

**Глоттохронология
в сравнительно-историческом языкознании.
Модели дивергенции языков**

**[Glottochronology in Comparative-Historical Linguistics
and the Models of Linguistic Divergence]**

М. Е. ВАСИЛЬЕВ, А. Ю. МИЛИТАРЁВ*

Москва, Российский Государственный Гуманитарный Университет

The paper addresses several issues with Morris SWADESH's method of dating linguistic divergence based on lexicostatistical calculations. After a brief explanation of the basic mathematical principles that constitute the foundation of the method, attention is drawn towards additional factors that hamper a simplistic application of Swadesh's formula. Some of them have already been dealt with in earlier works by S. A. STAROSTIN (such as the so-called «aging of words» and the difference in rates of stability among different items on the basic wordlist), but a somewhat different approach to their formulization is advocated.

Another factor that has not been previously considered on a serious mathematical basis is the possibility of interdependent development of languages that may happen for prolonged periods of time under specific conditions. This factor is also taken into account and given a mathematical expression.

* М. Е. ВАСИЛЬЕВ — студент кафедры иудаики ИСАА МГУ. А. Ю. МИЛИТАРЁВ — профессор этой кафедры (а также Института лингвистики и Института восточных культур и античности РГГУ), руководитель курсовых работ М. Е. ВАСИЛЬЕВА. Его участие в данном исследовании состоит в инициировании интереса М. Е. ВАСИЛЬЕВА к проблематике глоттохронологии, обсуждении теоретических и прикладных её аспектов в качестве лингвиста-компаративиста с двадцатипятилетним опытом применения глоттохронологического метода — от свадешевского до старостинского — для генетической классификации афразийских языков, а также в редактировании статьи и «переводе» лапидарной математической манеры изложения соавтора на язык, более понятный гуманитариию. Все новые концептуальные подходы и математический аппарат исследования разработаны М. Е. ВАСИЛЬЕВЫМ и его отцом — воронежским математиком Е. М. ВАСИЛЬЕВЫМ, из скромности не пожелавшим числиться в соавторах работы; он же взял на себя объёмный и кропотливый труд по машинной обработке и первичному анализу данных, а также отредактировал данную статью. Авторы приносят ему глубокую благодарность.

М. Е. ВАСИЛЬЕВ также благодарит А. В. ДЫБО, Г. С. СТАРОСТИНА, О. А. МУДРАКА и С. А. БУРЛАКА, ответивших на многочисленные вопросы и давших ряд ценных советов, использованных в статье.

А. Ю. МИЛИТАРЁВ выражает благодарность фондам, в течение ряда лет поддерживающих его исследования — Российскому фонду фундаментальных исследований (проект «От древних семитов к Библии и Корану: культурная трансформация с 5 тыс. до н. э. по 1 тыс. н. э. в свете лингвистических данных»), Российскому гуманитарному научному фонду (проект «Семитский этимологический словарь»), Российскому еврейскому конгрессу (проект «Вавилонская башня») и Институту Санга Фе (проект «Эволюция человеческих языков»).

Введение

Сравнительно-историческое языкознание (лингвистическая компаративистика) ставит перед собой следующие задачи:

- установление родственных связей между языками;
- построение генеалогической классификации языковых семей;
- реконструкция праязыковых состояний.

Из самой постановки этих задач для авторов данной работы вытекает идеологическая концепция их лингвистических исследований: языковая картина мира является результатом лингвогенеза, сочетающего в себе эволюционные процессы изменчивости и преемственности, на любом этапе которых эта картина рассматривается как временной срез ветвей единого генеалогического дерева языков мира.

Фундаментальный вклад в формирование указанной концепции, на которой построена московская школа макрокомпаративистики, сделал В. М. Иллич-Свитыч, по-новому разработавший проблему отдалённого родства ряда языковых семей Старого Света: индоевропейской, семитохамитской, картвельской, уральской, алтайской и дравидийской. Результатом его исследований стало обоснование ностратической теории, в которой В. М. Иллич-Свитыч смог объединить все вышеуказанные семьи языков по установленным им соответствиям в единую ностратическую макросемью. Эти результаты были опубликованы уже после смерти В. М. Иллич-Свитыча его коллегой В. А. Дыбо при участии ещё ряда лингвистов; начиная с 1971 г. тремя выпусками был издан труд В. М. Иллич-Свитыча «Опыт сравнения ностратических языков». Спустя четверть века (1995 г.) итогам развития ностратической теории был посвящён первый том Московского лингвистического журнала. Тем не менее, несмотря на распространение эволюционной концепции лингвогенеза, в настоящее время мы продолжаем сталкиваться с альтернативным подходом к возникновению языковой системы мира, опирающимся на плоскую, мозаичную картину многообразия языков, изменяемых в соответствии с гипотезой «смещения красок». Таким образом, задача утверждения эволюционных принципов в компаративистике по-прежнему остаётся актуальной.

Второй принципиальной основой компаративистики является её методология — сравнительно-исторический подход, в рамках которого нас будет интересовать лексикостатистический метод установления родства языков, разработанный в 1950-х годах М. Сводешом.

Фундаментальные принципы глоттохронологии М. Сводеша принято излагать в следующем порядке ([4: 21–25]):

1. В словарном составе любого языка мира можно выделить некоторое подмножество слов, отличающихся особой стабильностью. Эти слова называют основной или устойчивой частью языка.

2. Можно указать список некоторых базовых общечеловеческих понятий, которые в любом языке мира будут выражены словами из его основной части. Эти слова образуют основной список (ОС).

3. Доля слов из ОС, которые остаются неизменными, т. е. не замещаются другими этимологически тождественными словами за определённый интервал времени Δt , является постоянной величиной и зависит только от длины этого интервала, а не от того, когда этот интервал был выбран или слова какого языка рассматриваются.

4. Все слова из ОС могут с равной вероятностью остаться или исчезнуть из языка за определённый промежуток времени.

5. Вероятность того, что слово из ОС праязыка сохранится в ОС одного языка-потомка, не зависит от вероятности его сохранения в аналогичном списке другого языка-потомка.

Содержанию постулатов М. Сводеша соответствует основное математическое уравнение глоттохронологии, описывающее процесс замены слов в ОС языка:

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

где t — промежуток времени между точками сравнения, измеряемый в тысячелетиях;

N_0 — начальное количество слов в ОС: $N_0 = N(0)$, которое обычно полагают равным единице;

λ — интенсивность распада, в глоттохронологии часто называемая коэффициентом потерь слов из ОС (по М. Сводешу $\lambda = 0,14$);

$N(t)$ — количество слов от начального списка N_0 , сохранившихся в языке за промежуток времени t .

Это уравнение традиционно использовалось в физической теории радиоактивного распада и в гуманитарных кругах стало широко известно благодаря открытию в 1949 г. метода радиоуглеродного анализа («Carbon-14»), применяемого для датирования археологических находок. Однако, в отличие от триумфальных успехов радиоуглеродного анализа, попытки лингвистических датирований с помощью этого уравнения во многих случаях давали результаты, противоречащие историческим свидетельствам и опыту традиционного сравнительного языкознания, что привело, в конечном счёте, к весьма скептическому отношению научно-общества к лексикостатистике и, особенно, к глоттохронологии.

Принципиальные изменения в глоттохронологию внёс безвременно ушедший из жизни выдающийся лингвист, глава московской научной школы макрокомпаративистики С. А. Старостин. В результате проведённых им исследований было установлено, что причины указанных неудач заключаются не только в некорректном применении лексикостатистических методов из-за игнорирования заимствований и случайных подобию в сравниваемых языках, но и в принципиальном несо-

ответствии физических аналогий ряда постулатов М. Сводеша реальным лингвистическим процессам ([1–3]). Отказ от этих постулатов приводит к необходимости пересмотра основного уравнения глоттохронологии и построения его на иных исходных принципах. В частности, выводы, полученные в работах ([1–6]), позволяют предпринять попытку формирования основного уравнения глоттохронологии путём предварительного выбора общего вида математической модели распада языков на основе анализа лингвистических особенностей этого процесса и последующей идентификации параметров модели по совокупности исторически подтверждённых датировок основных списков.

