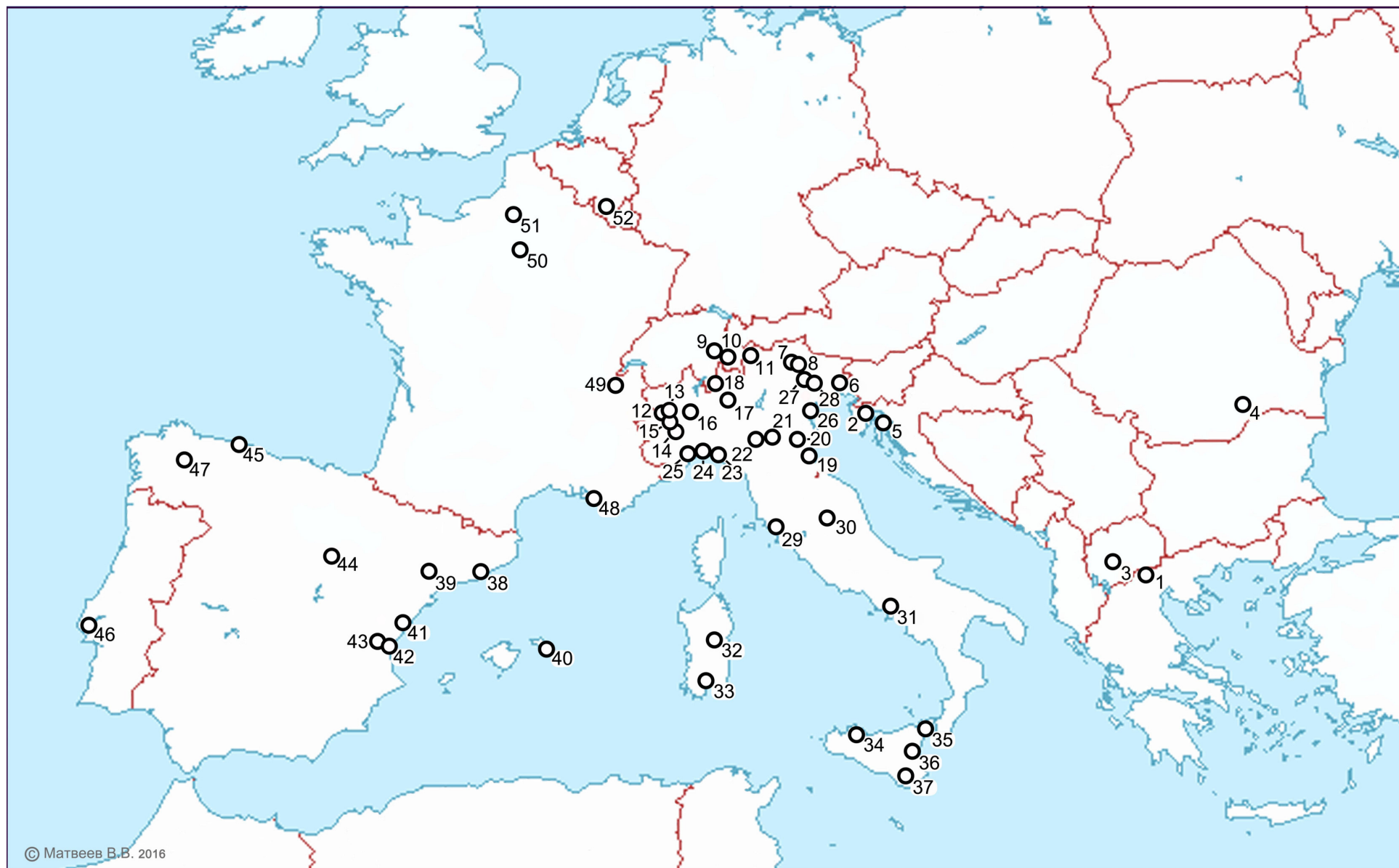


**XII традиционные чтения памяти С. А. Старостина**  
Институт восточных культур и античности РГГУ, 23-24 марта 2017 г.

Саенко М.Н., Васильев М.Е.  
Институт славяноведения РАН, Москва

**Точность глоттохронологического метода  
при датировании языковой дивергенции  
данные романских языков**

## Географическое распространение романских идиомов, используемых в исследовании



## Исходные данные для определения скорости дивергенции романских языков

(по данным 110-словных списков 52-х идиомов )

