

## МОДЕЛИРОВАНИЕ ПРОЦЕССА ЛЕКСИЧЕСКИХ ЗАМЕН НА ОСНОВЕ ТЕОРИИ ПОТОКОВ

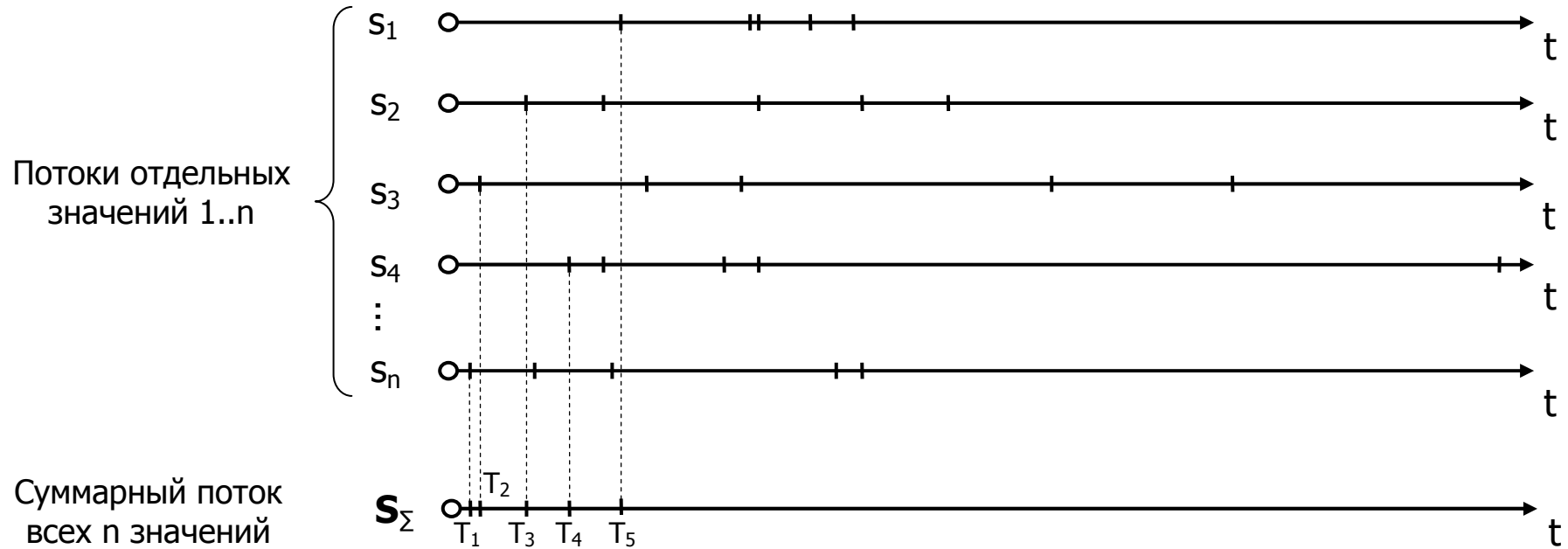
Процесс замен одного лексического значения  
как поток редких событий без последствия (поток Пуассона)



$\lambda$  – среднее число замен за единицу времени (тыс. лет);