Ниже излагаются математическое и лингвистическое обоснования таких моделей, их построение и обсуждение.

1. Модель общего распада языка

В разделе рассматривается математическое описание процесса распада ОС языка, проведён анализ соответствия известных формальных моделей лингвистическому содержанию этого процесса и предложен новый подход к построению глоттохронологических соотношений, сущность которого раскрывается на примере построения модели общего распада.

1.1. Математическое содержание процесса распада

Математическое содержание процесса распада ОС языка можно сформулировать в виде следующего утверждения: число распавшихся слов ∂N за отрезок времени ∂t пропорционально длительности ∂t этого отрезка и общему количеству нераспавшихся к данному моменту времени t слов $N(t)$ с коэффициентом пропорциональности λ :

$$-\partial N = \lambda N(t) \cdot \partial t.$$

Перепишав полученное соотношение в виде дифференциального уравнения

$$\frac{1}{\lambda} \cdot \frac{\partial N}{\partial t} + N(t) = 0,$$

и решив его, получим классическое уравнение глоттохронологии:

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

Зная число $N(t)$ слов из ОС, которые остались в данном языке, можно вычислить время, за которое объем ОС уменьшился от значения N_0 до $N(t)$:

$$t = -\frac{1}{\lambda} \cdot \ln\left(\frac{N(t)}{N_0}\right).$$

Поскольку $N_0 = 1$, то при вычислениях используют формулу

¹ Для удобства, при ссылках на математические выражения мы будем нумеровать наиболее важные из них по мере появления в тексте статьи.

$$t = -\frac{1}{\lambda} \cdot \ln(N(t)). \quad (2)$$

Практическое использование соотношений (1) и (2) для глоттохронологических вычислений показало, что во многих случаях наблюдается значительное расхождение расчётных значений t и датировок, полученных на основе культурно-исторических данных. Подробный анализ этих несоответствий проведён С. А. СТАРОСТИНЫМ ([1]), который установил, что причина неадекватности модели (1) заключается в неправомерности третьего и четвертого постулатов М. Сводеша, и уточнение модели (1) можно осуществить только путём пересмотра этих постулатов.

1.2. Пересмотр третьего постулата М. Сводеша

М. Сводеш в качестве базового значения коэффициента потерь предложил постоянное значение λ равное 0,14. Однако практически во всех исторически зарегистрированных ОС языков фактическое значение λ оказалось значительно меньшим ([1: 229–230]) и составило величину порядка 0,06 (табл. 1).

Таблица 1

Определение показателя распада λ для некоторых языков по историческим данным t и $N(t)$

Язык	t	$N(t)$	λ
Японский	1,2	0,93	0,06
Китайский	2,6	0,77	0,10
Английский	1,3	0,88	0,10
Немецкий	1,2	0,94	0,05
Французский	1,5	0,90	0,07
Испанский	1,5	0,91	0,06

При попытке использовать $\lambda = 0,06$ ([1: 231]) для вычисления времени распада других групп языков, как показано на примерах в табл. 2 и на рис. 1, мы снова получаем исторически неверные даты.

Таким образом, напрашивается вывод, что величина λ не является установившейся константой. Возникает следующая задача: как корректно определить коэффициент потерь λ ?

С. А. СТАРОСТИН отмечает [1: 233]: разница между теорией глоттохронологии и теорией, например, радиоактивного распада, для которой также применяется модель (1), заключается в том, что слова (корни), в отличие от атомных ядер, могут устаревать, т. е. чем больше рассматриваемый промежуток времени, тем больше вероятность исчезновения слова из исходного ОС. Тем самым предлагается пересмотреть третий постулат о постоянной доле p остающихся слов за равные промежутки времени. Было выдвинуто предположение, что доля p слов из ОС не постоянна, а меняется с течением времени: чем дольше слово «прожило» в языке, тем больше шансов, что оно скоро выпадет, т. е.

$$\lambda(t) = \lambda \cdot t,$$

где $\lambda(t)$ — коэффициент потерь в момент времени t ; λ — постоянный множитель (по С. А. Старостину $\lambda = 0,05$).

С учетом увеличивающегося роста потерь формула (1) трансформируется в формулу (3):

$$N(t) = N_0 \cdot e^{-\lambda t^2}. \quad (3)$$

Представленное на *рис. 1* сравнение кривых $N(t)$, построенных по выражениям (1) и (3), показывает, что уточненная формула (3) дает удовлетворительные результаты для недавних периодов времени и значительно менее правдоподобные — для древних периодов.

Попытка разрешить это противоречие приводит к необходимости пересмотреть четвёртый постулат.

1.3. Пересмотр четвёртого постулата М. Сводеша

В четвертом постулате М. Сводеша утверждается о равной вероятности исчезновения всех слов списка за одинаковые промежутки времени независимо от выбора интервала времени. На самом деле среди слов из ОС есть более устойчивые слова, сохраняющиеся на протяжении тысяч лет, а есть гораздо менее устойчивые. В силу этого наблюдается следующая зависимость: по мере выпадения слов из списка скорость распада λ уменьшается, поскольку начинает происходить повторная замена менее устойчивой части списка.

С. А. СТАРОСТИН предложил [1: 236, 237] учитывать корреляцию между λ и долей сохранившихся в ОС слов, т. е. полагать, что λ — величина, убывающая со временем пропорционально количеству остающихся слов, что описывается формулой:

$$\lambda(t) = \lambda \cdot N(t). \quad (4)$$

С одновременным учётом (4) и (3), получаем:

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

Отметим, что формула (5) демонстрирует противоречивость процесса лексических изменений: величина t^2 отражает резкий рост потерь, связанный с устареванием слов, а величина $N(t)$ — наоборот, замедление изменений, связанное с тем, что менее устойчивые слова исчезают из ОС, а более устойчивые сохраняются.

Кривая $N(t)$, построенная по выражению (5), также представлена на *рис. 1*, и свидетельствует о том, что введение второй поправки позволило получить лучший по сравнению с (1) и (3) вариант модели процесса распада ОС языка, который хорошо соответствует известным и наиболее вероятным предполагаемым датированиям в диапазоне $t < 4$ тыс. лет; однако при $t \geq 4$ тыс. лет мы по-прежнему получаем в результате расчета заниженные значения времени распада.

1.4. Критический обзор полученных результатов

Проведенный выше анализ процесса поэтапного улучшения глоттохронологических выражений на основе установления дополнительных взаимосвязей между темпом распада λ и другими параметрами позволяет выявить недостатки данного подхода и предложить способы их устранения.

1. Использование методики введения поправок в зависимость $N(t)$ для улучшения точности датирования в некотором определённом диапазоне t нецелесообразно, так как оно либо недостаточно эффективно в указанном временном диапазоне, либо приводит к ухудшению точности в другом диапазоне.

Действительно, рассмотрим переход от зависимости $N_1(t) = e^{-0,14t}$ к $N_2 = e^{-0,06t}$, предложенной С. А. Старостиным. Очевидно, что на интервале $t < 1 \dots 1,5$ тыс. лет график функции $N_2(t)$ (см. *рис. 1а, б*) соответствует опорным точкам гораздо лучше, чем $N_1(t)$ (сравним значения $N_1(t)$ и $N_2(t)$ во 2-ом и 3-ем столбцах *табл. 2* с контрольными значениями в последнем столбце: 0,11 и 0,25 — с 0,6 тыс. лет в первой строке, 0,51 и 1,21 — с 1,2 тыс. лет во второй, 0,75 и 1,76 — с 1,5 тыс. лет в третьей). Однако, при $t > 1,5$ тыс. лет кривая проходит значительно выше опорных точек, что свидетельствует о растущем отклонении значений функции $N_2(t)$ от известных (4,4 по сравнению с 2,6 тыс. лет, 10,6 по сравнению с 5,5 тыс. лет и 15,7 по сравнению с 11 тыс. лет в 4-ой, 5-ой и 6-ой строках *табл. 2* соответственно).