№ точки	Сравниваемые списки	Мин. % совп.	Средний % совп.	Макс. % совп.	Дата разделения, лет
1	Любые идиомы	-	100	-	0
2	Португальский – галисийский	-	97	-	1400
3	Румынский – мегленорумынский, истрорумынский, арумынский; Тосканский итальянский – умбрийский итальянский	88	89	92	1300
4	Старофранцузский (Кретьен де Труа, 1140 г. н.э.) – староитальянский (Данте, 1270 г.)	-	91	-	450 <sup>1</sup>
5	Франко-провансальский – пикардский, валлонский	86	88	89	850 <sup>2</sup>
6	Португальский – кастильский испанский	-	86	-	710
7	7.1. Португальский – фриульский, ладинские, руманшские, итало-романские, франко-провансальский, окситанский, галло-романские;	71	78	83	450 <sup>1</sup>
	7.2. Каталонские – фриульский, ладинские, руманшские, итало-романские, галло-романские;	73	79	85	
	7.3. Пикардский и валлонский – фриульский, ладинские, руманшские, итало-романские, каталонские, иберийские.	69	78	83	
	<b>Общие значения:</b>	<b>69</b>	<b>78,5</b>	<b>85</b>	
8	Румынские – остальные романские	61	69	78	300 <sup>3</sup>