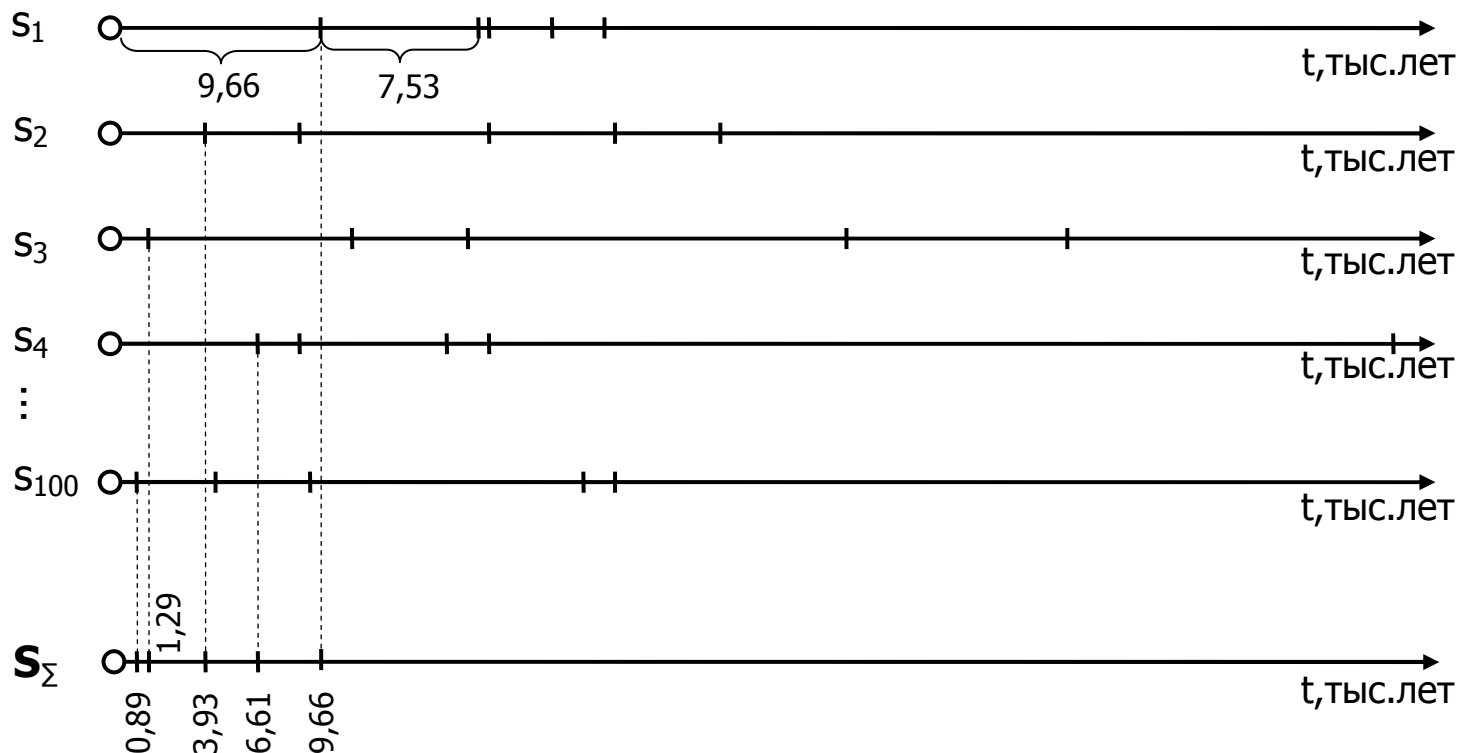
$T_{\text{ср}} = 1/\lambda$  – средний интервал времени между заменами.

Общий процесс замен в n-словном списке  
как совокупность замен каждого из значений  
(сумма элементарных потоков)

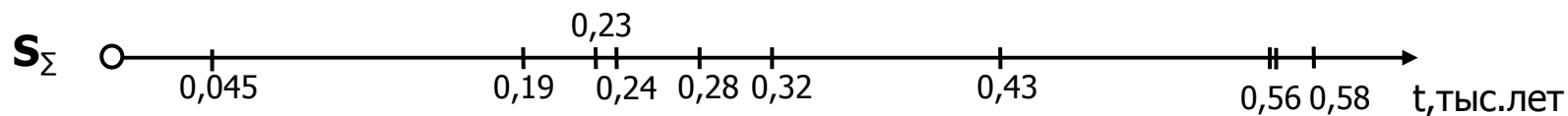


При количестве потоков  $n \geq 10$  число значений, оставшихся неизменными, имеет экспоненциальную зависимость от времени –  $N(\bar{T}_k) = e^{-\lambda \bar{T}_k}$ .