Для устранения этого недостатка С. А. Старостин вводит в зависимость $N_2(t)$ «ускоряющую» поправку $\lambda(t) = \lambda \cdot t$. Переход к $N_3(t) = e^{-0,05t}$ улучшил датирование в интервале $t = 1 \dots 2$ тыс. лет, но привел к чрезмерному «ускорению» функции распада, т. е. занижению дат при $t > 2$ тыс. лет (см. аналогично изменение графика функции $N_3(t)$ на *рис. 1а, б* и её значений в 4-ом столбце *табл. 2*); при этом для $t < 1$ оказалось, что процесс $N_3(t)$ не только не ускорился по отношению к $N_2(t)$, а наоборот, замедлился².

Для компенсации отмеченного выше чрезмерного «ускорения» в $N_3(t) = e^{-0,05t^2}$ С. А. Старостиным была предложена «замедляющая» поправка $\lambda(t) = 0,05 \cdot N(t) \cdot t$ и осуществлен переход к соотношению $N_4(t) = e^{-0,05 \cdot N_4(t) \cdot t^2}$, которое улучшило датировки для небольших t , но оказалось недостаточно эффективным для $t > 4$ тыс. лет (график функции на *рис. 1а, б* и датировки в 5-го столбце *табл. 2* указывают на растущее отклонение значений функции от контрольных в сторону занижения).

² Это явление легко объяснить математически тем, что квадратичная функция $\lambda(t) = t^2$ изменяется быстрее, чем функция $\lambda(t) = t$ лишь при $t > 1$, а при $t < 1$ возведение в квадрат соответствует замедлению процесса распада. Однако данная особенность квадратичной функции при $t < 1$ противоречит общей концепции постоянного убыстрения процесса распада, принятой в п. 1.2.

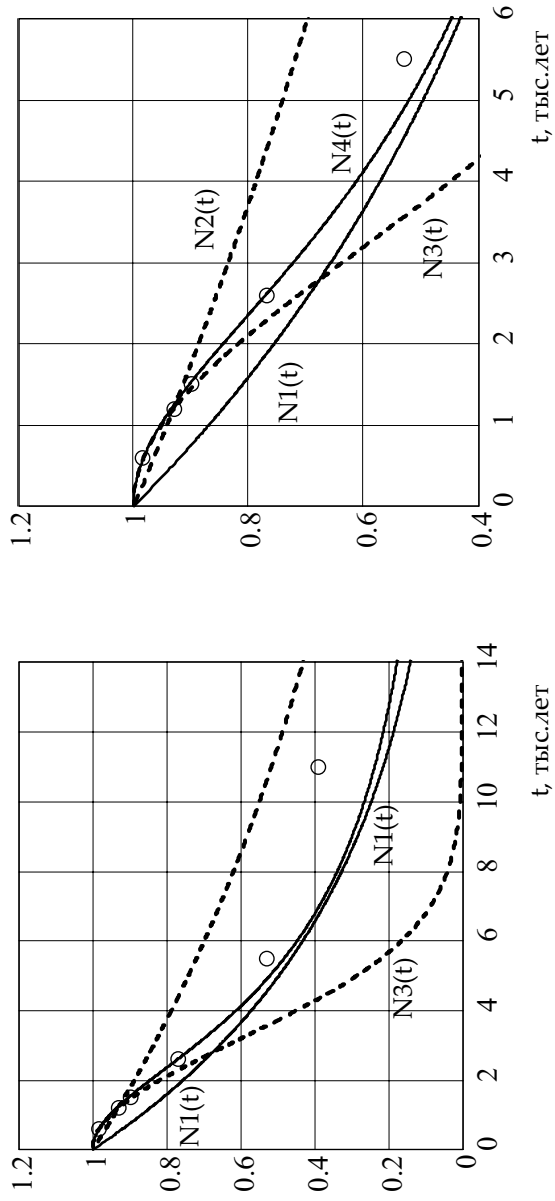


Рис. 1. Глоттохронологические модели М. Сводеша (N1) и

С. А. Старостина (N2, N3, N4):

$$N1(t) = e^{-0.14t}, N2(t) = e^{-0.06t}, N3(t) = e^{-0.05t^2}, N4(t) = e^{-0.05N4(t)t^2}.$$

Таким образом, формирование глоттохронологических выражений путём введения поправок, улучшающих точность датировок в том или ином диапазоне времени t , является нецелесообразным. Следовательно, математическую модель необходимо строить так, чтобы она обеспечивала наименьшее среднее отклонение от опорных точек (известных и предполагаемых внелингвистических датировок) на всём временном интервале её использования.

Таблица 2

Сравнительный анализ датирования распада языков с помощью различных глоттохронологических соотношений

Язык	N(t)	Время распада, найденное для различных моделей, тыс. лет				Известные и предполагаемые датировки распада ОС праязыка, тыс. лет
		N_1 $\lambda = 0,14$	N_2 $\lambda = 0,06$	$N_3 \lambda(t) \parallel$ $0,05t$	$N_4 \lambda(t) \parallel$ $0,05 \cdot N(t) \cdot t$	
Белорусский	0,985	0,11	0,25	0,55	0,55	0,6
Японский	0,93	0,51	1,21	1,2	1,25	1,2
Французский	0,90	0,75	1,76	1,45	1,53	1,5
Китайский	0,77	1,9	4,4	2,3	2,6	2,6
Индоевропейские языки (без анатолийской ветви)	0,53	4,5	10,6	3,6	4,9	5,5
Гипотетический праязык урало-алтайской макросемьи	0,39	6,7	15,7	4,3	6,9	≈ 11

2. Попытка интерпретации изменения скорости распада как изменения λ с течением времени приводит нас к противоречию в понимании характера происходящих изменений: увеличение скорости распада, связанное с устареванием слов, с одной стороны, и замедление скорости изменений, связанное с постепенным преобладанием в ОС наиболее устойчивых значений — с другой.

Таким образом, нам следует привести иные причины, которые могут объяснить изменение скорости распада и, в то же время, позволят избежать противоречий. Сформулируем их в виде следующих утверждений:

1) Все значения ОС можно разделить на несколько групп, обладающих различной устойчивостью. Тогда распад ОС можно представить как распад нескольких составляющих с различным коэффициентом распада λ . Очевидно, что с течением времени группы слов с меньшим значением λ , т. е. более устойчивые, начнут постепенно преобладать среди незаменившихся значений ОС, что соответствует общему замедлению процесса распада³.

³ Данное утверждение аналогично тезису, приведённому в п. 1.3. Однако на этот раз обсуждение ведётся в рамках уже новых предпосылок, а именно:

2) Среди значений ОС существует группа наиболее устойчивых значений, распадающихся так медленно, что коэффициент их распада λ можно принять за ноль. Будем называть эту группу «ядром» ОС. Тогда процесс лексических замен можно рассматривать как устремление неизменившейся части ОС к данному постоянному ядру, что численно выражается в замедлении процесса распада.

Приведённое объяснение мы обнаруживаем в работах [1–4], в которых прямо указывается на несостоятельность четвёртого постулата М. Сводеша и на существование в одном языке разделов ОС с различными коэффициентами распада⁴. В частности, в работе [3] СТАРОСТИНЫМ вводится понятие индекса стабильности, позволяющее ранжировать значения в ОС по степени их устойчивости, и экспериментально, на основе анализа ОС 14-ти языковых семей подтверждается неравномерность скорости развития различных разделов базисной лексики. При этом ему удалось установить, что «при любом способе (ранжирования устойчивости значений ОС — прим. авторов) однозначно выделяется максимально стабильная часть списка приблизительно в двадцать с небольшим значений» ([3: 8]).

1.5. Предлагаемый подход к построению модели

Опираясь на сделанные выводы, сформируем новый подход к построению глоттохронологической модели.