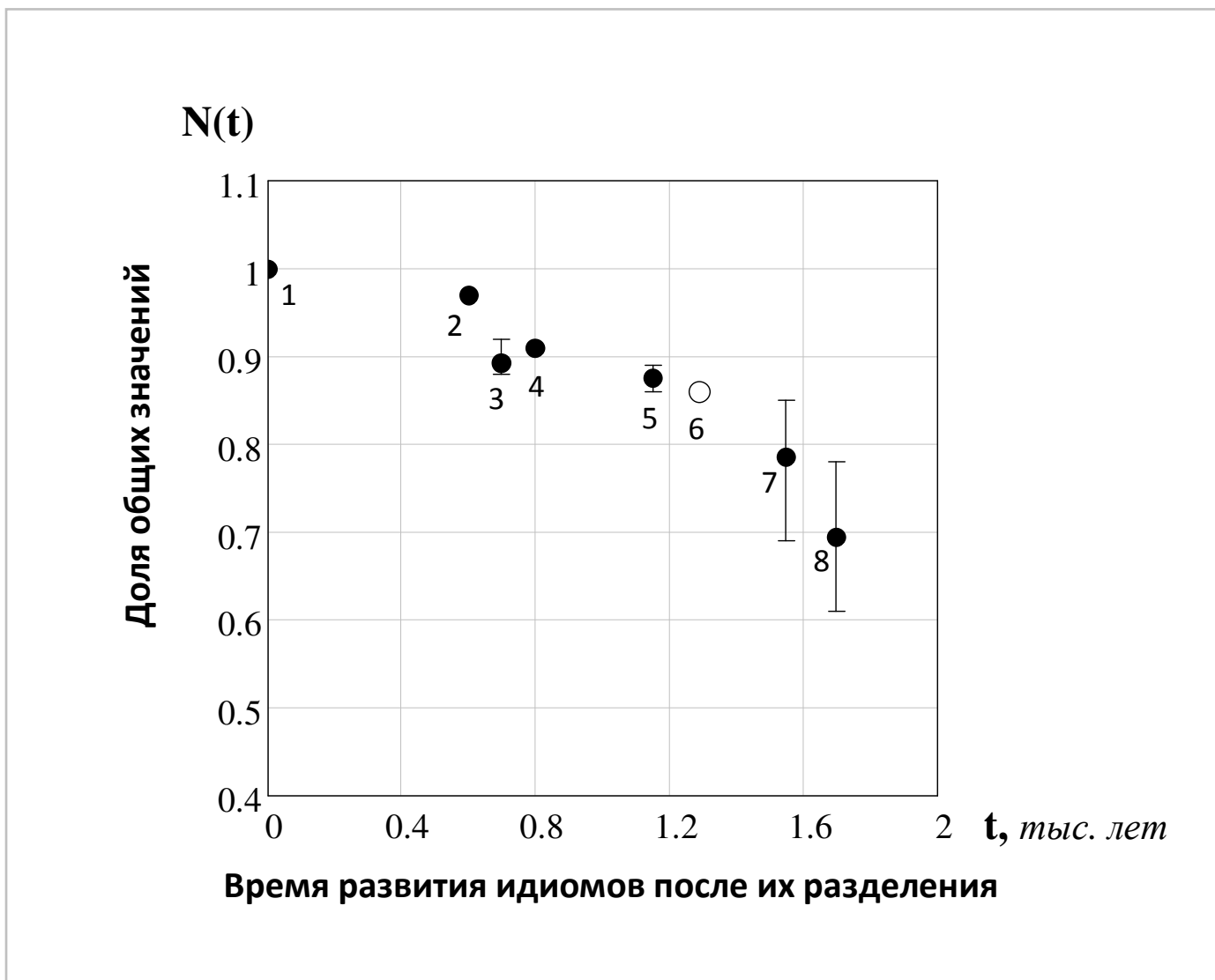
1 – распад Римской империи

2 – отделение Прованса от королевства Каролингов

3 – вывод римских легионов из Дакии

# Опорные точки для сравнения и калибровки моделей дивергенции

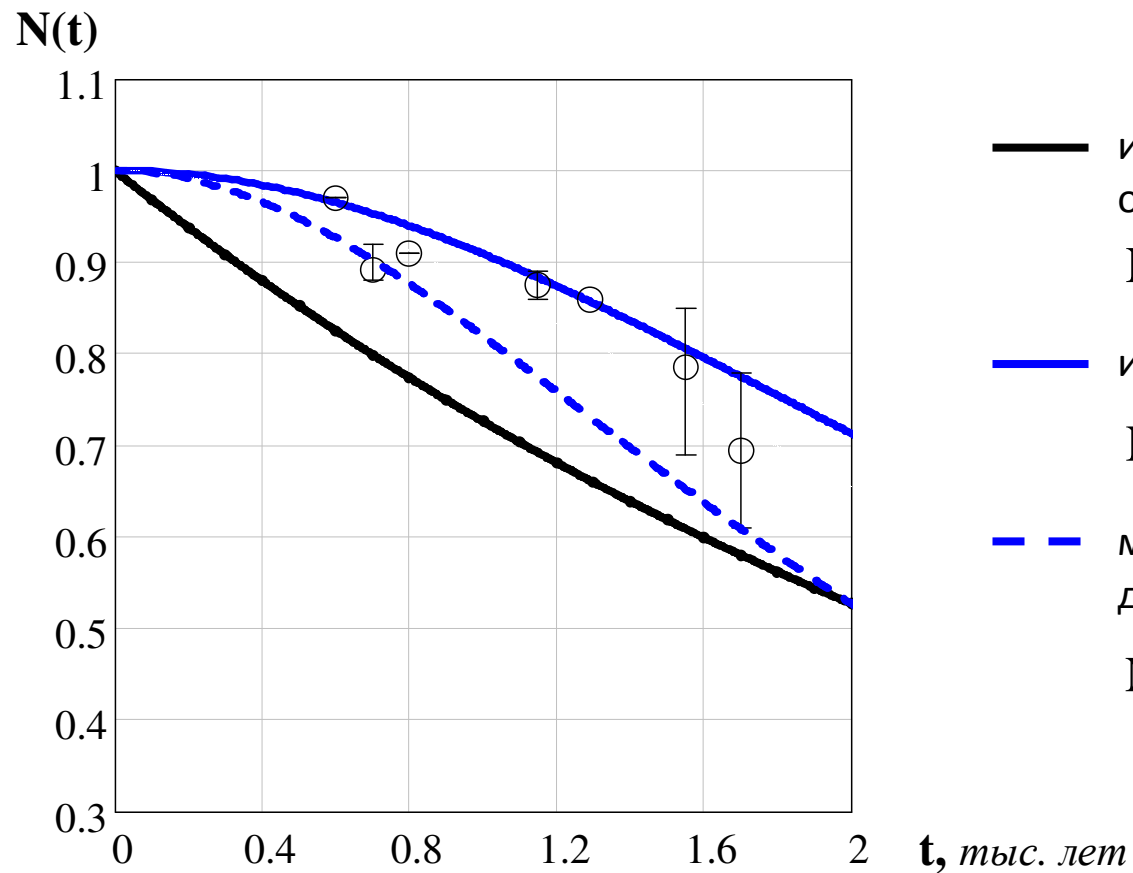
(по данным 110-словных списков 52-х романских идиомов)



## Основные задачи исследования

1. Сравнить результаты применения различных существующих глоттохронологических моделей (М. Сводеша, С.А. Старостина, потоковой модели) на материале романских языков.
2. При необходимости осуществить калибровку существующих моделей по имеющимся исходным данным (опорным точкам).
3. Определить фактическую и теоретическую погрешность рассматриваемых моделей при датировании процессов дивергенции в романских языках.
4. Сделать выводы о точности и границах возможного применения глоттохронологии на различных временных глубинах и языковом материале.