Моделирование процесса замен в стословном списке  
как суммы 100 потоков с одинаковой интенсивностью ( $\lambda=0,14$ )



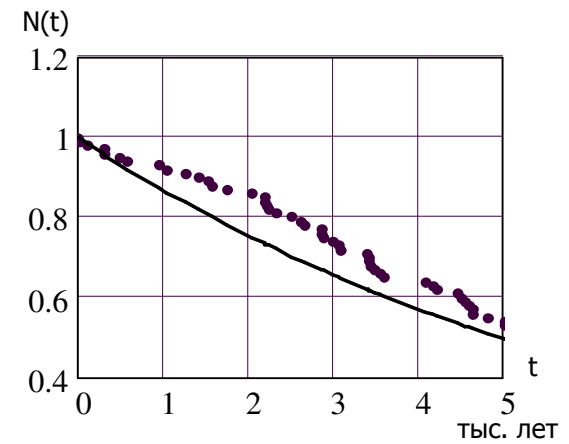
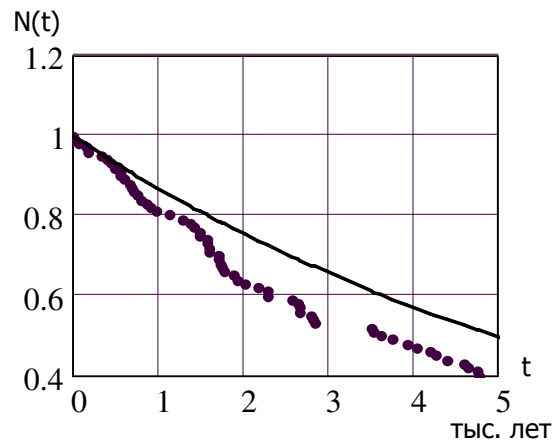
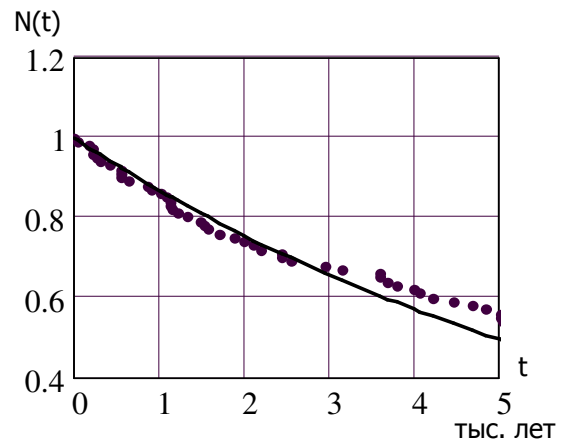
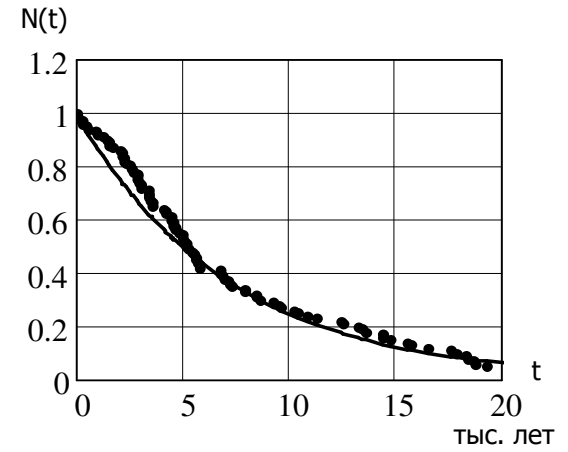
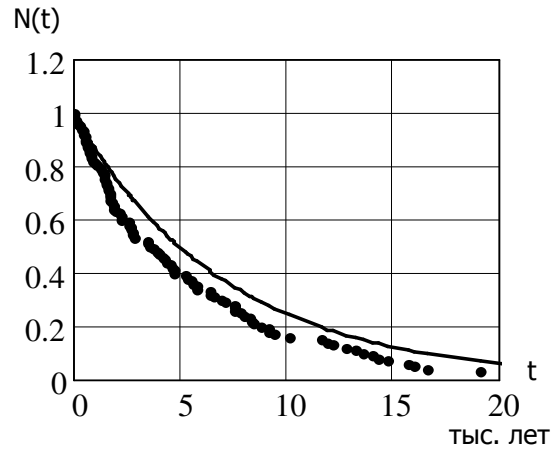
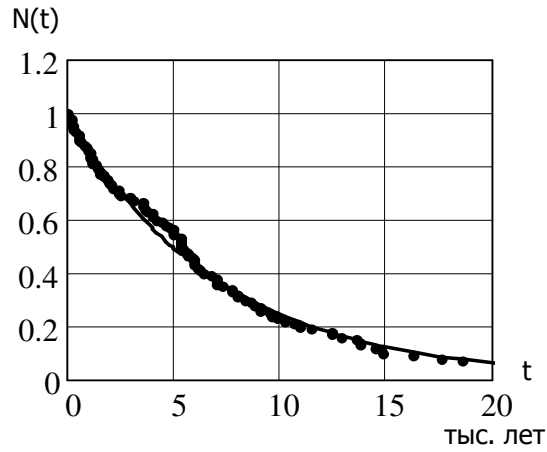
Распределение первых 10 замен суммарного потока на интервале 600 лет