1. Учитывая несостоятельность четвёртого постулата, будем рассматривать процесс лексических замен некоторого значения ОС как поток случайных событий, который описывается экспонентой $e^{-\lambda t}$ (см. п. 1.1) с некоторой постоянной интенсивностью λ . Тогда процесс общего распада языка можно представить как сумму нескольких элементарных потоков (обозначим их количество через i) с соответствующими интенсивностями λ_i ([4: 29]):

$$N_{\text{общ}}(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t} + \dots + c_i e^{-\lambda_i t} = \sum_i c_i e^{-\lambda_i t}, \quad (6)$$

где $\sum_i c_i = N_0$ а число слагаемых i теоретически равно числу значений в ОС (например, в случае со 100-словным списком $i = 100$). Однако, в соответствии с теоремой А. Я. ХИНЧИНА ([5: 58–68]⁵), сумму независимых, но интерпретация изменения скорости распада путём разбиения значений ОС на группы с различной устойчивостью; неизменность коэффициента распада λ в каждой из этих групп.

⁴ Заметим, что отказ от четвёртого постулата М. Сводеша неизбежно влечёт за собой отказ и от третьего постулата, т. к. последний справедлив только для модели вида (1), которая описывает процесс распада одной экспонентой.

⁵ В работе [5] математик А. Я. ХИНЧИН установил и доказал свойство сходимости суммы независимых, но сравнимых по интенсивности потоков случайных событий к одному потоку, интенсивность которого близка к сумме его составляющих.

сравнимых по интенсивности потоков случайных событий можно заменить некоторым одним потоком. Таким образом, следует ожидать, что реальное количество слагаемых i , которое определится в процессе расчёта параметров модели (6), будет значительно меньше 100.

Поэтому начальный вид математической модели процесса распада можно задать при помощи всего нескольких типовых слагаемых, например, трёх:

$$N(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t}.$$

Тогда, в результате расчёта неизвестных параметров модели ($c_1, c_2, c_3, \lambda_1, \lambda_2, \lambda_3$), мы сможем определить:

- а) сколько из слагаемых вида $c_i e^{-\lambda_i t}$ являются значимыми ($c_i \neq 0$);
- б) есть ли среди составляющих незначимые ($c_i = 0$);
- в) какие из значимых составляющих отражают процесс распада ($\lambda_i \neq 0$);
- г) есть ли среди значимых составляющих константа ($\lambda_i = 0$).

Если окажется, что все составляющие являются значимыми, то их количество можно будет увеличить.

2. Будем искать такие, заранее не известные параметры $c_1, c_2, c_3, \lambda_1, \lambda_2, \lambda_3$, при которых модель заданного вида $N(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t}$ обеспечит наименьшее совокупное расхождение расчётных и фактических значений $N(t)$ в каждой опорной точке.

1.6. Определение параметров модели общего распада

Ниже приводится таблица исходных данных, для составления которой использовались сведения из [1], <http://starling.rinet.ru>, а также материалы, любезно предоставленные А. В. Дыбо (табл. 3). В каждой строке указывается названия языка-предка и его потомка(ов), доля совпадений их ОС (столбец «Среднее значение»)⁶ и временной интервал между датировками ОС языка-предка и его потомков. Таким образом, каждая строка таблицы представляет собой опорную точку, которая состоит из значения времени (t_i), соответствующего продолжительности распада, и значения совпадения ($N_{ср,i}(t)$), соответствующего доле слов, сохранившихся в ОС за указанный интервал времени.

⁶ В случае, если имелись сведения более чем об одном потомке, проводилось попарное сравнение ОС языка-предка с ОС каждого из языков-потомков. При этом наименьшее и наибольшее значения совпадений указаны в столбцах «Нижнее значение» и «Верхнее значение» соответственно; для расчёта использовалось среднее арифметическое от всех значений, приведённое в столбце «Среднее значение».

Таблица 3

Опорные данные для расчёта модели общего распада

№ точки i	Сравниваемые языки	Нижнее значение $N_{н,i}(t)$	Среднее значение $N_{ср,i}(t)$	Верхнее значение $N_{в,i}(t)$	t_i , тыс. лет
1	Любые	—	1	—	0
2	Рунический тюркский — карханидский	—	0,97	—	0,3
3	Среднекорейский — современные потомки	0,94	0,95	0,96	0,5
4	Древневерхненемецкий — современный немецкий	—	0,95	—	1,1
5	Древнеанглийский — современный английский	—	0,87	—	1,1
6	Древнеяпонский — современные потомки	0,89	0,91	0,94	1,2
7	Народная латынь — современные романские	0,82	0,86	0,89	1,5
8	Древнегреческий — новогреческий	—	0,78	—	2,6
9	Древнекитайский — современные потомки	0,66	0,69	0,77	3

Чтобы определить параметры модели по исходным данным воспользуемся методом наименьших квадратов. Для этого необходимо найти такие $c_1, c_2, c_3, \lambda_1, \lambda_2, \lambda_3$, при которых сумма квадратов разностей между расчётными значениями модели $N(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t}$ и соответствующими значениями $N_{ср,i}(t)$ будет наименьшей для всей совокупности опорных точек.

Например, для строки 3 табл. 3 получаем выражение:

$$(c_1 e^{-\lambda_1 \cdot 0,5} + c_2 e^{-\lambda_2 \cdot 0,5} + c_3 e^{-\lambda_3 \cdot 0,5} - 0,95)^2 \rightarrow \min.$$

Составив подобные выражения для каждой строки таблицы и суммировав их, придем к задаче поиска минимума общей невязки⁷:

$$\sum_{i=1}^n (c_1 e^{-\lambda_1 t_i} + c_2 e^{-\lambda_2 t_i} + c_3 e^{-\lambda_3 t_i} - N_{ср,i})^2 \rightarrow \min^8.$$

⁷ Невязкой называется отклонение расчётных значений, полученных при моделировании процесса, от значений действительных.

⁸ Практически при поиске параметров следует также ввести ограничения, обеспечивающие получение неотрицательных параметров модели: $c_1 \geq 0, c_2 \geq 0, c_3 \geq 0, \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0$. Использование отрицательных параметров позволяет значительно уменьшить невязку, однако наличие в модели (6) отрицательных ($c_i < 0$) или растущих ($\lambda_i < 0$) составляющих необъяснимо в рамках предложенного в п. 1.5 подхода. Кроме того, в число ограничений может быть включена невязка для первой строки ($i = 1$ и $t_1 = 0$):

$$c_1 + c_2 + c_3 = 1.$$

С помощью математического пакета Mathcad получены следующие значения параметров модели:

$$c_1 = 0,196; c_2 = 0,211; c_3 = 0,593; \lambda_1 = 0,000; \lambda_2 = 0,135; \lambda_3 = 0,153;$$

Таким образом, полученная модель общего распада имеет три значимые составляющие. При этом одна из них ($c_1 = 0,196, \lambda_1 = 0,000$) — является константой, а две другие отражают два экспоненциальных процесса с близкими коэффициентами распада ($\lambda_2 = 0,135$ и $\lambda_3 = 0,153$). Искомая модель общего распада может быть приведена к простому виду (рис. 2)⁹:

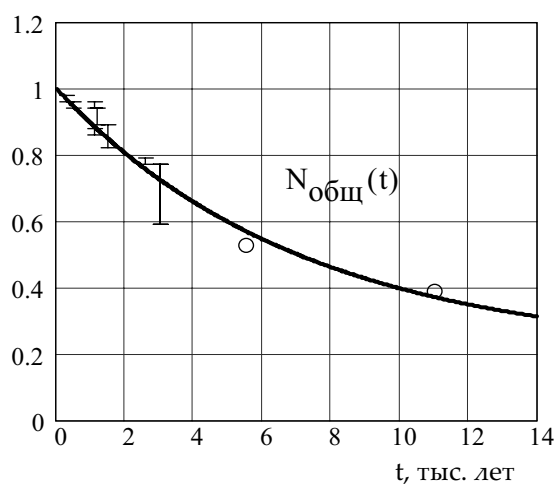


Рис. 2. График модели общего распада

$$N_{\text{общ}}(t) = 0,2 + 0,8e^{-0,14t}$$

$$N_{\text{общ}}(t) = 0,2 + 0,8e^{-0,14t} . \quad (7)$$

1.7. Обсуждение модели

Прежде всего, следует отметить наличие в модели устойчивого ядра ([3]), составляющего около 20% ОС и распадающегося настолько медленно, что коэффициент его распада λ можно принять за ноль. При этом, как уже говорилось, приближение $N_{\text{общ}}(t)$ к этому постоянному ядру интерпретируется как замедление общего процесса распада по сравнению с моделью (1).