## Сравнение исходных моделей и моделей, полученных по данным общего распада



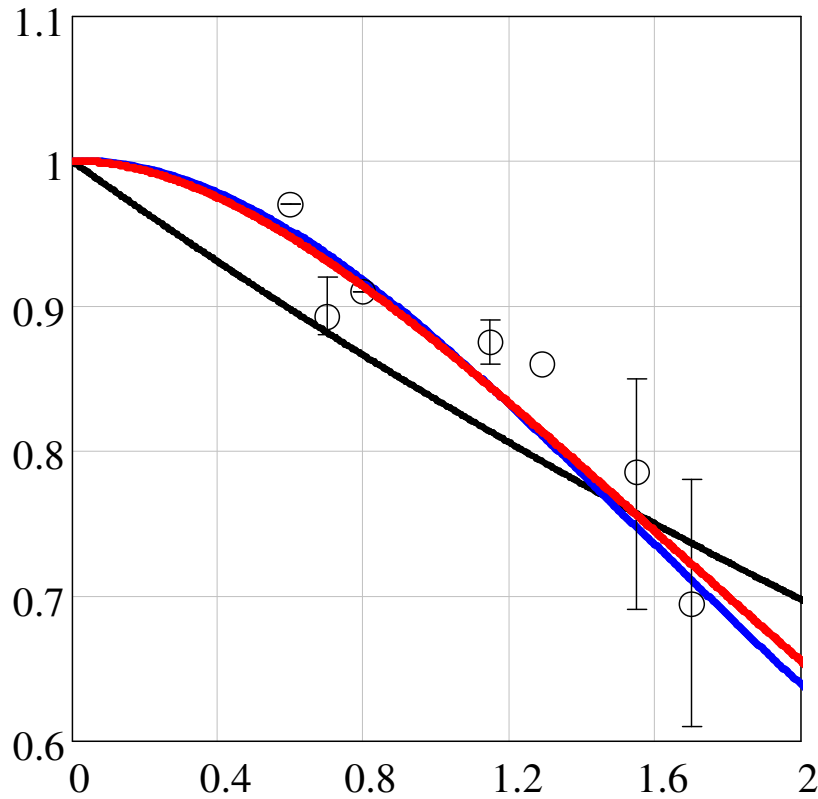
— исходная и калиброванная по данным общего распада модель Сводеша  
 $N_{Sw} = N_{SwA}(t) = e^{-2 \cdot 0,16 \cdot t}$ , ( $\epsilon = 8,55$ )

— исходная модель Старостина  
 $N_{St}(t) = e^{-2 \cdot 0,05 \cdot \sqrt{N_{St}} \cdot t^2}$ , ( $\epsilon = 2,15$ )

- - модель Старостина, калиброванная по данным общего распада  
 $N_{StA}(t) = e^{-2 \cdot 0,11 \cdot \sqrt{N_{StA}} \cdot t^2}$ , ( $\epsilon = 7,21$ )