Средний интервал между заменами  $T_{cp} = 580/10 \approx \mathbf{60}$  лет;

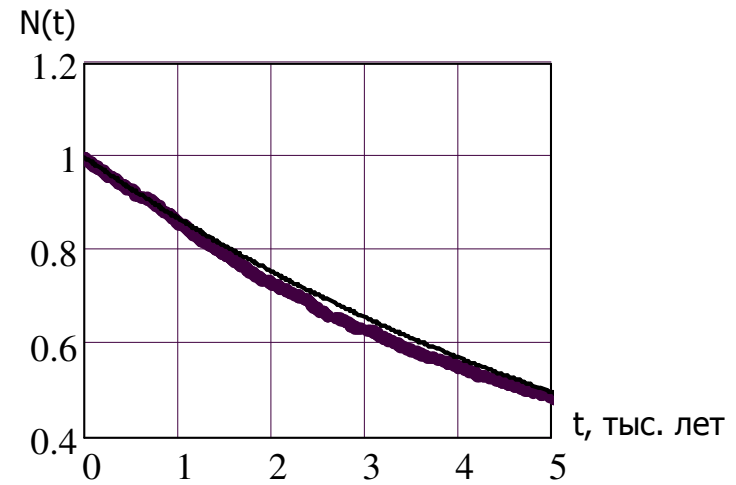
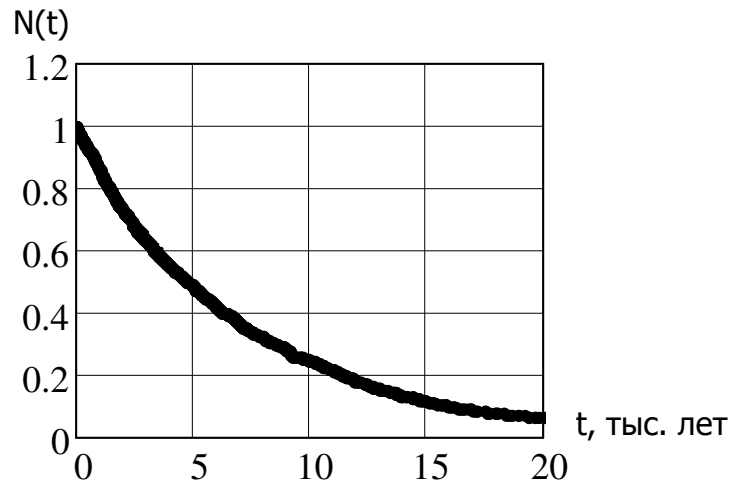
Расчетный интервал между заменами  $T_p = 1/100\lambda = 1/(100 \cdot 0,14) = 0,714$  тыс. лет  $\approx \mathbf{71,4}$  года.

Результаты компьютерного моделирования для списка из 100 значений с одинаковой и постоянной скоростью замен (стабильностью)



— — расчетная кривая  $N(t) = e^{-\lambda t}$ ,  $\lambda=0,14$   
••••• — кривая суммарного потока  $S_{\Sigma 1..100}$

Результаты моделирования для списка из 1000 значений  
с одинаковой и постоянной стабильностью



— — расчетная кривая  $N(t) = e^{-\lambda t}$ ,  $\lambda = 0,14$   
••••• — кривая суммарного потока  $S_{\Sigma 1..1000}$

