410	±460
Фриульский — руманшские	84,3	480	850	±350
Фриульский — лигурийские	86,7	480	960	±310
Фриульский — сицилийские	86,8	480	960	±310
Фриульский — португальский/галисийский	83,0	480	790	±360
Фриульский — галло-романские	82,0	480	740	±370
Южнославянские — восточнославянские[40]	77,0	480	510	±430
Путунхуа — миньские идиомы[41]	63,5	–110	–90	±580
Балкано-романские — основной массив романских	69,6	271	180	±520

не превышает 200 лет, (см., например, рис. 14), что подтверждает адекватность используемой модели и её параметров рассматриваемому процессу дивергенции.

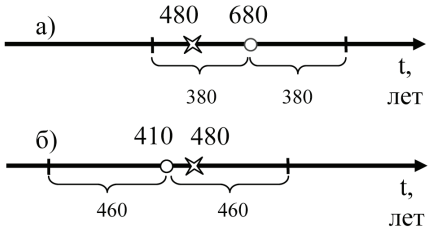
Безусловно, более показательной (и методически корректной) была бы апробация модели на другом языковом материале (который ранее не использовался при её калибровке) и на других интервалах времени. В качестве примера можно привести результаты датирования дивергенции китайских, славянских, а также балкано-романских языков (последние три строки табл. 4). Во всех трех случаях полученные датировки оказались очень близки к предполагаемой фактической дате разделения. Подобные примеры сви-

[40] В сравнении участвовали списки орбаничского чакавского, градищанского кайкавского и люблянского словенского, с одной стороны, с туровским белорусским и деулинским русским — с другой.

[41] Использовались списки путунхуа, цзяньоу и хайнаньского, составленные Г. С. Старостиным и Е. А. Кузьминой. Проценты совпадений для обоих примеров приводятся по данным из «Глобальной лексикостатистической базы данных», представленным на сайте проекта http://starling.rinet.ru/new100/main.htm (по состоянию на 01.05.2017) .

детельствуют о том, что эффективное использование полученной модели в теории не ограничено только романскими языками или определенным временным отрезком.

Рисунок 14. Иллюстрация взаимного расположения некоторых фактических и расчётных дат распада относительно доверительных интервалов: а) руманшские — сицилийские; б) португальский/галисийский — галло-романские.



Таким образом, точность глоттохронологических расчетов определяется в первую очередь не свойствами моделей, а случайным характером процесса лексических замен, который проявляется в существенном разбросе фактических долей совпадений, полученных для пар языков с одинаковыми интервалами распада. Величина этого разброса вносит основной вклад в конечную погрешность получаемых датировок.

Заключение

Подводя итоги, сформулируем основные результаты проведенного исследования в виде нескольких обобщающих выводов и положений:

1. Сравнительный анализ существующих глоттохронологических методов показывает, что наилучшие результаты при датировании процесса дивергенции достигаются при использовании модели С. А. Старостина и потоковой модели (после их предварительной калибровки). При этом попытки построения моделей дивергенции на основе данных общего распада, как это подразумевается методикой М. Сводеша и С. А. Старостина, приводит к абсурдным результатам и указывает на несостоятельность используемого в них постулата Сводеша о независимом развитии языков-потомков после разделения. Таким образом, моделирование процессов диверегенции должно учитывать возможность согласованного изменения в лексике родственных языков, при котором в списках разделившихся идиомов происходят замены одних и тех же значений.
2. Калибровка рассмотренных моделей по исходным данным позволила добиться хорошего численного совпадения расчетных и фактических датировок. При этом отдельные примеры показывают, что калиброванные модели могут эффективно применяться для датирования языковой дивергенции в других языковых семьях и на различных временных глубинах.
3. Точность глоттохронологических расчетов определяется в первую очередь не свойствами моделей, а вероятностным характером процесса лексических замен, который выражается в существенном разбросе фактических значений, величина которого и вносит основной вклад в конечную погрешность получаемых датировок. В силу случайного характера лексических замен определение времени разделения языков возможно только в пределах некоторого доверительного интервала с заранее выбранной вероятностью попадания фактической даты в этот интервал. Таким образом, корректная датировка дивергенции двух идиомов должна представлять собой не конкретное значение, а интервал значений с соответствующей величиной вероятности. Например, вместо «1000 лет назад», следует указывать «1000±350 лет назад с вероятностью 70%».