## Сравнение калиброванных моделей

$N(t)$



— калиброванная модель Сводша

$$N_{SwC}(t) = e^{-2 \cdot 0,09 \cdot t}, (\varepsilon = 0,86)$$

— калиброванная модель Старостина

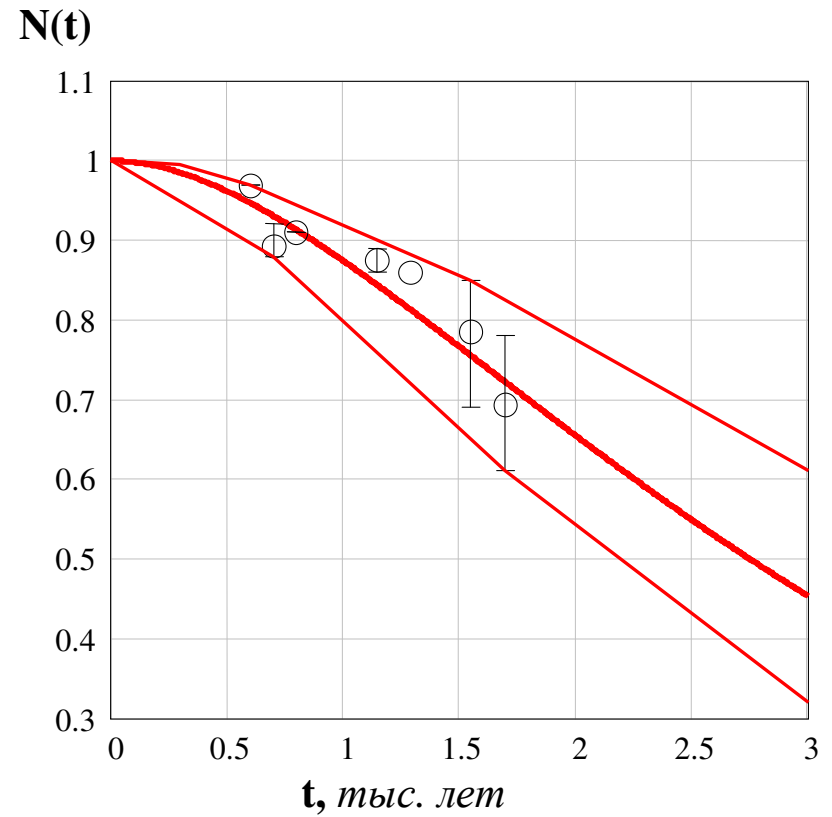
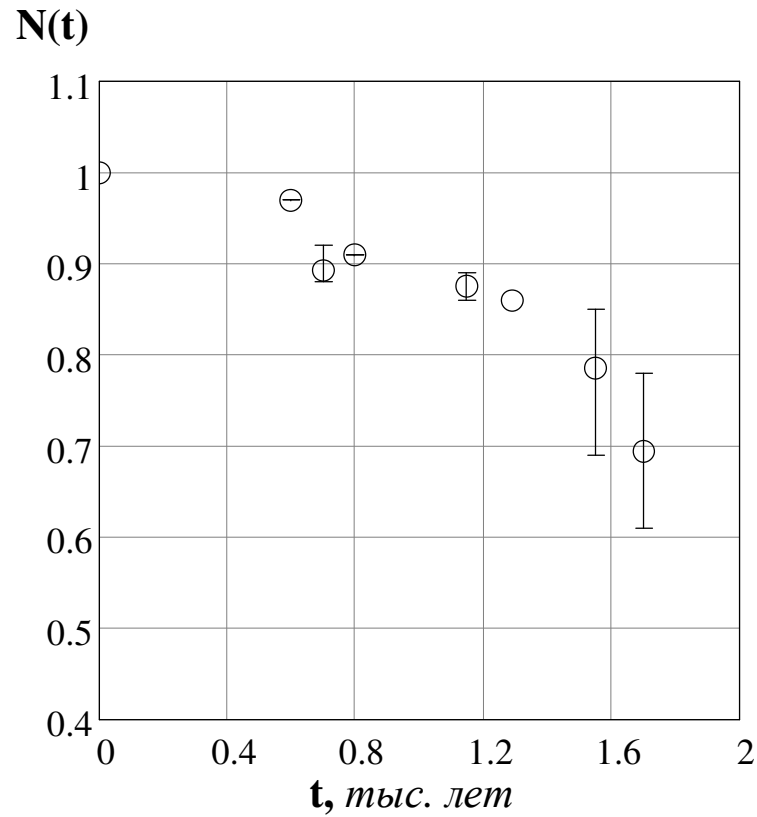
$$N_{StC}(t) = e^{-2 \cdot 0,07 \cdot \sqrt{N_{St}} \cdot t^2}, (\varepsilon = 0,89)$$

— потоковая модель

$$N_P(t) = e^{-0,61t} (1 + 0,61t), (\varepsilon = 0,67)$$

# Оценка точности глоттохронологических моделей

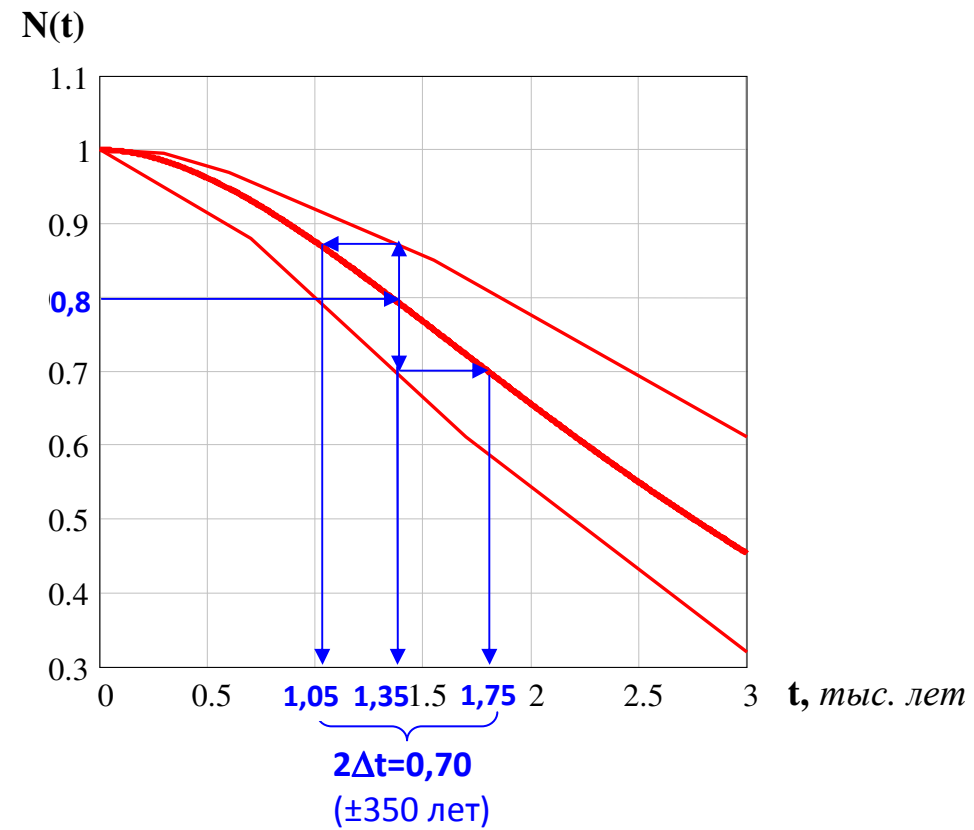
## 1. Разброс значений в исходных лексикостатистических данных