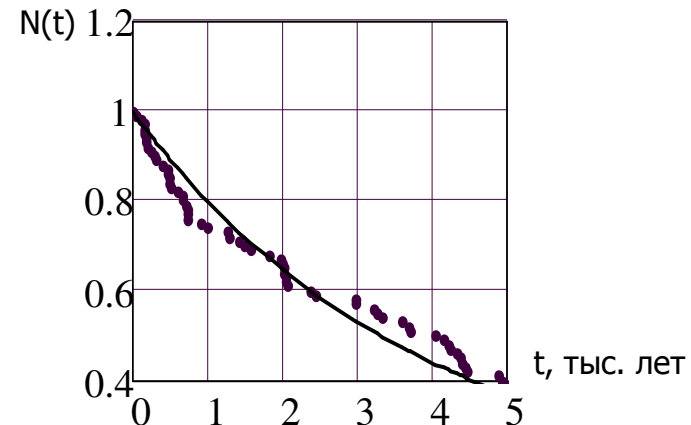
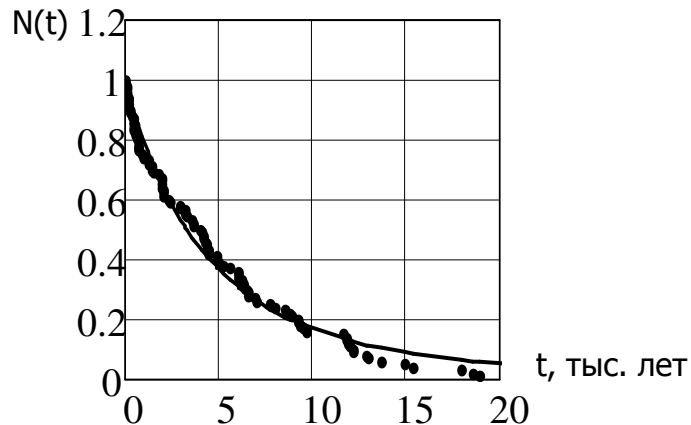
## Моделирование процесса замен для нескольких групп значений с различной, но постоянной стабильностью

Параметры модели [Merwe 1966]:

Номер группы	1	2	3	4	5	6	7	8
Количество значений в группе	2	7	17	24	24	17	7	2
Интенсивность замен $\lambda$	0,041	0,073	0,117	0,174	0,248	0,342	0,462	0,616

Результирующая формула: 
$$N(t) = 0,02e^{-0,041t} + 0,07e^{-0,073t} + 0,17e^{-0,117t} + 0,24e^{-0,174t} + 0,24e^{-0,248t} + 0,17e^{-0,342t} + 0,07e^{-0,462t} + 0,02e^{-0,616t} \approx e^{-0,2t}$$

### Результаты компьютерного моделирования

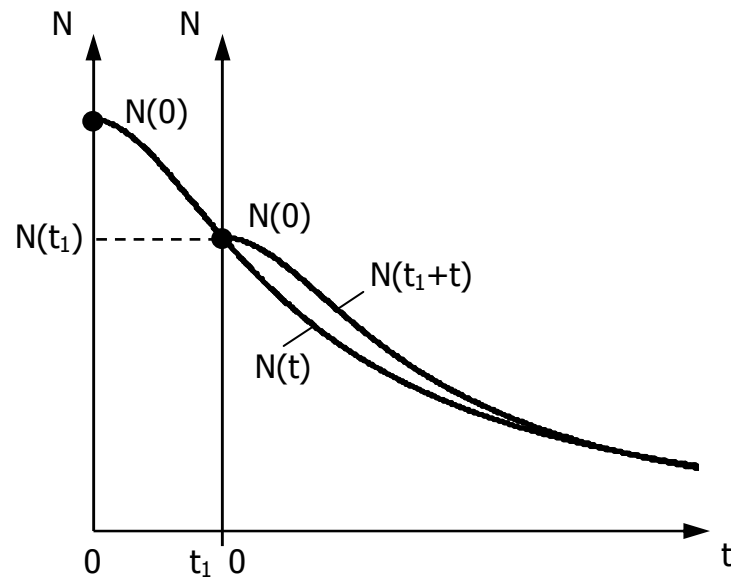


— расчетная кривая  $N(t)$   
 — кривая суммарного потока  $S_{\Sigma 1..100}$  (сумма 2-х потоков с  $\lambda=0,041$ , 7-ми потоков с  $\lambda=0,073$ , 17-ти потоков с  $\lambda=0,117$  и т.д.)