4. Теоретическая оценка доверительных интервалов, полученная на основе моделирования процесса дивергенции в виде потока лексических замен, позволила установить, что по мере увеличения времени дивергенции относительное значение этого интервала уменьшается и стремится к некоторому постоянному значению. Например, для заданной вероятности p=0,7 и периоде дивергенции 500 лет доверительный интервал составляет ±50% от этого периода, при 2000 лет — 26%, а к 4000 лет приближается к 20-процентному уровню.
5. Сравнение теоретических погрешностей моделей с фактическим разбросом известных данных, полученных для романских языков, свидетельствует о том, что на временном интервале до 2 тыс. лет погрешности датировок, вызванные случайным характером замен, являются доминирующими и носят объективный характер — т. е. не могут быть существенно снижены (в статистическом смысле) за счет дальнейшего уточнения стословных списков или привлечения дополнительных данных.
6. Установленная зависимость ширины доверительного интервала от числа лексических значений в списках сравниваемых языков показывает, что при расширении списка значение доверительного интервала пропорционально уменьшается. Например, при периоде дивергенции 1350 лет величина доверительного интервала для 200-словного списка в два раза меньше, чем для 110-словного. Таким образом, увеличение размера списков в теории позволяет существенно повысить точность глоттохронологических расчётов.
7. Дальнейшее повышение теоретической точности и надёжности глоттохронологических моделей возможно в первую очередь за счет привлечения дополнительных данных (опорных точек) для калибровки моделей на материале различных языковых семей на разных временных глубинах.

Литература

Алексеев, А. Н. 2013. Ранние кочевники в Якутии. Вестник Северо-Восточного федерального университета им. М. К. Аммосова 5(10): 62—69.

Арапов, М. В., М. М. Херц. 1974. Математические методы в исторической лингвистике. Москва: Наука.

Багаев, М. Х. 2015. К вопросу об этнокультурной общности на северо-восточном Кавказе в VIII–IV тыс. до н.э. В.: Н. Ярычев (ред.). 4-я ежегодная итоговая конференция профессорско-преподавательского состава Чеченского государственного университета. 28 февраля 2015 г.: 126–128. Грозный: Чеченский государственный университет.

Васильев, М. Е. 2010. Об использовании лексического критерия для построения генеалогической классификации. В: З. М. Шаляпина (ред.). Востоковедные чтения 2008. Бюллетень обшества востоковедов РАН. Тɒvды межинститутской научной конференции. Выпуск 17: 530–572. Москва: Институт востоковедения РАН.

Васильев, М. Е., А. И. Коган. 2013. К вопросу о восточнодардской языковой общности. Вестник РГГУ. Серия: Филология. Вопросы языкового родства. № 16 (117): 149—177.

Васильев, М. Е., А. Ю. Милитарев. 2008. Глоттохронология в сравнительно-историческом языкознании. Модели дивергенции языков. Orientalia et Classica: Труды Института восточных культур и античности 19: 509—536.

Васильев, М. Е., М. Н. Саенко. К вопросу о точности глоттохронологии: датирование процесса лексических замен по данным романских языков. Вестник РГГУ. Серия: Филология. Вопросы языкового родства. 14(4): 259–278.

Васильев, М. Е., Г. С. Старостин. 2014. Лексикостатистическая классификация нубийских языков: к вопросу о нильско-нубийской языковой общности. Вестник РГГУ. Серия: Филология. Вопросы языкового родства. № 16 (138): 51—72.

Вентцель, Е. С., Л. А. Овчаров. 1969. Теория вероятностей. Москва: Наука.

Иллич-Свитыч, В. М. 1966. Мнимые и действительные возможности лексикостатистики. В: Основные проблемы эволюции языка: 160–162. Самарканд: Фан.

Клейн, Л. С. 2010. Время кентавров. Степная прародина греков и ариев. С.–Петербург: Евразия.
Нарумов, Б. П. 2001. Арумынский язык/диалект. В: И. Челышева (ред.). Языки мира. Романские языки: 636–656. Москва: Academia.
Сводеш, М. 1960. Лексикостатистическое датирование доисторических этнических контактов. Новое в лингвистике 1: 23–52.