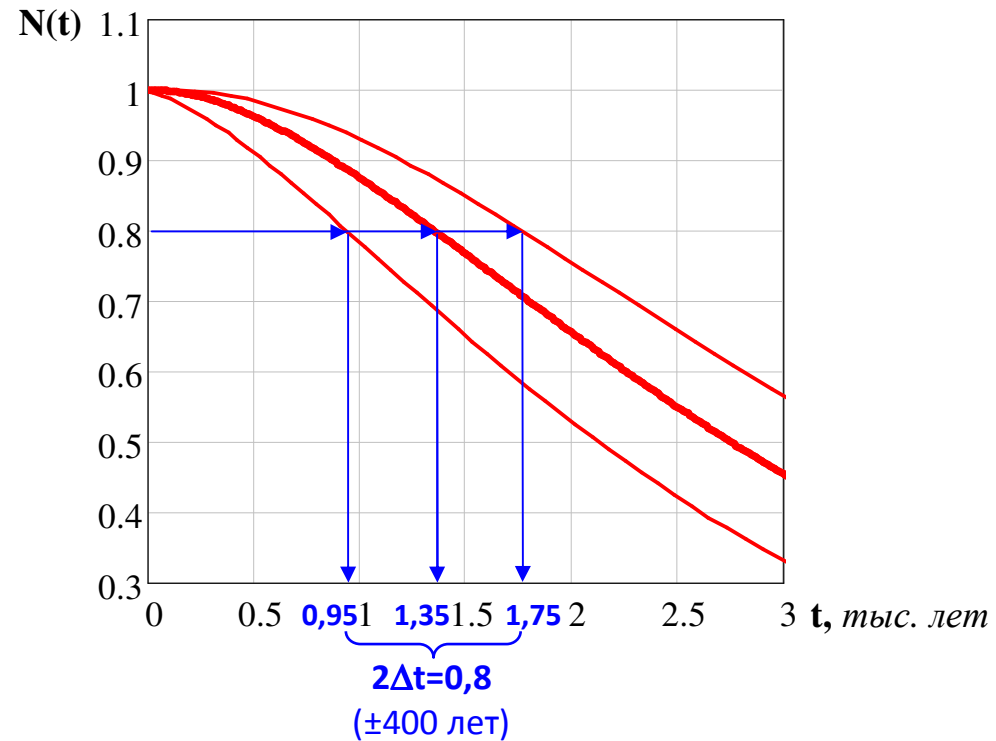
## 2. Фактическая погрешность исходных данных