⁹ Для приведения модели к простому виду (7) две составляющие с различными коэффициентами $\lambda_2 = 0,135$ и $\lambda_3 = 0,153$ были сведены в одну с $\lambda = 0,14$. Проведённое численное и графическое сопоставление выражения (7) и исходного выражения показало, что различия между ними пренебрежимо малы по сравнению с величиной доверительного интервала $N_{н,i}(t) \dots N_{в,i}(t)$.

Как и ожидалось, наличие полученной модели устойчивого ядра опровергает четвёртый, а, следовательно, и третий постулат Сводеша о постоянной доле слов в ОС, сохраняющихся за равные промежутки времени.

Адекватность числовых значений параметров модели обеспечивается самим способом их получения — методом наименьших квадратов отклонений от реальных опорных данных, причём разброс в расположении этих данных (см. рис. 2) свидетельствует о статистическом характере функции $N_{\text{общ}}(t)$, усреднённый характер которой не отражает индивидуальных особенностей развития конкретных языковых семей.

Сопоставим новую модель общего распада $N_{\text{общ}}(t) = 0,2 + 0,8e^{-0,14t}$ с известной моделью $N_4(t) = e^{-0,95 \cdot N_4(t) \cdot t^2}$, которая была получена путём введения поправок, предложенных С. А. Старостиным (рис. 3). Очевидно, что модель $N_4(t)$ хорошо совпадает только с отдельными опорными точками, в то время как $N_{\text{общ}}(t)$ занимает средневзвешенное положение относительно всех опорных точек и обеспечивает минимальную среднюю невязку датировок в целом — в широком диапазоне t .

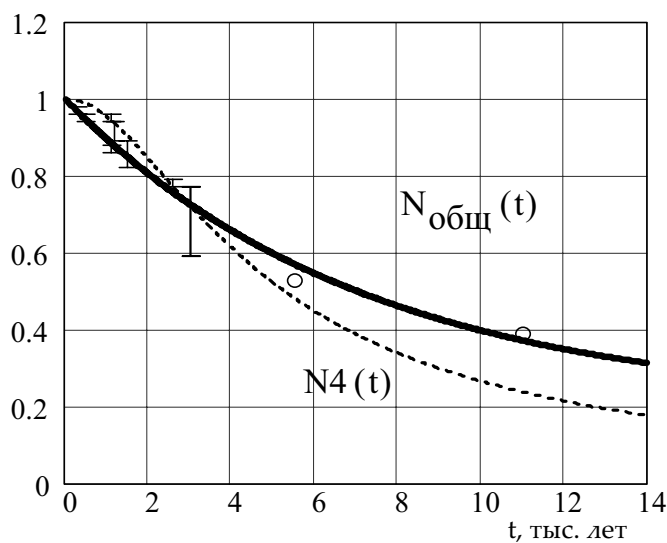


Рис. 3. Сравнение моделей общего распада
 $N_{\text{общ}}(t) = 0,2 + 0,8e^{-0,14t}$, $N_4(t) = e^{-0,95 \cdot N_4(t) \cdot t^2}$

Модель (7) может быть использована при реконструкции ОС языков прямым вычислением $N_{\text{общ}}(t)$ для заданного интервала времени t . Напри-

мер, доля общих слов в ОС современных потомков и праязыка индоевропейской семьи, предположительно распавшегося 6 тыс. лет назад, составит

$$N_{\text{общ}}(t) = 0,2 + 0,8e^{-0,14 \cdot 6} = 0,55;$$

а для праурало-алтайского языка с предполагаемой продолжительностью распада $t = 11$ тыс. лет

$$N_{\text{общ}}(t) = 0,2 + 0,8e^{-0,14 \cdot 11} = 0,37.$$

Разумеется, эти результаты носят статистический характер.

2. Модель относительного распада языков

2.1. Пересмотр пятого постулата М. Сводеша

Пятый постулат утверждает, что развитие любых двухязыков-потомков одного праязыка происходит независимо. Это даёт возможность, зная долю общих слов $N_n(t)$ в ОС нескольких (n) родственных языков, определить, с какого времени эти языки существуют самостоятельно (т. е. время разделения).

Например, рассмотрим два родственных языка А и В. Процесс распада каждого из них независим от другого и описывается классическим уравнением глоттохронологии (1):

$$N_A(t) = N_B(t) = N(t) = e^{-\lambda t}.$$

Тогда дивергенцию языков А и В относительно друг друга можно представить в виде произведения:

$$N_2(t) = N_A(t) \cdot N_B(t) = N(t) \cdot N(t) = e^{-\lambda t} \cdot e^{-\lambda t} = e^{-2\lambda t}, \quad (8)$$

или $N_2(t) = N_2(t)$.

Таким образом, модель $N_2(t)$ относительного распада языков получается непосредственно из модели общего распада $N(t)$.

Переписав выражение (8) относительно t , получаем простое выражение, позволяющее вычислить время, прошедшее с момента разделения языков А и В, по известной доле совпадений $N_2(t)$ между их ОС:

$$t = -\frac{1}{2\lambda} \cdot \ln N_2(t),$$

Несмотря на очевидную привлекательность рассмотренного подхода к описанию процесса относительного распада, правомерность лежащего в его основе пятого постулата ставится под сомнение многими исследователями. Так, в работе [4: 27, 29] отмечается, что непосредственно после разделения наблюдается согласованное развитие языков-потомков, при котором в их основных списках происходят одни и те же замены. То есть, несмотря на территориальное разделение, в развитии ОС языков продолжает прослеживаться единство. При этом их независимая эволюция проявляется по мере ослабления общих тенденций развития и начинает превалировать в процессе распада лишь спустя некоторое время.

Этот факт можно проиллюстрировать на примере достаточно непродолжительного по времени (0,5 тыс. лет) распада среднекорейского языка (табл. 4).

Таблица 4¹⁰
Согласованное изменение стословных списков потомков среднекорейского языка после разделения

Значение ОС	Средне- корейский, АКО	О-в Чонсандо [Чхонсандо] (청산도, 靑山島), CHN	Пхёнъян- [Пхёнъян-] намдо (평안[평양] 남도, 平安[平壤] 南道), PNM	Кёнсандо (경상도, 慶尙道), KJN	Хамгёндо (함경도, 咸鏡道), HMG	Чеджудо (제주도, 濟州島 [濟州道]), CJD	Сеул (서울), SEU	Канвондо (강원도, 江原道), KWD
‘перо’	čís	thɔl	thɔl	"	thɔl	"	thɔl	thɪl
‘волос’	thərí, thə̀rək	khal	k̄al	khál	khál	–	qhal	"
‘колени’	mùrùp(h), mùràp	"	"	céngéj	"	"	"	oɥumpé
‘полный’	kăták- / kătáik-	"	"	"	cháuda	"	čáwúda	"
‘голова’	mərí	"	"	"	"	kol	"	kol
‘кожа’	káčòk, káčh	"	"	"	"	koptegi	"	"

" — слово не изменилось или было заимствовано.

¹⁰ 10-словные списки корейских языков, использованные при составлении таблицы, представлены в базе данных kor.dbf, созданной в рамках проектов «Вавилонская башня» и «Evolution of Human Languages» и доступной на веб-сайте <http://starling.rinet.ru>.

Анализ таблицы показывает, что относительно 110-словного списка среднекорейского языка в ОС языков-потомков произошла замена шести значений:

одно (кожа) — лишь в одном Чеджудо;

три (колени, полный, голова) — в двух языках (из них два — на одних те же слова);

два значения (перо, волос) — в пяти языках из семи заменились также на одни и те же слова.

Таким образом, значительная доля замен в ОС недавно образовавшихся потомков происходит согласованно и замедляет процесс дивергенции этих языков. Очевидно, данное согласованное развитие не учитывается в модели общего распада $N_{\text{общ}}(t)$, что приводит к необходимости построения отдельной модели относительного распада языков.