## Выполнение третьего постулата Сводеша и свойство хронологической инвариантности

Модель с изменяемой скоростью замен  
(ускоряющей и замедляющей поправкой)

$$N(t) = N(0)e^{-0,05N(t)t^2}$$

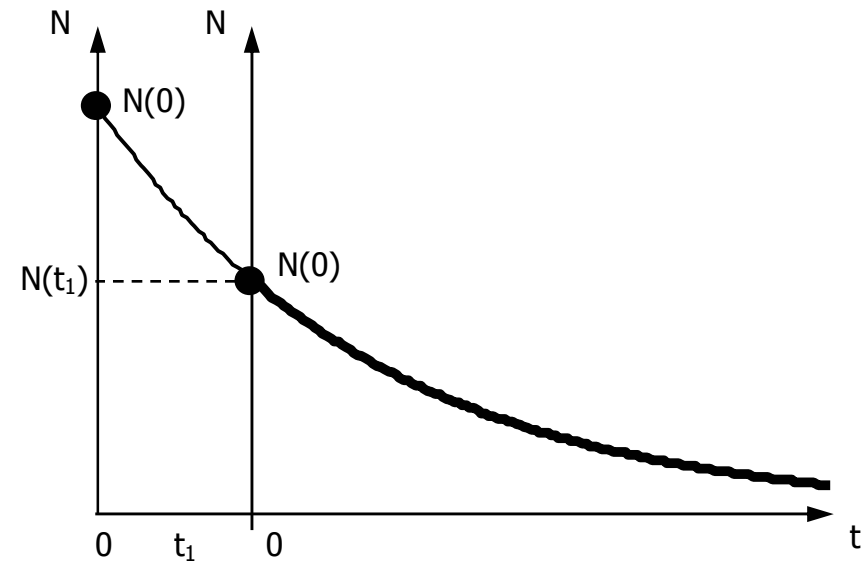


$$N(t_1+t) = N(0) \cdot e^{-\lambda \cdot N(t_1+t) \cdot (t_1+t)^2} \neq N(t_1) \cdot e^{-\lambda \cdot N(t_1+t) \cdot t^2} -$$

третий постулат Сводеша и условие временной инвариантности **не выполняются**

Экспоненциальная модель  
с постоянной скоростью замен

$$N(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + c_3 e^{-\lambda_3 t} + \dots + c_i e^{-\lambda_i t} = \sum_i c_i e^{-\lambda_i t}$$



$$N(t_1+t) = N(0) \cdot e^{-\lambda(t_1+t)} = N(0) \cdot e^{-\lambda t_1} \cdot e^{-\lambda t} = N(t_1) \cdot e^{-\lambda t} -$$

третий постулат Сводеша и условие временной инвариантности выполняются

## Полученные результаты и применимость модели

1. Предложенная модель, описывающая процесс лексических замен как суммарный поток случайных событий без последствия, полностью согласуется с эмпирическими представлениями о характере этого процесса и, в частности, удовлетворяет условиям постоянства скорости замен (третий постулат Сводеша) и хронологической инвариантности (независимости от рассматриваемого интервала времени).
2. Исследование свойств модели наглядно подтверждает возможность и целесообразность выделения в базовой лексике нескольких групп значений с различной, но постоянной скоростью изменений. При этом количество этих групп и интенсивность замен в каждой из них определяется путем калибровки модели по известным лингвистическим данным.
3. После предварительной калибровки модель может быть использована для проведения всех видов глоттохронологических расчетов, а именно: датирование процесса лексических замен в одном языке с течением времени и датирование относительного процесса замен в двух или нескольких родственных языках.
4. Благодаря статистической основе и известным математическим свойствам модели её применение позволяет численно оценить величину погрешностей при глоттохронологических или лексикостатистических расчетах, а также формировать доверительный интервал с заданной вероятностью для полученных значений (в частности, абсолютных датировок и относительного расположения узлов генеалогического дерева).
5. Совокупность результатов моделирования указывает на возможность использования предложенной модели в качестве базовой для апробации, сравнения и анализа эффективности как уже известных, так и новых методов построения и датирования генеалогических деревьев.