### Неопределенность датировки для выбранного значения $N(t)$



### 3. Теоретическая погрешность расчетных датировок

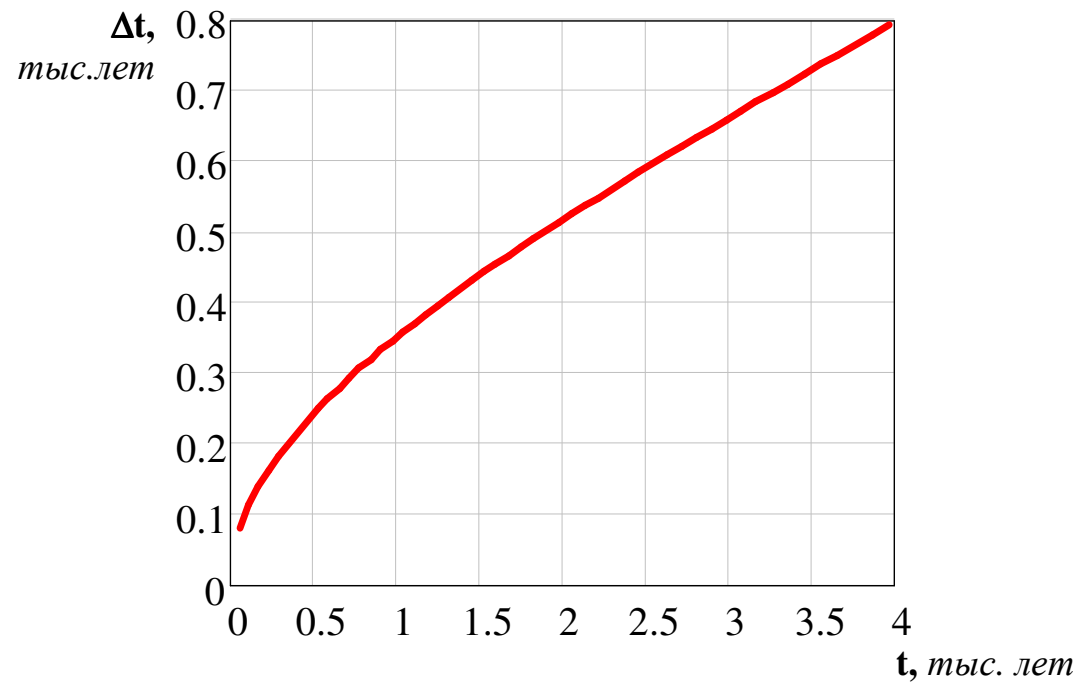
Доверительный интервал, рассчитанный для модели  $N(t)=e^{-0,16t}(1+0,61\cdot t)$  с заданной вероятностью  $p=0,7$



### 3. Теоретическая погрешность расчетных датировок

#### Изменение величины доверительного интервала со временем

для модели  $N(t) = e^{-0,16 \cdot t} (1 + 0,61 \cdot t)$  с вероятностью  $p = 0,7$



#### Расчетные значения доверительного интервала

t, тыс. лет	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8	2	2,5	3	3,5	4
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

#### 4. Датирование романских языков с помощью полученных моделей

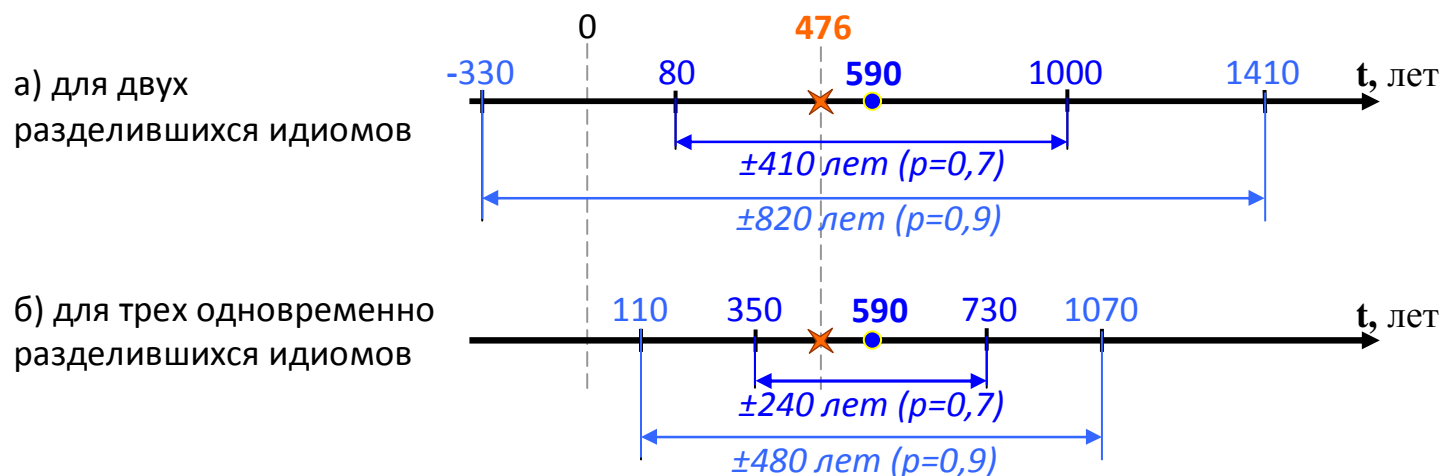
##### Датирование разделения пары языков (или групп языков)