2.2. Математическое содержание процесса относительного распада

Рассмотрим два родственных языка-потомка, и обозначим через $N(t)$ число совпадающих слов в ОС двух языков на момент времени t , прошедшего с момента разделения этих языков.

Тогда процесс их относительного распада можно представить в виде двух составляющих

$$N(t) = N_c(t) + N_n(t), \quad (9)$$

где $N_c(t)$ — значения, которые изменяются в ОС обоих языков-потомков согласованно и замещаются одинаковыми лексемами. Будем называть их связанной составляющей относительного распада;

$N_n(t)$ — значения, которые замещаются в ОС языков-потомков независимо друг от друга различными лексемами. Назовём их независимой составляющей.

Определим содержание каждой из этих составляющих.

Связанная составляющая соответствует той части ОС языков-потомков, в которой действуют общие законы развития, унаследованные от языка-предка. Фактически данные законы отражают собой систему семантических связей и переходов языка-предка, внутри которой все основные лексемы, соответствующие значениям ОС имеют одного или нескольких «конкурентов», претендующих на то же место в словесном списке. Именно их количеством и силой «конкуренции» определяется частота и последовательность происходящих замен. В момент распада данная система без изменений наследуется всеми языками-потомками и продолжает независимо функционировать в каждом из них так, как если бы разделения не происходило, и все они оставались бы одним языком. Именно этим объясняется согласованное развитие языков-потомков, при котором мы наблюдаем одни и те же замены, происходящие с одинаковой частотой и последовательностью.

С другой стороны, сразу после распада система каждого языка начинает изменяться, перестраиваться под воздействием специфических условий той внешней среды, в которую он попал. Происходят изменения в составе конкурирующих лексем, а также их иерархии. Однако эти, уже произошедшие изменения, начинают проявляться в ОС языков только по мере происходящих замен. Поэтому мы продолжаем фиксировать совпадение некоторого значения в ОС языков-потомков, в то время как сегменты системы, соответствующие данному значению в каждом из языков, уже не совпадают и подчиняются различным законам развития. Следовательно, данное значение ОС будет замещаться в каждом из языков-потомков различными лексемами независимо друг от друга, что по определению означает его переход из связанной составляющей в независимую составляющую.

Собственно же дивергенция стословных списков языков-потомков есть проявление несогласованного развития значений внутри независимой составляющей.

Построим математическую модель, соответствующую приведённой интерпретации процесса относительного распада. При этом первоначально будем исходить из наиболее простых соображений и усложнять модель только по мере необходимости.

Допустим, что количество значений в связанной составляющей $N_c(t)$ убывает экспоненциально, т. е. за каждый отрезок времени dt число значений связанной составляющей N_c уменьшается на величину dN_c пропорциональную длительности этого отрезка и общему количеству значений $N_c(t)$, не изменившихся на данный момент времени t , с коэффициентом пропорциональности μ , т. е.

$$-dN_c = \mu N_c(t) \cdot dt.$$

Перепишем полученное соотношение в виде дифференциального уравнения:

$$\frac{1}{\mu} \cdot \frac{\partial N_c}{\partial t} + N_c(t) = 0,$$

Решив его, получим:

$$N_c(t) = N_c(0) \cdot e^{-\mu t},$$

где t — время, прошедшее с момента разделения языков (в тысячелетиях); $N_c(0) = N_0$ — количество значений из ОС, входящих в связанную составляющую на момент разделения¹¹.

Убывание функции $N_c(t) = N_0 \cdot e^{-\mu t}$ соответствует переходу значений из связанной составляющей в независимую с некоторым коэффициентом μ (чем больше μ , тем быстрее уменьшается $N_c(t)$).

¹¹ Очевидно, что в момент распада в связанную составляющую по определению входят все значения ОС. Т. е. $N_c(0) = N_0 = 1$.

Действительно, в соответствии с выражением (9) изменение независимой составляющей $N_H(t)$ можно представить в виде разности:

$$\begin{aligned} N_H(t) &= N(t) - N_C(t), \\ N_H(t) &= N(t) - N_0 \cdot e^{-\mu t}, \end{aligned}$$

т. е. с уменьшением связанной составляющей $N_C(t)$ возрастает доля независимо развивающихся значений $N_H(t)$.

В свою очередь, изменение составляющей $N_H(t)$ отражает уже собственно процесс расхождения языков-потомков. Предположим, что её убывание происходит также экспоненциально: уменьшение числа совпадающих слов ∂N пропорционально времени ∂t и $N_H(t)$ с некоторым коэффициентом пропорциональности η — коэффициентом потерь совпадающих слов, т. е.

$$-\partial N = \eta N_H(t) \cdot \partial t,$$

или
$$-\partial N = \eta(N(t) - N_0 \cdot e^{-\mu t}) \partial t.$$

Приходим к неоднородному дифференциальному уравнению

$$\frac{1}{\eta} \cdot \frac{dN}{dt} + N(t) = N_0 \cdot e^{-\mu t}. \quad (10)$$

В результате решения уравнения (10)¹², получаем выражение, описывающее относительный распад языков-потомков:

$$N_{\text{отн}}(t) = N_0 \left(\frac{\mu}{\mu - \eta} \cdot e^{-\eta t} + \frac{\eta}{\eta - \mu} \cdot e^{-\mu t} \right).$$

При этом допустим возможность существования совпадающей части среди значений, входящих в устойчивое ядро ОС каждого из языков-потомков. Для этого введём в полученное выражение постоянную составляющую c_0 и получим модель относительного распада

полного вида:
$$N_{\text{отн}}(t) = c_0 + c_1 \left(\frac{\mu}{\mu - \eta} \cdot e^{-\eta t} + \frac{\eta}{\eta - \mu} \cdot e^{-\mu t} \right), \quad (11)$$

и упрощённого вида (при $\mu = \eta$):
$$N_{\text{отн}}(t) = c_0 + c_1 \cdot e^{-\eta t} (1 + \eta \cdot t), \quad (12)$$

в которых $c_0 + c_1 = N_0$.

2.3. Определение параметров модели относительного распада

Исходные данные для расчёта параметров модели относительного распада получены из тех же источников, которыми мы пользовались при заполнении табл. 3 (см. п. 1.6). Строение таблиц также аналогично. Отличие заключается в том, что сравнение ОС производится не между языком-предком и его потомками, а между всеми языками-потомками попарно. При этом величина t_1 соответствует времени, прошедшему с момента разделения.

¹² Полный вывод см. в Приложении.

Таблица 5

Опорные данные для расчёта модели относительного распада

Номер точки i	Сравниваемые языки	Нижнее значение $N_{н,i}(t)$	Среднее значение $N_{ср,i}(t)$	Верхнее значение $N_{в,i}(t)$	t_i , тыс. лет
1	Любые	—	1	—	0
2	Современные потомки среднекорейского	0,92	0,96	1	0,5
3	Белорусский — украинский (потомки восточнославянского)	—	0,97	—	0,6
4	Современные исландский, риксмол, датский, шведский, гвестаг	0,93	0,95	0,96	1
5	Современные японские языки (исключая диалекты островов Рюкю)	0,92	0,95	0,98	1,4
6	Современные потомки праславянского (белорусский, украинский, русский — сербский, болгарский, польский, македонский)	0,76	0,83	0,88	1,4
7	Современные потомки народной латыни (французский, испанский, румынский)	0,84	0,86	0,87	1,5
8	Современные огузские языки	0,84	0,9	0,96	1,7
9	Современные индо-иранские языки	0,47	0,56	0,65	3,2
10	Современные потомки праиндоевропейского (персидский, армянский, новогреческий, албанский, французский)	0,23	0,29	0,32	6