References

Alekseev, A. N. 2013. Ranniye kochevniki v Yakutii. Vestnik Severo-Vostochnogo federal'nogo universiteta im. M. K. Ammosova 5(10): 62—69.
Arapov, M. V., M. M. Herz. 1974. Matematicheskiye metody v istoricheskoy lingvistike. Moskva: Nauka.
Bagaev, M. Kh. 2015. K voprosu ob etnokul'turnoy obshchnosti na severo-vostochnom Kavkaze v VIII–IV tys. do n.e. In: N. Yarychev (ed.). 4 ezhegodnaya itogovaya konferenciya professorsko-prepodavatel'skogo sostava Chechenskogo gosudarstvennogo universiteta 28 fevral'a 2015 goda: 126–128. Grozniy: Chechenskiy gosudarstvennyj universitet.
Illich-Svitych, V. M. 1966. Mnimye i deystvitel'nye vozmozhnosti leksikostatistiki. In: Osnovnye problemy evolyutsii yazyka: 160–162. Samarkand: Fan.
Klein, L. S. 2010. Vremya kentavrov. Stepnaya prarodina grekov i ariev. S.-Petersburg: Evraziya.
Narumov, B. P. 2001. Arumynskiy yazyk/dialekt. In: I. Chelysheva (ed.). Yazyki mira. Romanskie yazyki: 636–656. Moskva: Academia.
Starostin, S. 2000. Comparative-historical linguistics and lexicostatistics. In: Colin Renfrew et al. (eds.). Time Depth in Historical Linguistics: 233–259. Cambridge: McDonald Institute for Archaeological Research.
Swadesh, M. 1960. Leksikostatisticheskoye datirovaniye doistoricheskikh etnicheskikh kontaktov. Novoye v lingvistike 1: 23—52.
Vasilyev, M. E. 2010. Ob ispol'zovanii leksicheskogo kriteriya dlya postroyeniya genealogicheskoy klassifikatsii. In: Z. Shalyapina (ed.). Vostokovednyye chteniya 2008. Bvulleten' Obshchestva vostokovedov RAN. Trudy mezhinstitutskoy nauchnoy konferentsii 17: 530-572. Moskva: Institut vostokovedeniya RAN.
Vasilyev, M. E., A. Yu. Militaryov. 2008. Glottokhronologiya v sravnitel'no-istoricheskom yazykoznanii. Modeli divergentsii yazykov. Orientalia et Classica: Trudy Instituta vostochnykh kultur i antichnosti 19: 509—536.
Vasilyev, M. E., A. I. Kogan. 2013. K voprosu o vostochnodardskoy yazykovoy obshchnosti. Journal of Language Relationship 10: 149–177.
Vasilyev, M. E., G. S. Starostin. 2014. Leksikostatisticheskaya klassifikatsiya nubiyskikh yazykov: k voprosu o nil'sko-nubiyskoy yazykovoy obshchnosti. Journal of Language Relationship 12: 51–72.
Vasilyev, M. E., M. N. Saenko. K voprosu o tochnosti glottokhronologii: datirovaniye protsessa leksicheskikh zamen po dannym romanskikh yazykov. Journal of Language Relationship 14(4): 259–278.
Venttsel, E. S., L. A. Ovcharov. 1969. Teoriya veroyatnostey. Moskva: Nauka.

Mikhail Vasilyev, Mikhail Saenko. How accurate can glottochronology be? Dating language divergence on the basis of Romance data.

The paper is a sequel to an earlier study by the authors, in which they discussed the accuracy of linguistic datings arrived at by the glottochronological method on the basis of data from 110-item wordlists for Romance languages. The object of this second part of the study is the dating of linguistic divergence, i.e. determining the separation dates for two or more modern languages. In this paper, we compare several traditional as well as newly offered models for the glottochronological process, with special attention paid to the margin of error and reliability of glottochronological calculations on different time depths. The results of the study allow for a realistic assessment of the degree of accuracy in the glottochronological dating of the divergence of Romance languages and lead to a number of practical conclusions that will be useful for the application of glottochronology to any other linguistic material.

Keywords: glottochronology, lexicostatistics, Swadesh wordlist, Romance languages.