№	Сравниваемые языки	Средняя доля совпадений	Предполагаемая дата разделения (лет)	Расчётная дата разделения (лет)	Доверит. интервал, $p=0,7$ (лет)	Доверит. интервал, $p=0,9$ (лет)
2.1	Руманшские – лигурийские	0,788	476	590	$\pm 410$	$\pm 820$
2.2	Руманшские – сицилийские	0,807	476	680	$\pm 380$	$\pm 760$
2.3	Руманшские – португальский/галисийский	0,783	476	570	$\pm 420$	$\pm 840$
2.4	Руманшские – галло-романские	0,771	476	518	$\pm 430$	$\pm 860$
2.5	Лигурийские – сицилийские	0,878	?	1020	$\pm 290$	$\pm 580$
2.6	Лигурийские – португальский/галисийский	0,803	476	660	$\pm 400$	$\pm 800$
2.7	Лигурийские – галло-романские	0,770	476	513	$\pm 430$	$\pm 860$
2.8	Сицилийские – португальский/галисийский	0,815	476	720	$\pm 370$	$\pm 740$
2.9	Сицилийские – галло-романские	0,802	476	660	$\pm 400$	$\pm 800$
2.10	Португальский/галисийский – галло-романские	0,747	476	410	$\pm 460$	$\pm 920$
2.11	Фриульский – руманшские	0,843	?	850	$\pm 350$	$\pm 700$
2.12	Фриульский – лигурийские	0,867	?	960	$\pm 310$	$\pm 620$
2.13	Фриульский – сицилийские	0,868	?	960	$\pm 310$	$\pm 620$
2.14	Фриульский – португальский/галисийский	0,830	?	790	$\pm 360$	$\pm 720$
2.15	Фриульский – галло-романские	0,820	?	740	$\pm 370$	$\pm 740$
2.16	Балканские – основной массив романских	0,696	271	183	$\pm 520$	$\pm 1040$
2.17	Исландский – фарерский	0,940	IX век	1350	$\pm 200$	$\pm 400$

## Датирование одновременного разделения трех языков (или групп языков)

№	Сравниваемые языки	Средняя доля совпадений	Предполагаемая дата разделения (лет)	Расчётная дата разделения (лет)	Доверит. интервал, $p=0,7$ (лет)	Доверит. интервал, $p=0,9$ (лет)
3.1	Руманшские – лигурийские – португ./галис.	0,791	476	610	$\pm 240$	$\pm 480$
3.2	Руманшские – португ./галис. – галло-романские	0,767	476	500	$\pm 250$	$\pm 500$
3.3	Лигурийские – сицилийские – галло-романские	0,817	?	730	$\pm 210$	$\pm 420$
3.4	Лигурийские – португ./галис. – галло-романские	0,773	476	530	$\pm 250$	$\pm 500$
3.5	Сицилийские – лигурийские – руманшские	0,824	?	760	$\pm 210$	$\pm 420$
3.6	Сицилийские – португ./галис. – галло-романские	0,788	476	590	$\pm 240$	$\pm 480$
3.7	Португ./галис. – фриульский – галло-романские	0,799	476	644	$\pm 230$	$\pm 460$
3.8	Португ./галис. – сицилийские – руманшские	0,802	476	660	$\pm 230$	$\pm 460$
3.9	Фриульский – руманшские – лигурийские	0,833	?	800	$\pm 210$	$\pm 420$
3.10	Фриульский – сицилийские – галло-романские	0,830	?	786	$\pm 210$	$\pm 420$

### Соответствие фактических и расчетных датировок, а также их доверительных интервалов



## Основные выводы

1. Из рассмотренных моделей наиболее точно процесс дивергенции описывается моделями Старостина и потоковой моделью после их предварительной калибровки. При этом попытка получения коэффициентов моделей на основе данных общего распада, как это подразумевается моделями Сводеша и Старостина, приводит к абсурдным результатам и указывает несостоятельность их исходных посылок.
2. Калибровка параметров моделей по исходным данным позволяет добиться хорошего численного совпадения расчетных и фактических датировок на временном интервале до 2 тыс. лет.
3. В силу значительного разброса в исходных данных, определение даты разделения возможно только в пределах некоторого доверительного интервала оценок с заданной вероятностью.
4. Большая относительная величина доверительного интервала на небольших временных дистанциях ставит под сомнение практическую ценность глоттохронологии при датировании дивергенции языков, разделившихся менее 1,5 тыс. лет назад. При этом верификация расчетных датировок для больших глубин невозможна из-за отсутствия достаточного количества данных, необходимых для уточнения параметров моделей.
5. Существенное повышение точности глоттохронологических расчетов возможно в первую очередь за счет расширения используемых списков, а также привлечения дополнительных данных для уточнения моделей на материале различных языковых семей и разных временных интервалах.