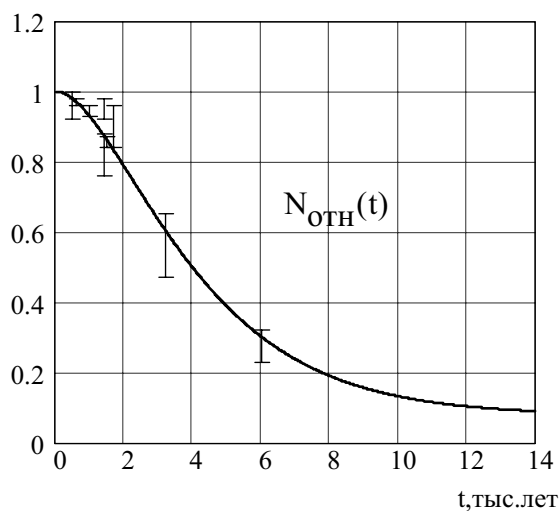


Рис. 4. Кривая процесса относительного распада $N_{\text{отн}}(t)$

В результате применения метода наименьших квадратов отклонений (см. п.1.6), для модели вида

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

были получены следующие значения параметров:

$$c_0 = 0,079; c_1 = 0,921; \eta = 0,454; \mu = 0,457.$$

Таким образом, модель относительного распада состоит из двух составляющих: постоянной (c_0), которая соответствует совпадающей части ядра величиной примерно в 8 значений, и распадающейся (c_1) — около 92 значений. Подставив значения параметров в выражение (11), получим модель полного вида:

$$N_{\text{отн}}(t) = 0,08 + 0,92 \cdot (158,494 \cdot e^{-0,454t} - 157,494 \cdot e^{-0,457t}),$$

которая, в силу практического равенства $\eta = 0,454$ и $\mu = 0,457$, может быть представлена в упрощённой форме (12) с $\eta = \mu = 0,45$:

$$N_{\text{отн}}(t) = 0,08 + 0,92 \cdot e^{-0,45t} (1 + 0,45t). \quad (13)$$

График функции представлен на рис. 4.

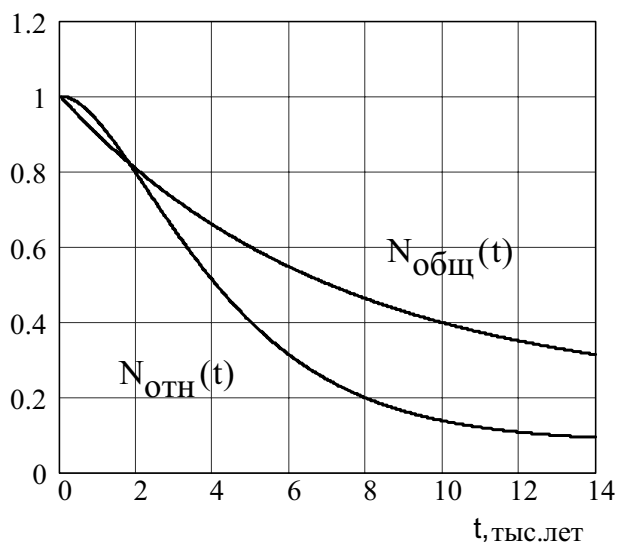


Рис.5. Сравнение моделей общего $N_{\text{общ}}(t)$ и относительного $N_{\text{отн}}(t)$ распада языков

2.4. Обсуждение модели относительного распада

В первую очередь сравним полученную модель (13) с моделью (7) общего распада (рис. 5).

Анализ рис. 5 указывает на сравнительно замедленный относительный распад родственных языков в пределах первых полутора-двух тысячелетий после их разделения и на проявляющееся затем всё более заметное расхождение лексического состава их ОС, постепенно замедляющееся по мере приближения к некоторому, примерно 8-процентному, общему ядру ОС.

Взаимное временное расхождение процессов общего и относительного распада раскрывает причины противоречий, связанных с использованием модели общего распада для определения дат разделения языков ([1: 232]): заниженные датировки для значений $N(t)$, близких к единице, и завышенные — для малых $N(t)$ у давно разделившихся языков.

Получение модели относительного распада в более простой форме (12) из возможных двух форм (11) и (12) свидетельствует о примерном равенстве интенсивности μ процесса уменьшения числа связанно развивающихся слов $N_c(t)$ в ОС родственных языков-потомков, и интенсивности η процесса уменьшения числа совпадающих независимо развивающихся слов ОС.

Следует отметить, что появление достаточно надёжных опорных точек для $t > 6$ тыс. лет (см. табл. 5 и рис. 4) может существенно повлиять на значение постоянной составляющей $c_0 = 0,08$ модели (13); показатель $\eta = 0,45$ при этом существенно не изменится.

С практической точки зрения важно обратить внимание, что формула (13) аналитически неразрешима относительно величины t . Другими словами, мы не можем определить время, прошедшее с момента распада, подставив количество совпадений в некоторое выражение, так как последнего не существует. Поэтому для глоттохронологических вычислений с помощью этой модели необходимо пользоваться математическими программами с готовым набором численных методов решения уравнений или таблицей, содержащей значения времени t при заданной величине совпадений ОС — $N_{отн}(t)$.

В качестве примера рассмотрим определение даты распада диалектов островов Рюкю. Из таблицы совпадений основных списков этих диалектов определяется средняя доля общих слов $N_{отн}(t) = 0,87$.

Таблица 6

Таблица совпадений ОС диалектов островов Рюкю				
Языки	NAS	SHU	HAT	YON
NAS		0,92	0,89	0,84
SHU	0,92		0,84	0,9
HAT	0,89	0,84		0,9
YON	0,84	0,9	0,9	

[NAS — Надзэ (名瀬), SHU — Сюри (首里), HAT — Хатэрума (波照間), YON — Ёнагуни (与那国)]

Составляется уравнение

$$0,87 = 0,08 + 0,92e^{-0,45t} \cdot (1 + 0,45t),$$

решение которого в пакете Mathcad имеет вид

```
t := 0
Given
0,87 = 0,08 + 0,92·e-0,45t·(1 + 0,45·t)
Find(t) = 1,46
```

Полученное значение 1,46 тыс. лет равно искомому времени, прошедшему с момента распада этих диалектов, или дате распада 540 г. н.э.

Фрагмент таблицы с заранее вычисленными решениями уравнения относительного распада представлен в табл. 7. Используя её, лег-

ко убедиться, что заданному значению $N_{\text{отн}}(t) = 0,87$ действительно соответствует искомое значение $t = 1,46$ тыс. лет.

Таблица 7

Значения функции относительного распада
и соответствующие им значения аргумента t .

$$N_{\text{отн}}(t) = 0,08 + 0,92e^{-0,45t}(1 + 0,45t)$$

$N_{\text{отн}}(t)$	t , тыс. лет	$N_{\text{отн}}(t)$	t , тыс. лет	$N_{\text{отн}}(t)$	t , тыс. лет	$N_{\text{отн}}(t)$	t , тыс. лет
0,90	1,24	0,84	1,67	0,79	2,00	0,73	2,40
0,89	1,32	0,83	1,74	0,78	2,07	0,72	2,46
0,88	1,39	0,82	1,81	0,77	2,14	0,71	2,53
0,87	1,46	0,81	1,87	0,76	2,20	0,70	2,60
0,86	1,53	0,80	1,94	0,75	2,27	0,69	2,66
0,85	1,60			0,74	2,33		

Сопоставим предлагаемую модель $N_{\text{отн}}(t)$ с известной моделью

$$N_2(t) = e^{-2 \cdot 0,05 \sqrt{N_2(t)} \cdot t^2},$$

построенной на базе 5-го постулата (см. рис. 6). Очевидно, что модель $N_2(t)$ хорошо соответствует лишь отдельным опорным точкам¹³ на небольших интервалах t . В отличие от неё, новая модель $N_{\text{отн}}(t)$ занимает средневзвешенное положение и обеспечивает наименьшее общее отклонение от известных опорных точек на всём временном интервале.

Вместе с тем, на графике хорошо видно, что любые значения, полученные с использованием данной модели, носят статистический, усреднённый характер. Поэтому они могут использоваться только в качестве предварительных оценок датировки распада и требуют дальнейшего уточнения путём привлечения для расчётов дополнительных внелингвистических сведений.

¹³ Исходные данные представлены на графике в виде вертикальных отрезков с верхней и нижней ограничительными чертами, которые маркируют верхнее $N_{в,i}(t)$ и нижнее $N_{н,i}(t)$ значения соответствующих строк табл. 6.

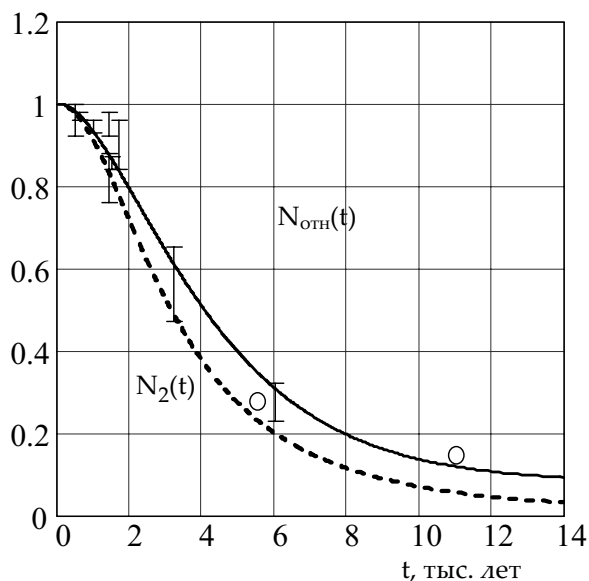


Рис. 6. Сравнение моделей относительного распада:

$$N_{отн}(t) = 0,08 + 0,92 \cdot e^{-0,45t} (1 + 0,45t);$$

$$N_2(t) = e^{-2,05\sqrt{N_2(t)} \cdot t^2}$$

Выводы

1. Построение математической модели глоттохронологического процесса распада следует осуществлять путём определения её параметров по имеющейся совокупности опорных данных, которые состоят из установленных или достаточно очевидных исторических датировок и известных процентных совпадений ОС языков.

2. При использовании исходных данных следует учитывать их статистическую погрешность, которая определяет в итоге погрешность получаемых результатов модели.

3. В результате критического анализа третьего и четвёртого постулатов М. Сводеша обосновывается возможность построения глоттохронологической модели общего распада языка в виде суммы нескольких

слагаемых, соответствующих группам значений ОС с различной интенсивностью распада:

$$N_{\text{общ}}(t) = c_0 + c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t} + \dots$$

При этом среди слагаемых выделяется наиболее стабильная, практически не изменяющаяся устойчивая группа c_0 — ядро ОС.

Процесс общего распада ОС языка может быть представлен в виде суммы двух компонент: постоянной составляющей, соответствующей устойчивому ядру ОС ($\approx 20\%$ списка), и распадающейся составляющей ($\approx 80\%$ списка):

$$N_{\text{общ}}(t) = 0,2 + 0,8 e^{-0,14t}.$$

Интенсивность распада изменяющейся компоненты списка составляет $\lambda = 0,14$ и совпадает с постоянной М. Сводеша.

4. Исходя из установленной ошибочности пятого постулата М. Сводеша о независимом развитии языков-потомков после их разделения, обоснована необходимость построения отдельной модели, описывающей процесс связанного развития ОС языков-потомков.

Для этого предлагается представить процесс распада в виде суммы двух экспоненциально убывающих слагаемых: связанной составляющей $N_c(t)$ и независимой составляющей $N_n(t)$

$$N(t) = N_c(t) + N_n(t).$$

В результате получена следующая модель относительного распада языков:

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

и её упрощённый вид, справедливый при $\eta = \mu$

$$N_{\text{отн}}(t) = c_0 + c_1 \cdot e^{-\eta t} (1 + \eta \cdot t),$$

где $c_0 + c_1 = 1$; η — коэффициент потерь совпадающих слов в ОС сравниваемых языков; μ — коэффициент связанности распада ОС языков.

Определение этих параметров по фактическим числовым данным позволило получить следующее конкретное выражение для этой модели:

$$N_{\text{отн}}(t) = 0,08 + 0,92 \cdot e^{-0,45t} (1 + 0,45t).$$

5. Значения параметров моделей могут изменяться в зависимости от исходных данных, используемых при расчёте. Однако главным итогом работы представляются не значения параметров, а качественные выводы, объясняющие природу изменений в базовой лексике языка. При этом следует подчеркнуть, что предложенные соображения никоим образом не умаляют значение полученных ранее (в первую очередь, С. А. Старостиным) результатов, а привлечение новых данных

будет стимулировать дальнейшую работу по уточнению уже существующих моделей и разработке новых, более адекватно описывающих предполагаемые реальные лингвистические процессы.

Литература

1. S. STAROSTIN. *Comparative-Historical Linguistics and Lexicostatistics*. // *Time Depth in Historical Linguistics*. The McDonald Institute for Archaeological Research Publications: Printed by Blue Print, Cambridge, 2000. Volume 1, pp. 233–259.
2. А. МИЛИТАРЕВ. *Once More About Glottochronology and the Comparative Method: the Omotic-Afrasian case*. // *Orientalia et Classica*, Труды Института восточных культур и античности. М.: РГГУ, 2005, с. 339–408.
3. С. А. СТАРОСТИН. *Определение устойчивости базисной лексики*. М.: (Письменное сообщение), 2005. 9 с.
4. М. В. АРАПОВ, М. М. ХЕРЦ. *Математические методы в исторической лингвистике*. М.: Наука, 1974. 168 с.
5. А. Я. ХИНЧИН. *Работы по математической теории массового обслуживания*. М.: Физматлит, 1963. 236 с.
6. С. А. БУРЛАК, С. А. СТАРОСТИН. *Сравнительно-историческое языкознание*. М.: Academia, 2005. 432 с.

Приложение

Решение неоднородного дифференциального уравнения

$$\frac{1}{\eta} \cdot \frac{dN}{dt} + N(t) = N_0 \cdot e^{-\mu t}, \quad (10)$$

представляется в виде суммы

$$N_{\text{отн}}(t) = n_1(t) + n_2(t)$$

общего решения $n_1(t)$ однородного уравнения

$$\frac{1}{\eta} \cdot \frac{dN}{dt} + N(t) = 0,$$

а именно:

$$n_1(t) = C_1 \cdot e^{-\eta t},$$

и частного решения $n_2(t)$, вид которого зависит от соотношения между μ и η .

В простейшем случае можно предположить, что коэффициент μ связанности распада языков равен коэффициенту η потерь совпадаю-

щих слов в ОС этих языков, т. е. $\mu = \eta$. При этом условии частное решение $n_2(t)$ имеет вид

$$n_2(t) = t \cdot C_2 \cdot N_0 \cdot e^{-\eta t}.$$

Подстановка $N(t) = n_2(t)$ в (10) позволяет найти неизвестную константу: $C_2 = \eta$, и получить решение

$$N_{\text{отн}}(t) = C_1 e^{-\eta t} + t \cdot \eta \cdot N_0 \cdot e^{-\eta t},$$

где $C_1 = N(0) = N_0$, т. е.

$$N_{\text{отн}}(t) = N_0 e^{-\eta t} (1 + \eta t).$$

В общем случае, при $\mu \neq \eta$ частное решение $n_2(t)$ имеет вид

$$n_2(t) = C_2 \cdot N_0 \cdot e^{-\mu t},$$

подстановка которого в (10) даёт значение C_2 :

$$C_2 = \frac{\eta}{\eta - \mu}.$$

Результирующее выражение

$$N_{\text{отн}}(t) = C_1 e^{-\eta t} + \frac{\eta}{\eta - \mu} \cdot N_0 \cdot e^{-\mu t}$$

при $t = 0$ превращается в равенство

$$N_0 = C_1 + \frac{\eta}{\eta - \mu} \cdot N_0,$$

позволяющее определить C_1 :

$$C_1 = \frac{\mu}{\mu - \eta} \cdot N_0.$$

Для этого случая уравнение относительного расхождения двух языков:

$$N_{\text{отн}}(t) = N_0 \left(\frac{\mu}{\mu - \eta} \cdot e^{-\eta t} + \frac{\eta}{\eta - \mu} \cdot e^{-\mu t} \right).$$